Identity authentication through 3D face analysis

Charles Beumier

Résumé en français

Cette thèse s'intitule « Authentification d'identité par analyse tridimensionnelle du visage ». Elle porte sur la vérification d'identité d'une personne par analyse de sa surface faciale.

Introduction

La vérification d'identité des personnes est un domaine en pleine expansion. La mobilité des personnes et le besoin d'accessibilité aux données ne cessent de croître. La sécurité d'accès aux données ou bâtiments est une nécessité pour garantir la protection des personnes et des biens. Elle requiert traditionnellement l'intervention humaine sur base de laisser passer prenant la forme d'une information connue de l'utilisateur (un mot de passe) et/ou de la possession d'un objet (une carte d'accès). Les risques ou dommages occasionnés par la perte, le vol ou la divulgation d'un de ces éléments de sécurité motivent les développements technologiques visant à exploiter des caractéristiques inhérentes et discriminantes des utilisateurs. Le nombre de solutions commerciales est élevé, exploitant par exemple la parole, l'iris, les empreintes digitales ou la main, pour n'en citer que quelques unes, mais le succès commercial est encore entaché de l'effet psychologique négatif de ces solutions. C'est dans ce cadre que la reconnaissance de visage garde son attrait issu de l'acceptation des personnes à être filmées.

Malheureusement, les performances d'un système de reconnaissance faciale sont comparativement faibles. L'analyse des images bidimensionnelles de visage souffre principalement de la dépendance à l'angle de vue et de l'influence prépondérante de l'éclairage. C'est pourquoi la reconnaissance tridimensionnelle des visages s'inscrit comme une solution prometteuse et encore faiblement exploitée aujourd'hui. D'une part, l'information de volume est faiblement dépendante de l'angle de prise de vue. D'autre part, une normalisation des niveaux de couleur mesurés est rendue possible par la connaissance des normales à la surface tridimensionnelle.

De plus, les deux types d'information, géométrique et en niveau de couleur (texture), couvrent l'ensemble des données relatives à la surface faciale. Etant donné que ces données sont assez complémentaires, leur utilisation conjointe (appelée fusion) augmente les performances en accroissant la robustesse. En effet, les données de texture sont localisées là où les niveaux de couleur subissent de grandes variations permettant leur localisation. C'est le cas des narines, des sourcils, de la bouche et des yeux. Par contre, les données volumétriques du visage sont souvent mieux estimées dans les zones à faible texture.

Enfin, les images en profondeur permettent une localisation plus aisée des visages, sur base par exemple de la position du nez, et d'une façon générale grâce aux grandes discontinuités de profondeur avec l'entourage.

La thèse s'articule en trois chapitres. Le premier présente une introduction à la reconnaissance des visages et motive l'utilisation de la troisième dimension. Le deuxième chapitre décrit en détail les prototypes de capture 3D développés au cours de cette thèse en retraçant leur évolution. Le dernier chapitre décrit les approches poursuivies pour comparer les surfaces faciales, intégrant l'information de texture et présentant les résultats de comparaison.

Chapitre 1 : La reconnaissance de visages

Introduction

Le premier chapitre introduit la reconnaissance de visages comme un moyen de remplacer les traditionnels codes ou clés de sécurité qui peuvent s'oublier ou se perdre. Malgré les différentes solutions commerciales exploitant des éléments biométriques, la reconnaissance des visages a l'avantage d'être acceptée des utilisateurs. Ce chapitre présente les critères d'évaluation d'un système de reconnaissance en général ainsi que de la reconnaissance de visage en particulier. Les approches en reconnaissance de visage repris dans la littérature sont ensuite classées en trois groupes, pour l'exploitation aussi bien des vues faciales que des profils. Finalement, l'approche tridimensionnelle apparaît comme une évolution naturelle de la recherche entreprise au laboratoire de traitement des signaux et des images (SIC) de l'Ecole royale militaire, dans le but de répondre aux limitations classiques d'une approche bidimensionnelle.

Aptitude humaine

Les méthodes de reconnaissance des visages des systèmes existants sont relativement peu performantes par rapport aux autres solutions biométriques. Leur faiblesse provient principalement de la similitude des éléments constitutifs et de leur configuration et de la dépendance des images quant à l'illumination et à l'angle de prise de vue. Mais le potentiel est relativement élevé compte tenu du fait que l'analyse des visages est considérée comme l'aptitude humaine la plus performante en matière de reconnaissance. Certains résultats d'étude par les neuropsychologues justifient des approches tantôt globales, tantôt locales, selon que l'on considère la forme générale ou les détails particuliers du visage.

Critères d'évaluation

Tout système de reconnaissance numérique doit être évalué sous différents critères pour vérifier son adéquation à l'application envisagée. Dans le cadre de nos recherches sur les visages, l'objectif principal est de vérifier l'identité d'une personne sur base de son visage dans un scénario coopératif : la personne veut être reconnue pour être acceptée. Les critères repris concernent :

- le taux de reconnaissance garantissant un certain niveau de sécurité;
- le prix et les contraintes techniques telles que les besoins en mémoire, la vitesse et la résolution du capteur;
- la convivialité et le degré de coopération requis;
- la sensibilité à l'éclairage, à la posture ou à l'expression du visage.

Chacun de ces critères fait l'objet d'une discussion. D'autres critères d'adéquation repris dans la littérature sont mentionnés. Ainsi une caractéristique biométrique est utile si elle est universelle, unique, permanente dans le temps et si elle peut être mesurée. Les applications doivent en outre considérer leur performance, leur acceptabilité et la possibilité de techniques frauduleuses.

Classification des approches

Une classification classique des approches en reconnaissance de visage distingue les méthodes globales des méthodes par caractéristiques. Les méthodes globales traitent l'image ou une transformation de l'image dans son entièreté. Les méthodes par caractéristiques analysent l'image et en extraient des caractéristiques servant à la reconnaissance. Ces deux formes de reconnaissance coexistent dans le cerveau humain. L'approche globale mène à des algorithmes concis mais n'adressant habituellement que modestement les problèmes de dépendance à l'illumination et à l'angle de prise de vue. Pour l'être humain, le processus global concerne l'appréciation générale en configuration et forme de la tête conduisant instantanément à une impression de familiarité.

L'approche par caractéristiques se construit autour d'algorithmes de détection et de mesures d'éléments constitutifs qui vont permettre de situer le visage dans l'espace des caractéristiques. Le développement algorithmique est habituellement plus fastidieux et plus dépendant de l'application mais il permet un contrôle accru sur base des résultats en cours de développement ou d'exploitation. Le processus humain correspondant est l'analyse détaillée des composants internes tels les yeux, la bouche et le nez, par ordre d'importance.

Etant donné la complémentarité des deux approches, certains auteurs proposent leur combinaison. Une troisième classe de méthodes peut être envisagée en empruntant à la classification des techniques de reconnaissance des formes la dénomination de « méthodes basées sur un modèle ». Un modèle général est adapté sur base de caractéristiques afin de permettre une comparaison plutôt globale profitant d'éventuelles contraintes introduites dans le modèle. Les références bibliographiques en reconnaissance de visages de face et de profil sont présentées dans cette thèse selon cette distinction en trois classes.

Historique au sein du SIC

L'expérience en reconnaissance de visage au sein du SIC a commencé en 1989 par une évaluation des possibilités en reconnaissance d'un réseau de neurones de type Kohonen. C'était donc une approche globale qui, bien que fonctionnant à l'échelle d'une dizaine de personnes acquises dans des conditions similaires d'éclairage et de point de vue, était

incapable de « généraliser » à d'autres vues sans nécessiter une étape de normalisation ou d'énormes moyens en apprentissage et en mémoire.
Quand on s'est proposé d'aborder au mieux la reconnaissance des visages sans contrainte de méthode, on s'est d'abord tourné vers l'analyse de la vue de face, par extraction de caractéristiques et classification. Malgré un succès rapide à l'échelle de quelques personnes, le nombre de caractéristiques simples à extraire, stables et discriminantes s'est avéré relativement réduit, proscrivant cette approche pour une application pratique encourant des variations naturelles d'illumination, de point de vue et d'attitude du sujet.
Dans notre recherche de solutions pratiques et automatiques, nous nous sommes tournés vers l'analyse de profils, les avantages escomptés étant la facilité d'extraction du contour, la faible dépendance au point de vue et la possibilité de traitements rapides et simples d'une courbe plane. Nos travaux ont conduit à la réalisation d'un prototype d'identification automatique de profils en temps réel avec un taux de reconnaissance de l'ordre de 95% pour une population d'une quarantaine de personnes. Le succès complet de ce développement nous a mené à considérer la surface du visage.

Motivation pour une approche 3D

L'analyse de la surface faciale est la continuation logique de l'analyse des profils. L'espoir est ainsi de profiter de toute l'information géométrique des visages, tout en limitant les difficultés classiques d'illumination et d'angle de prise de vue. La localisation des visages se trouve simplifiée par la connaissance de la profondeur et les leurres ou tentatives d'imposture sont plus ardues à réaliser.
Toutefois, l'approche classique d'analyse en niveau de gris ou de couleur n'est pas à négliger. Cette dernière apporte une information complémentaire à la géométrie, dans les régions texturées où la capture 3D peut s'avérer délicate. Inversement, les zones à faible texture se prêtent bien à la capture tridimensionnelle. C'est ainsi qu'une approche de type fusion, combinant la texture et la géométrie, inclut non seulement toute l'information de surface disponible des visages mais garantit par complémentarité plus de robustesse. En outre, les limitations classiques affectant l'analyse de la texture dues à l'illumination et au point de vue peuvent être partiellement compensées par l'information géométrique.
Le nombre de publications couvrant la reconnaissance de visages par exploitation des images de profondeur est assez réduit. Aucune classification n'a donc été entreprise dans cette thèse et les exemples sont présentés un par un. Il est essentiel dans ces travaux que les données 3D soient de grande qualité. La conclusion principale est qu'une contribution importante est encore attendue dans ce domaine et, eu égard à l'intérêt porté à notre base de données de visages 3D, de nombreux travaux semblent en cours.

Conclusions

Suite au succès remporté par nos travaux sur la reconnaissance automatique de profils, la reconnaissance 3D de la surface faciale du visage s'inscrit comme une continuation naturelle, d'autant plus qu'elle est une alternative prometteuse pour limiter les difficultés classiques d'analyse automatique du visage. Ainsi, une approche tridimensionnelle est faiblement dépendante du point de vue, principalement dans le cas de vues quasiment frontales, et peut être rendue faiblement sensible aux conditions d'illumination, particulièrement dans le cas d'un éclairage propre au système d'acquisition. Enfin, l'information en profondeur permet une détection aisée du visage dans la zone de capture.

Chapitre 2 : Acquisition 3D

Introduction

Notre objectif est de réaliser un système capable de capturer la surface du visage à des fins de comparaison, et répondant aux critères de fidélité, de précision et de confort pour l'utilisateur, tout en ne requérant qu'une faible et brève coopération de sa part.

L'acquisition par lumière structurée est présentée et motivée comme solution à notre application. Ensuite, le calibrage du système, étape nécessaire à l'obtention de mesures fidèles, est décrit. Les deux sections suivantes détaillent le traitement des images conçu pour extraire respectivement l'information 3D et la texture. L'évolution historique justifie la réalisation des trois prototypes réalisés au cours de cette thèse. Enfin, les bases de données constituées au cours de ce travail sont décrites et servent de base au développement réalisé en reconnaissance de visage.

Acquisition par lumière structurée

Après une présentation synthétique des systèmes d'acquisition d'images de profondeur, on montre que la lumière structurée apparaît comme l'approche la plus séduisante pour capturer des visages. Elle consiste à projeter un faisceau de lumière contenant des éléments de forme et de position connues afin de pouvoir localiser les points illuminés de la scène à l'intersection des rayons projetés et des rayons atteignant la caméra.

Les caractéristiques rendant la lumière structurée intéressante comprennent son faible coût, limité à l'adjonction d'un projecteur, sa grande vitesse, dans la mesure où une seule image est nécessaire, et la possibilité de récupérer la texture, éventuellement par une deuxième image. Le fait de disposer d'une propre source de lumière permet de rendre le système moins sensible aux conditions ambiantes, bien qu'une exposition aux fortes luminosités rendrait la lumière projetée peu visible. Le principal inconvénient de la lumière structurée est la nécessité de localiser et d'identifier les éléments projetés, ce qui peut nécessiter des traitements ardus. En outre, le projecteur requiert de l'espace et de la puissance électrique.

De nombreux systèmes exploitant la lumière structurée ont été développés. Ils diffèrent principalement par le type d'éléments projetés. Nous conférons dans cette thèse un rôle triple à la forme projetée : les éléments doivent permettre leur localisation dans l'image, leur identification parmi leurs homologues, et dans une moindre mesure, la possibilité même imparfaite de capturer la texture de la surface. Ce dernier point vise à la capture simultanée, par la même image, de la surface et de ses propriétés de couleur. Une revue de la littérature est présentée sous ces trois aspects.

Les diapositives réalisées pour les besoins de cette thèse sont ensuite largement décrites, indiquant en quoi elles satisfont aux exigences de localisation, d'identification et éventuellement de texture. Trois diapositives ont été retenues, menant aux prototypes A, B et C. Elles sont toutes trois basées sur des lignes verticales, apportant précision en localisation, et sur un mode différent d'identification (épaisseur, couleur et points). Les techniques originales de codage développées en vue de l'identification des lignes sont amplement décrites.

Le choix de la caméra et l'agencement caméra/projecteur sont justifiés à la fin de cette section.

Calibrage

Le but du calibrage est de déterminer les paramètres liés à la caméra, au projecteur et à leur agencement, de telle sorte que la capture 3D soit aussi fidèle que possible. L'approche consiste à capturer un objet de référence possédant des caractéristiques connues et d'optimiser les paramètres afin de retrouver par mesure ces caractéristiques. Le succès du calibrage dépend de la qualité du modèle du système, de la qualité de l'objet de référence, de la précision en détection des caractéristiques de l'objet et de l'algorithme d'optimisation des paramètres.

Deux formulations mathématiques sont utilisées pour le modèle. La première dérive les coordonnées X, Y, Z par trigonométrie et contient 12 paramètres optimisés globalement. La deuxième utilise la géométrie projective et résout successivement le calibrage des paramètres de la caméra et celui des paramètres restant du système.

L'évolution des prototypes A à C a conduit à des objets de référence différents. Il s'agit d'objets plans dont les points de référence sont respectivement les coins d'un carré, les coins d'un damier et les intersections d'une grille. Mis à part le prototype A, ces objets fournissent des points précis et facilement détectables. Ils permettent également une bonne visibilité des lignes projetées.

La détection des caractéristiques de l'objet comprend dans notre cas la localisation des coins de référence, l'identification des lignes de lumière projetées et l'identification des coordonnées symboliques des coins dans l'objet de référence. Les coins sont localisés, avec une précision subpixelique, par intersection des droites passant par les bords des carrés. Par construction, les lignes de lumière projetées sont presque verticales dans l'image. Autour de chaque coin détecté, quelques lignes avoisinantes sont localisées, afin de permettre l'estimation du numéro de la ligne, en valeur décimale, qui passe par le coin. Les coordonnées symboliques des coins s'obtiennent par voisinage, connaissant le coin de référence (0,0) marqué sur l'objet. Dans le cas d'objets plans, comme les lignes projetées sont espacées régulièrement, une approximation linéaire liant la position des coins et leur numéro de ligne projetée permet de filtrer les coins douteux.

L'algorithme d'optimisation des paramètres est propre à chaque prototype. Il consiste à minimiser une erreur de mesure basée sur la position mesurée des points (dépendant des paramètres) et leur position théorique sur l'objet de calibrage. Le prototype A cherche un minimum par approximations successives des paramètres. Le prototype B utilise l'algorithme du Simplexe. La méthode du prototype C commence par l'optimisation des paramètres liés à la caméra, de manière à fournir la position tridimensionnelle des coins de référence par rapport au repère caméra. Ces coins munis de leur numéro de ligne projetée permettent de déduire les plans de projection correspondant aux lignes de lumière émanant du projecteur. La méthode C prévoit également l'adaptation des positions théoriques des coins dans l'objet de référence (« Auto-calibrage ») permettant de compenser les imperfections de ce dernier.

Une interface graphique a été développée afin de permettre la supervision de la détection des coins, l'initialisation des paramètres sur base de mesures, le contrôle de l'optimisation en cours et la manipulation des données de calibrage. A tout instant, l'interface montre la valeur courante des paramètres, l'image sélectionnée pour l'interactivité avec la liste de ses coins, et la liste des images. Chaque coin et chaque image peuvent être activés ou désactivés pour l'optimisation et les erreurs estimées de chaque coin de l'image sélectionnée sont affichées.

Extraction 3D

Afin d'extraire l'information 3D des images acquises, les lignes projetées sont d'abord détectées grâce à un détecteur de transition horizontale. Chaque ligne reçoit alors une étiquette dépendant d'une propriété différente pour chaque prototype (Prototype A : épaisseur, B : couleur et C : position de points). L'identification de chaque ligne est issue de la distribution des étiquettes sur quelques lignes voisines. Quelques filtres sont ensuite appliqués pour améliorer l'identification des lignes (élimination des points isolés et cohérence de la surface). Enfin, la position image et le numéro de ligne sont convertis en position 3D.

Estimation de la texture

Le but est d'obtenir une information couleur (ou niveau de gris) de la surface à partir de l'image contenant les lignes projetées. Nous disposons ainsi du volume et de la texture en parfaite correspondance. Le motif très contrasté du prototype A ne permet pas une bonne estimation de la texture. Par contre, le prototype B utilise les lignes blanches intermédiaires pour compenser les niveaux des lignes de couleur. Le prototype C permet une compensation grâce aux zones grises séparant les lignes contenant les points.

Historique

L'aboutissement au prototype C provient de l'évolution du système d'acquisition. Ainsi le prototype A fut développé en 1997 dans le cadre du projet européen M2VTS. Il a été utilisé en 1998 et 1999 pour acquérir une base de données (voir section suivante) qui a servi aux tests de reconnaissance (chapitre suivant). L'étape de calibrage en est la principale faiblesse.

En 2001, dans le cadre du projet français BIOMET, l'approche de calibrage a été fortement améliorée, incluant un meilleur objet de référence et une interface de contrôle des points détectés et de l'optimisation des paramètres. L'utilisation d'une caméra numérique de plus haute résolution et d'un flash a conduit à la réalisation d'un prototype compact. L'introduction de la couleur dans le motif de projection a permis d'aborder la mesure de texture. Ce prototype B, alors encore en cours de développement, a été utilisé à la fin du projet BIOMET pour l'acquisition de visages 3D. Deux problèmes sont apparus. D'une part la mauvaise réalisation de la diapositive couleur et le manque de contrôle de l'éclairage (flash) ont conduit pour un cinquième des images à des captures de faible qualité. D'autre part, les images de calibrage de faible variabilité ont montré la faiblesse d'un calibrage global des paramètres.

En 2003, nous avons reconsidéré l'approche de calibrage par optimisation des paramètres de caméra dans un premier temps, et du système dans un deuxième temps. Un nouvel objet de référence, dont le contraste est meilleur, a été conçu. Finalement, une nouvelle technique de codage des lignes sur base de la position verticale de points a permis de reconsidérer la projection en noir et blanc. Le prototype C qui en découle profite du contraste accru apportant précision et robustesse dans la détection des lignes.

Bases de données

Deux bases de données ont été acquises. Le prototype A a servi à constituer la base '3D_RMA' comportant deux sessions de 120 personnes. Chaque session contient trois images de volume d'angles de prises de vue différents, dont la première, frontale, est couplée avec une image en niveau de gris sans projection. Le prototype B a été utilisé dans la troisième campagne du projet BIOMET pour capturer des données sur un groupe de 81 personnes sous six angles de vue différents. Le prototype C est encore en développement et n'a pas servi à acquérir une base de données.

Conclusions

La projection de lumière structurée comme moyen de capture tridimensionnelle se prête parfaitement à l'acquisition de la surface faciale :

- la précision, de l'ordre du mm, est suffisante pour la comparaison de visage;
- le temps de capture et de traitement, de l'ordre de trois secondes, sied aux utilisateurs;
- le surcoût se limite au prix d'un projecteur de diapositives et d'une diapositive;
- la coopération des utilisateurs est naturelle.

Les trois prototypes développés satisfont aux exigences en localisation, codage et restitution de texture (sauf prototype A) par acquisition d'une seule image de l'objet 3D à capturer. Le prototype A utilise au mieux la qualité d'image des capteurs de résolution réduite (768x576). Le prototype B exploite la qualité accrue en couleur et en résolution des nouvelles caméras permettant un codage de projection plus compact et une meilleure estimation de la texture. Le prototype C introduit un codage original et compact en niveaux de gris permettant le contraste accru de la projection et donc une localisation plus précise dans les images des caméras couleurs à masque de BAYER.

Chapitre 3 : La reconnaissance tridimensionnelle de visages

Introduction

La reconnaissance tridimensionnelle réalisée dans cette thèse consiste en une comparaison géométrique des surfaces faciales à laquelle a été adjointe une comparaison en niveau de gris. La comparaison géométrique mesure la distance moyenne entre profils obtenus par coupe plane des surfaces faciales à comparer. La comparaison des niveaux de gris mesure la similitude des différences locales en niveaux de gris le long des profils après un moyennage perpendiculairement aux profils. Enfin, une approche de fusion par combinaison linéaire des deux sources de comparaison permet d'accroître les performances en reconnaissance. Les résultats et sources d'erreur clôturent ce chapitre.

Comparaison 3D

Après avoir considéré des méthodes de localisation montrant l'influence de la qualité limitée des données et la difficulté de définir des points ou zones précises dans un visage (à l'exception du nez), une approche globale a été développée. Son principe consiste à minimiser la distance entre deux surfaces faciales en adaptant les paramètres de rotation et de translation.

Dans un premier temps, afin de réduire la charge de calcul, des profils 2D provenant des coupes planes par plans parallèles distants de 1 cm ont été extraits (adaptés pour chaque valeur de rotation et translation). La distance globale est mesurée comme la valeur moyenne des distances entre profils 2D correspondants des deux surfaces faciales. Chaque distance est calculée comme la surface séparant les deux profils divisée par la longueur du profil.

Dans un deuxième temps, essayant de profiter de la symétrie inhérente aux visages, seuls le profil central et deux profils latéraux (à 3 cm de part et d'autre du profil central) ont été extraits. Par variation de deux paramètres de rotation et d'un de translation, la proéminence du profil central et la similitude des deux profils latéraux sont maximisées. Le profil central et le profil latéral (moyenne de deux profils latéraux) sont ainsi obtenus une fois pour toute sur base intrinsèque au visage. Leur comparaison avec un autre visage revient à chercher les trois paramètres de rotation et translation restant qui minimisent la distance entre profils centraux et latéraux respectivement. Par utilisation de la pente locale le long des profils, la comparaison se ramène à déterminer le décalage (un seul paramètre) entre deux profils offrant le minimum de différence de pente locale. Le temps de calcul se trouve largement diminué, par l'extraction intrinsèque des profils, éventuellement ‘offline’ pour toutes les surfaces de référence, et par un traitement à un paramètre lors de la comparaison.

Comparaison de texture

L’information de texture complète favorablement la géométrie du visage. L’acquisition par lumière structurée étant sensible à la texture de la surface, un traitement particulier peut être réservé aux zones à forte texture. La mise en correspondance des surfaces faciales bénéficie des données en texture par l’adjonction de conditions initiales et de contraintes lors de l’optimisation. Enfin, les données géométriques et de texture sont assez complémentaires et offrent par leur combinaison de meilleures performances en reconnaissance. C’est ce dernier point qui a été exploité dans cette thèse.

L’expérimentation en reconnaissance de texture ayant été réalisée pour le prototype A, les mesures en texture ont d’abord utilisé l’image en niveau de gris obtenue avec le projecteur éteint, juste avant d’acquérir la première capture 3D. Les profils centraux et latéraux étant disponibles, les mesures en texture se font aux points image correspondants, par moyenne des niveaux des pixels avoisinant (perpendiculairement au profil) pour réduire le bruit et la sensibilité à la localisation. Pour compenser l’effet de l’éclairage, les différences locales des niveaux de gris le long du profil sont utilisées. Il en résulte un renforcement des zones à variation de texture.

La distance entre deux profils en niveaux de gris est obtenue en calculant la somme des différences en valeur absolue des valeurs homologues (point de même distance au nez). Pour tenir compte de l’imprécision de localisation du nez, plusieurs décalages sont considérés, et celui offrant l’erreur minimale est retenu.

Combinaison du 3D et de la texture

Les scores de comparaison géométrique et en niveau de gris des profils centraux et latéraux ont été combinés linéairement afin de conduire à des valeurs minimales des taux de fausses acceptations et de faux rejets. Les coefficients de la combinaison linéaire sont adaptés pour offrir la meilleure séparation des scores de tests ‘client’ et ‘imposteur’ d’une partie des données de comparaison. Le reste de ces données est utilisé pour l’évaluation des performances avec les coefficients obtenus.

Résultats

De nombreux tests ont été effectués avec le prototype A. Pour l’approche globale, des taux d’erreurs égales (EER) de 2 à 5 % ont été obtenus, selon les sessions de données considérées. La méthode basée sur les profils central et latéral a donné des résultats plus modestes. Elle a toutefois facilité l’adjonction de l’analyse en niveau de gris menant, après fusion, à des taux d’erreurs égales d’un peu plus d’1% dans le cas de données frontales ou par fusion de plusieurs vues.

Les expériences de comparaison des données BIOMET (prototype B) ont mené à des taux d'erreurs égales de 2 % pour la méthode globale et de 7.5 % pour l'approche par profils central et latéral. Ces taux sont meilleurs que pour le prototype A car ils concernent des données pour lesquelles l'intervention manuelle s'est limitée à quelques cas pour l'extraction de profils et aucunement pour la comparaison. L'analyse en niveau de gris devrait contribuer à une importante réduction de ces taux d'erreurs.

Sources d'erreur

Les résultats obtenus dépendent des différents acteurs intervenant dans le processus complet de reconnaissance. Ainsi, la coopération des utilisateurs est primordiale pour éviter les difficultés liées au positionnement et au respect des conditions (expression neutre, pas d'écharpe). Ensuite, le prototype de capture 3D commet des erreurs systématiques liées au matériel (e.g. optique) et à son calibrage et des erreurs plus aléatoires dépendant des conditions d'acquisition (texture, couleur) et de la qualité du processus d'analyse des images. Quant à l'algorithme de reconnaissance, impliquant la minimisation de distances, il induit dans la comparaison géométrique ou en niveaux de gris des erreurs ou imprécisions dues à la présence possible de minima locaux. Finalement, les performances mesurées dépendent fortement des données présentes dans la base de données et de la représentativité de la population considérée au sein de celle-ci.

Conclusions

Une méthode de reconnaissance 3D de visages a été présentée dans ce chapitre. Elle combine linéairement les scores issus de la comparaison de l'information géométrique et en niveau de gris extraite des images striées acquises par le système d'acquisition 3D. La comparaison géometrique s'effectue par comparaison de profils plans, obtenus de préférence après une phase de normalisation qui exploite la symétrie naturelle des visages. La comparaison en niveaux de gris considère les valeurs de gris le long des profils extraits pour l'analyse géométrique, après un moyennage effectué perpendiculairement au profil. Les profils en niveau de gris à comparer s'obtiennent par différence le long des profils pour réduire l'influence de la lumière ambiante.

Les résultats obtenus confirment la validité de l'approche et l'intérêt de combiner l'information de volume et de texture, même simplement. La méthode globale est plus robuste et plus discriminante. Mais l'approche par profils central et latéral a reçu plus d'attention grâce au temps de calcul réduit.

Conclusions et perspectives

Cette thèse présente une approche complète de comparaison tridimensionnelle de visages. L'objectif poursuivi est d'exploiter l'information de volume et sa robustesse afin de réduire les limitations classiques en reconnaissance de visage, à savoir la dépendance à l'angle de prise de vue et à l'illumination. Deux contributions majeures charpentent ce travail : l'acquisition de surface faciale et la comparaison des surfaces acquises.

L'acquisition par projection de lumière structurée convient à la capture de surface faciale. La surface acquise par une image frontale couvre la majeure partie de la surface utilisable. La précision en profondeur de l'ordre du millimètre semble satisfaire aux besoins en reconnaissance. La rapidité d'acquisition, l'attitude et le positionnement face au système remplissent les conditions élémentaires d'une utilisation pratique. L'évaluation qualitative lors de l'acquisition de base de données a révélé la bonne qualité générale des systèmes développés mais aussi les limitations inhérentes à la présence, par ordre d'importance, de barbe ou moustache, du nez et des yeux.

La reconnaissance des visages acquis par le système a été réalisée par comparaison de surface requérant l'optimisation des six paramètres de translation et rotation. Dans l'approche globale, les six paramètres sont modifiés afin de réduire la distance entre profils homologues obtenus par coupes planaires parallèles. Dans l'approche par profils central et latéral, la symétrie intrinsèque des visages est exploitée pour normaliser trois de ces six paramètres et extraire les profils central et latéral. Les trois paramètres restant sont optimisés lors de la comparaison des profils extraits. Une transformation des courbes en mesure de pente locale permet une comparaison rapide ne dépendant que d'un seul paramètre. Une comparaison en niveaux de gris a été intégrée dans cette deuxième approche en mesurant les niveaux de gris le long des profils géométriques extraits et en les normalisant par différence le long des profils pour réduire l'influence de la lumière ambiante ou projetée.

Les tests de comparaison ont montré les possibilités de reconnaissance, par extraction et comparaison automatique des profils. Les performances sont d'autant meilleures que les vues sont frontales. Ces dernières impliquent en effet une meilleure visibilité du visage et une information plus cohérente entre acquisitions. L'adjonction des données relatives à la comparaison des niveaux de gris augmente largement les performances grâce à l'information complémentaire ajoutée pour un coût modique lié à l'acquisition ou à la comparaison. Les principales causes d'erreur proviennent des défauts de capture de la surface faciale réduisant la qualité de l'information ou la symétrie des visages, et les problèmes d'optimisation lors de la comparaison des profils, liés à la présence de minima locaux.

Les résultats de reconnaissance des visages 3D peuvent être améliorés. De façon générale, l'acquisition profiterait d'une analyse en niveaux de gris visant à éviter les zones difficiles à acquérir par lumière structurée. De plus, le prototype C, conçu mais non encore exploité, devrait bénéficier d'une meilleure reconstruction de la surface faciale et, de façon secondaire, de la texture. Ces améliorations conduiraient immanquablement à une comparaison plus robuste et plus performante des profils géométriques et en niveaux de gris. Une comparaison simultanée de l'information géométrique et de texture conduirait également à de meilleures performances en réduisant l'incidence des minima locaux. Enfin, les résultats de reconnaissance dépendent fortement des contraintes d'acquisition et de la fidélité des données de références. Des représentations plus complètes, lors de l'enregistrement ou lors de tests, éventuellement par acquisition de séquences, assureraient une comparaison plus robuste. Sans réelle contrainte pratique, des acquisitions purement frontales, somme toute naturelles pour l'utilisateur, augmenteraient la qualité des surfaces acquises et de leur comparaison.

Acknowledgements

I would first like to thank Prof Dr. Ir. Marc Acheroy, director of this thesis, and head of the *Signal & Image Center (SIC)* of the Royal Military Academy in Brussels where I am employed. Thanks to his management assets, he has been able to find the contacts and projects to allow me to go on working in the field of Face Recognition. Through his social skills, he was able to create a research team with a good spirit and a good atmosphere inviting people to stay.

The harmony of the research team is also due to the participation of all its members. Let me thank everyone for their help at any level. More specifically, I am thankful to Vinciane for her continuous support and highly relevant advice, to Pascal for deep and constructive discussions and Yann for his expertise in calibration.

The last three years have been spent with the improvement of the 3D acquisition system, thanks to a collaboration with the *Image and Signal Processing department* of the ENST (Ecole Nationale Supérieure des Télécommunications) in Paris. Let me thank people who made this collaboration possible such as Prof. Dr. Ir. Francis Schmitt, who accepted to be director of this thesis at ENST, and Prof. Dr. Ir. Henri Maître, head of the department. I am also thankful to all people at ENST for their kindness and willingness to help me in any respect. I was particularly happy to share personal and technical discussions with Francis Schmitt and Hans Brettel and to receive support in technical and administrative matters from Marc Sigelle. I enjoyed the participation in the BIOMET project, which allowed me to further develop the 3D acquisition system, and which gave me the opportunity to establish valuable contacts.

The *Systèmes Logiques et Numériques* department of the Université Libre de Bruxelles (ULB), headed by Prof. Dr. Ir. Philippe Van Ham, is in close contact with SIC. I am delighted to have benefitted from the university's expertise in 3D capture and analysis, namely in the persons of Nadine Warzée and Cédric Laugerotte.

Last but not least, I would like to thank my family and the people around me not already cited for their patience and support all along this thesis.

Contents

Introduction

The emergence of information access (internet) and the increasing automatisation of daily tasks raise the problem of security. The traditional password and PIN (Personal Identification Number) as identity verification are outdated and we are looking more and more for biometric solutions, which are not subject to loss or theft. Some of them are technically proven, such as the fingerprint or iris scan, but tend to be perceived as intrusive, and thus not acceptable by the user. Face analysis appears as an appropriate alternative, provided that adequate recognition rates (in terms of false acceptance and false rejection) can be guaranteed.

This thesis evaluates the 3-dimensional analysis of the human face as a way to achieve person recognition. The classical frontal analysis typically requires illumination and pose normalisation. Using makeup or presenting a picture in front of the camera easily fools frontal systems. The profile approach, which analyses the geometry of the external outline composed of rather static parts (such as the forehead, the nose and the chin) is naturally more adapted to a correct normalisation. This explains the success of many profile systems, but the information provided is limited to a 2-dimensional curve and can be fooled by a simple drawing.

Our research is based on the assumption that dealing with more geometrical information - as opposed to grey-level values – is an appropriate way to address the common problems of illumination and pose normalisation. A system analysing the geometrical aspect of the facial surface will be able to determine the pose, compensate for illumination effect like shading and gather more information. Moreover, volume acquisition and analysis will be easier in the smooth parts (forehead, cheeks and chin), precisely where grey-level information lacks. And last but not least, 3D analysis allows for certain quantitative measures such as a nose length, leading to performances in a field where the human ability is poor (one can judge of the nose length, but can not accurately estimate its size at sight).

The volume approach for face recognition is not new, but still relatively recent, certainly due to the high computational power required by 3D capture, display and analysis. 3D sensing, originally performed by scanners, was slow, expensive, and possibly dangerous when using LASER. In the last few years, new developments have contributed to the creation of 3D acquisition systems that are fast, cost-effective and accurate. As a consequence, the amount of 3D research in face analysis has increased.

The purpose of this thesis was to design a realistic face recogniser based on volume analysis. It had to be as automatic as possible, low-cost and fast while meeting high recognition performance standards. The additional possibilities such as fusion with grey-level information or dynamic analysis are possible ways to increase performance. Dealing with range images simplifies the localisation of the face. Also, the persons to be identified are assumed to be cooperative: they wish to gain access to the secure service or area and will not hide from the camera.

By the time of the early developments of this thesis, 3D face recognition approaches reported in the literature relied on high quality facial surfaces and available 3D acquisition systems were expensive and rather slow. We decided to realise a low cost and fast acquisition prototype based on structured light. A geometrical approach for 3D face recognition was developed to show the 3D recognition ability, even for low quality facial surfaces. Texture information was integrated in order to improve recognition performance with low cost and development penalty. Finally, the original 3D acquisition prototype was improved to better acquire the geometry and the texture, so that three original prototypes were designed and developed.

The document is split into three parts. Chapter one introduces face recognition in the context of cooperative access control and gives the reasons for adopting the 3-dimensional face recognition approach. The second chapter provides a detailed description of the three prototypes developed along the thesis, regarding hardware, calibration and acquisition software. The third chapter describes the 3D face recognition algorithm, integrating texture information, and presents recognition results.

The main achievements of this thesis are:

- The design and development of three prototypes for 3D face acquisition based on structured light, with original projected patterns
- The acquisition of two 3D databases
- The development of a 3D face recognition method based on 3D surface distance
- The integration of texture information in the 3D face recognition method
- The realisation of a graphical interface for calibration allowing control of reference points and parameter optimisation.

Chapter 1 Face recognition

1.1 Introduction

The famous human ability for face recognition has intrigued many researchers and we will mention some results of the work carried out by neuroscientists and psychophysicists in section 1.2. This natural ability may explain why many computing paradigms have addressed face recognition as a challenge.

The interest in automatic face analysis, however, is older than the birth of computers. The "Nature" magazine reported on the work of Sir Galton [Galton10] (early 20th century) on "*Numeralized profiles for classification and recognition*" for forensic purposes. The need for biometrical security is certainly much higher nowadays and concerns nearly everyone, due to the increasing mobility of people and information.

The computer evolution brings with it better recognition systems, but also leads to greater data accessibility and a higher demand for automatisation, both of which increase security needs. This thesis contemplates biometric recognition in realistic situations, using automatic (computerised) recognition of humans in cooperative scenarios. The complications which arise in unconstrained scenarios are considerable. Therefore, only applications implying user cooperation (willingness to enter a room or access certain services) are analysed here. Section 1.3 details typical system requirements for a practical solution.

Biometric clues reduce practical problems, such as loss or theft, of trivial security means such as PIN (Person Identification Number), password or physical objects (keys). Many technical solutions exist today and some are already commercialised (fingerprint, face, hand geometry, iris, retina, speech, ... refer to [Jain99]).
But user preferences work in favour of face or speech related approaches, since these are non-intrusive and people may be more used to being video-recorded. This is why a brief comparison of biometrical systems is given in section 1.4, while a more complete presentation of the face and profile activities is given in section 1.5. The latter includes a summary of the face recognition research carried out in our department. 3D analysis, which is the principal subject matter of this thesis, is the natural continuation of frontal and profile based approaches. It complements those earlier developments and adds interesting new features.

1.2 Human abilities

The study of the human face recognition ability [Ellis86a, Young89] has recognised this aptitude as a ***dedicated*** high-performance ***brain process*** that we can use as a reference for computer developments. For instance, discovering five types of cells in the macaque brain, each selective for one prototypical view (face, profile, back, head up, head down) [Perrett86] could support Beymer's work on face recognition under varying poses, based on 15 views used to model a person's face [Beymer94].

According to [Blanc-Garin86, Ellis86, Laughery89], the brain makes complementary use of featural and global (called 'holistic') approaches to recognise human faces. The ***global*** approach is used as a rapid indication of feeling of familiarity [Blanc-Garin86] and is obviously used at long distance where the resolution lacks. External parts (hair, outline) have a significant impact on the global impression and are particularly important for the recognition of unfamiliar faces.

On the other hand, the statistical (***featural***) approach, which analyses inner components, is used critically at short distance. Identification of familiar faces, as well as communication, naturally use this way as it relies on expressive parts (eyes, eyebrows and mouth). As described in [Blanc-Garin86], the feature analysis, concentrating on each part to extract characteristics, takes greater effort and more time. Considering the importance of facial parts for human recognition of faces, precedence is given to the head outline, followed by the eyes. The feature saliency of the mouth is not very important and the nose seems to play nearly no role at all in frontal face recognition [Haig86, Fraser86].

1.3 Typical system requirements

We detail in this section the typical requirements a face recognition system has to fulfil in real applications. The list is not exhaustive and is not ordered by importance. The weight of the different criteria is highly dependent on the applications. When meaningful, a comparison with the human abilities is given.

1.3.1 Face detection

Face related processing such as face tracking, expression analysis or face recognition must be preceded by a phase of face detection and localisation. Should the face be improperly localised, any dependent face processing algorithm would fail. This topic is a very active field of research as reviewed in [YangM02]. In many studies, either face detection and localisation is assumed to be already resolved, or the face recognition algorithm encompasses face detection.

1.3.2 Recognition

Recognition ability is of course what is expected from the systems under analysis. However, the performance of current face recognition applications is far from being equal to human capabilities, especially when some of the criteria described below are considered. But most specific applications do not need such a high performance. The cooperation of the user, the control of some parameters (distance, light, limited number of persons) and the addition of other sources of information such as speech make the recognition rates of current face modalities viable.

We commonly distinguish two recognition tasks, namely *verification* and *identification*. For verification, the person claims his or her identity by some other channel such as voice, typing, password or badge. The system has to check the validity of the claim. This normally leads to a reduction in the number of comparisons. On the other hand, an identification system has to guess who is standing in front of the camera. The whole database of references is then implicated in the comparison.

The performance of a recognition system is classically quantified by the error rates. Two types of errors are evaluated: false acceptance (FA), corresponding to the acceptance of an impostor and false rejection (FR), the erroneous rejection of a client. Relative to the total number of tests, these errors are respectively called False Acceptance Rate (FAR) and False Rejection Rate (FRR), normally expressed as a percentage.
Every recognition system is a trade-off between a low FAR and a low FRR. The application can give some clues to making a good choice. For instance, the FAR must be very low in a high security environment or for an Automatic Teller Machine, where the penalty for accepting an inappropriate person is very high, although a client can accept a second or third trial when rejected. On the contrary, forensic applications need low false rejection rates to ensure a criminal will be caught, even though more falsely accepted candidates have to be examined.

The curve representing the FAR versus FRR is called the Receiver Operating Curve (``ROC curve'', Figure 1). It is established by computing the FAR and FRR for several values of the parameter controlling the trade-off. One particular point is usually extracted from this curve: the EER (``Equal Error Rate''), which is the point where the FAR equals the FRR (intersection with the bisectrix of the axes). This rate gives an estimation of the performance and can be used (with caution) as comparison criterion.

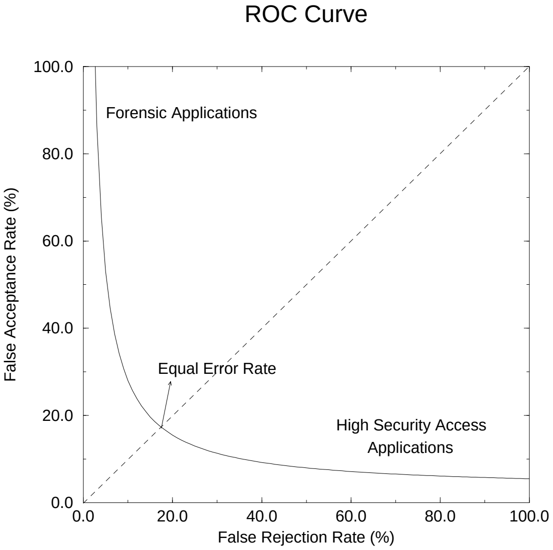


Figure 1: Typical ROC curve

Another performance value that is often determined is the recognition ratio, which equals 100 - FAR - FRR, expressed as a percentage. The value of course depends on the chosen operating point (value of the parameter controlling the trade-off).

The evaluation of the performances of a recognition system depends heavily on the tests performed, their conditions and on the representation of the clients. A representative database is a key element for a correct estimation of what the performances can be in real conditions, provided that the test conditions (lighting, position, attitude, etc.) are similar. A good representation of individuals must take into account the versatility of people present in the population as well as the natural variation in the presentation of each person. For this, several shots per person are very valuable, so that posture, illumination or face variation can be accounted for.

1.3.3 Applicability and user-friendliness

In the research community, there is room for investigating certain promising methods, although they may currently be inadequate for the user under real conditions. The natural evolution of science, however, may resolve some of the current problems regarding applicability over time.

The second aspect, user-friendliness, is a commercial criterion rather than one specific to research. Nevertheless, a user-friendly solution facilitates testing due to better user cooperation. Face analysis seems to be well accepted by users. No contact is required. Cooperation is normally natural. No learning is necessary.

1.3.4 Price

The production price of a technique, if primordial for a commercial application, is of secondary concern to the researcher. But in the spirit of a realistic application, we have to ensure that the price remains acceptable. With the high demand for security systems and the possibility of producing large volumes, software development costs are of little concern. As a result, the marginal price only relates to the hardware including the computer, the image subsystem and the interface subsystem (screen), which are increasingly available at lower cost and higher quality. The need for specific hardware is often the critical element when it comes to price. For instance, high resolution or non-video standard solutions are expensive, although prices for common hardware for image capture and display are very reasonable.

1.3.5 Speed and memory

Practical applications have to consider speed. We believe that a latency of a few seconds is acceptable for the user. Of course, the continuous increase of computing power reduces response delays, but an optimised algorithm requiring several minutes should be rejected for the moment.

Memory consumption is to be kept low. In recent years, the hardware requirements considered speed and memory, because memory was expensive and swapping slowed down the process. The size and speed of nowadays memories have increased and will certainly continue to do so. Except for the storage of large databases, available memory sizes are now very comfortable.

1.3.6 Illumination

The human eye sees a very large range of light intensities. The human brain is accustomed to dealing with noise and intensity changes in recognition tasks. Thanks to a general model, it can also integrate illumination changes caused by reflections.

On the contrary, cameras have a limited range and resolution in intensities. They are more subject to noise and saturation. Although profile based systems using the contour are little sensitive to illumination (assuming the contrast with the background is sufficient), frontal approaches have to consider illumination changes. Control of the lighting conditions (possibly with artificial lighting) eases software development and generally improves performance.

1.3.7 Pose

Translation, rotation and scaling of the face barely affect the human ability for face recognition. Automatic frontal face recognition programs do not tackle this pose problem easily, especially for Left/Right or Up/Down (depth) rotations. Translation and scale are generally handled by easy but additional processing. The cooperation of the individual and, to a lesser extent, the proper installation of the camera help normalise the acquisition relatively to pose.

1.3.8 Expression

Expression gives the face its own way of communication and is so important in social interaction that it possesses its own research area [Ekman78, Salzen86, Donato99]. In face recognition, it is seen as a disturbing effect to be minimised through user cooperation. However, a dynamic analysis (of a smile for instance) could reveal much of the identity, although face expressions would have to be produced (on demand) by the user.

1.3.9 Resolution

This criterion influences the image quality in terms of spatial accuracy.

As presented in section 1.2, a holistic approach only needs a coarse resolution. The gain of a fine resolution concerns feature analysis.

A fine image grain brings more facial details although some might be irrelevant. But a more detailed image consumes more memory and requires more computation for analysis.

The resolution offered by conventional hardware fulfils image quality needs. However, this supposes a correct installation of the camera and a minimum cooperation of the user to ensure that the facial region lies within the image and has enough pixels.

1.3.10 User cooperation

Face recognition systems require a certain degree of cooperation from the user to achieve their optimal performances. Cooperation typically concerns pose (distance, centring and rotation - see subsection 1.3.7), expression (usually required to be neutral - see subsection 1.3.8) and decoration (concerning scarf, hat or glasses).

1.4 Some other biometric solutions

About a ten of biometric features have already led to commercialised products to identify or perform verification of people: face, fingerprint, hand geometry, hand vein, iris, retinal scan, signature, voice, keyboard typing, gait, ... By definition, they use a physiological or behavioural characteristic of the person. To be useful for verification, they must be universal (shared by every person), unique, permanent (with time), collectable (leading to quantitative measures). Specific applications must still consider their performance (including constraints and environmental influences), acceptability by the user and the possibility of fraudulent techniques. See Table 1 for a comparison of biometric features relatively to these criteria (from [Jain99]).

Highly secure systems such as the ones based on fingerprint, iris or retinal scan benefit from the uniqueness and permanence of the underlying biometric properties they use [Jain97]. Unfortunately, these systems are little accepted by the user since they require physical contact or specific action (non invasive solutions seem however possible [Wildes97]). On the contrary, face or voice recognition techniques are well accepted.

Biometrics	Universality	Uniqueness	Permanence	Collectability	Performance	Acceptability	Circumvention
Face	H	L	M	H	L	H	L
Fingerprint	M	H	H	M	H	M	H
Hand Geometry	M	M	M	H	M	M	M
Keystrokes	L	L	L	M	L	M	M
Hand Vein	M	M	M	M	M	M	H
Iris	H	H	H	M	H	L	H
Retinal Scan	H	H	M	L	H	L	H
Signature	L	L	L	H	L	H	L
Voice Print	M	L	L	M	L	H	L
Odour	H	H	H	L	L	M	L
DNA	H	H	H	L	H	L	L
Gait	M	L	L	H	L	H	M
Ear	M	M	H	M	M	H	M

Table 1: Comparison of biometrics technologies (L=Low; M=Medium; H=High)

1.5 Classical approaches: frontal or profile

1.5.1 Face recognition: history at the SIC

We first started experiments in the field of face recognition in 1989 [Perneel90], when we tried to verify the ability of an artificial neural network approach for classification. The network was a Kohonen map that processed raw face images (no pre-processing). Although successful for about ten people, the program revealed its weaknesses when we acquired more images a few days later, with different illumination and pose. Clearly, the global approach showed its limitations regarding pose and illumination changes and the acquisition of additional images with different poses or illuminations would have taken too much memory.

In 1991, we were asked to address the face recognition problem, by any practically proven approach. We naturally started with a frontal analysis, thinking more information was to be found there. According to the sensitivity of our previous global approach, we preferred to consider a feature extraction procedure. Image resolution was limited (typically 70x100 pixels for the head) so we tried to analyse the configuration of the facial components. The head was first localised horizontally from high transitions of the horizontal profile that was obtained from the average of vertical grey-level projections. Then the horizontal average of pixels in the vicinity of the head horizontal centre gave a vertical grey-level profile from which we could easily localise the vertical position of the hair, forehead, eyebrows, eyes, nostril, mouth and sometimes chin. Our features were derived from ratios of distances between these vertical references. Although fully automatic and successful, the set of features was limited and sensitive to rotations. Additional features could have been extracted, but with uncertainty regarding correct detection. This would complicate classification and probabilities would have to be estimated. Finally, we decided to consider another modality.

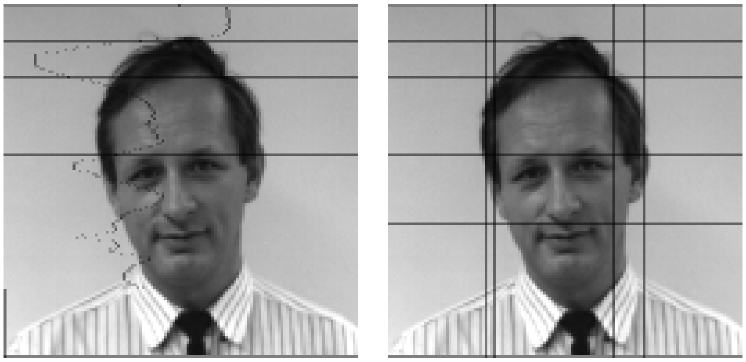


Figure 2: Vertical grey-level profile and horizontal/vertical references

In 1993, the profile appeared to be the appropriate solution to build up an automatic and fast solution. Firstly, the external profile contour delivers a lot of information based on rather rigid parts (forehead, nose, chin), which is little dependent on rotation. Cooperation is natural. Processing is easy and memory needs are low as the information is contained in a 1-dimensional curve. Lighting is nearly unconstrained, provided we have sufficient contrast for contour detection, which was simplified in our case by a uniform white background.

In a first development, curvature was used to localise reference points along the profile. The features consisted of distances between those references and curvature values. Later, we preferred to extract curvature (local) and angular (global) values along the profile using the nose and eye for rotation and scale normalisation.

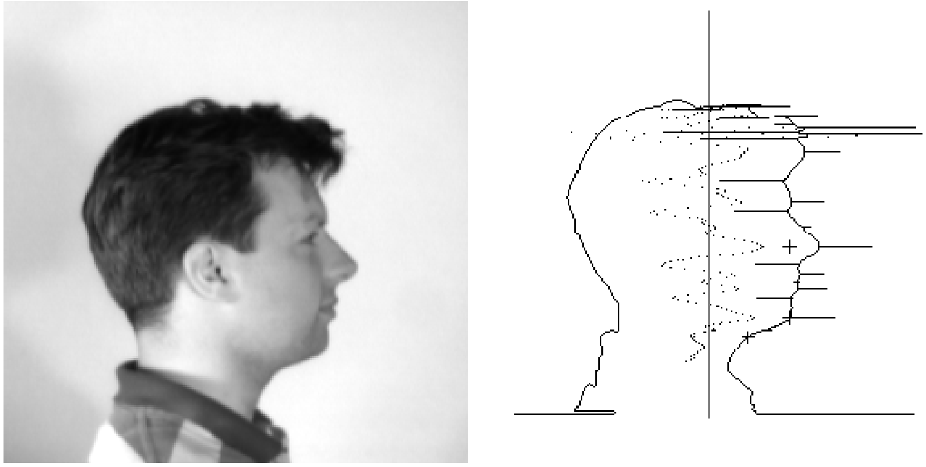
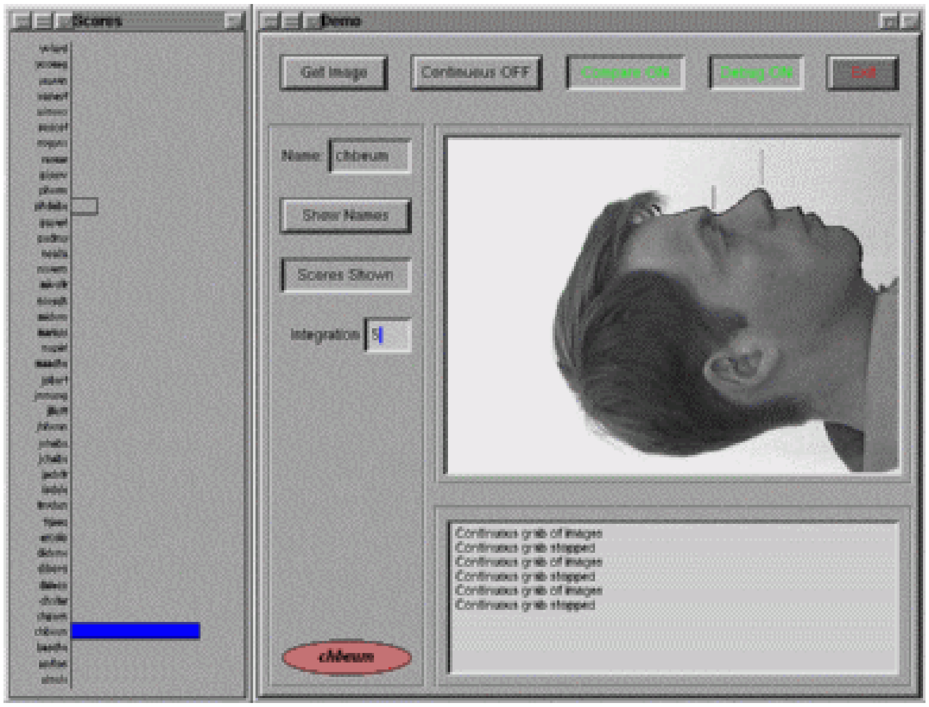


Figure 3: Profile image and extrema of curvature along the profile

In 1995, within the framework of the M2VTS (Multi-modal Verification Tools for Security Applications) project of the ACTS European programme, the profile solution served as a first prototype for the verification platform. The system is fast: acquisition, contour detection and feature extraction takes about 150 msec on a Pentium 200 MHz. Feature comparison is so fast it allows more than 5000 profile comparisons per second. For the database of our lab consisting of 41 persons, 270 profiles, comparison of one profile with all the database profile features takes about 50 msec. In that identification mode (comparison with the whole database), four profile images can be identified each second. The performance with sufficient reference images ("sufficient database representation") was measured several times, in different experiments, reaching an EER lower than 10%, with most errors stemming from bad eye or nose localisations. Acquiring profile images during a few seconds and asking the user to move his head if he is not recognised, the temporal fusion of subsequent identifications brought the EER inferior to 3%.



Name: chbeum

Integration 5

```
Continuous grab of images
Continuous grab stopped
Continuous grab of images
Continuous grab stopped
Continuous grab of images
Continuous grab stopped
```

Figure 4: Graphical interface of automatic profile identification system

Improvements could be brought to the current system by extracting more features (grey-level, for instance), combining features with adapted weights, using another decision rule, and/or optimising temporal fusion. But the main conclusion is that geometrical information brings new robust features (compared to grey-level) and lighting has little or no influence.

The main problems with the profile approach are:

- The dependency on the background to extract a correct profile;
- the important impact of glasses (proscribed in the above experiments);
- and the easiness with fooling the system with a picture or even a drawing.

A full facial surface will bring much information without those main problems.

1.5.2 Frontal and profile face analysis

a) A tri-modal classification

Computerised face recognition techniques can be separated into two classes: ***featural*** and ***holistic***. Featural approaches consider the extraction of characteristics relevant for the discrimination of the face. On the contrary, global methods compare whole images or a transformed version of them [Brunelli93, Kamel93, Lam98].

The same dichotomic distinction is made to explain the ***human brain process*** during the recognition of faces [Parkin86, Blanc-Garin86, Sergent86]. First, a face is categorised as a face by a holistic analysis considering lower frequencies and configuration. This rather quick process also delivers a feeling of familiarity with the face. Secondly, inner components of the face are analysed to make finer distinctions. This second process is slower, but is crucial for identification.

A ***feature analysis*** has the advantages of a deep control (each feature can be observed, normalised and weighted separately) and a high data reduction (the values to be stored for each reference are limited to a few numbers). It is commonly more susceptible to errors from occlusion due for example to glasses or hair. It requires more development details and a classification engine. ***Holistic*** approaches are conceptually easy, mainly consisting of a direct or indirect comparison. However, normalisation regarding viewpoint and illumination is a matter of concern and irrelevant data (like reflections) is hardly eliminated.

The two approaches seem to exhibit opposite advantages, a fact that has led to their ***combination***. For instance in Lam *et al.* [Lam98], facial points are first localised thanks to the eyes and mouth corners, the head outline, the nose and the eyebrows. These feature points are used for facial comparison. In case of face similarities relative to those features, eye, eyebrows and mouth windows are compared by correlation.

Trying to exploit advantages of the featural and holistic approaches has led to a third class of recognition techniques. In the general framework of pattern recognition, Mundy [Mundy91] presents the advantages of this in-between class that he calls ``***Model-based"*** recognition. The principle consists in adapting a model according to information extracted from the image. The model represents a holistic view of the object to be recognised and its adaptation is realised according to the extraction of image features. The model-based approach has the advantage of constraining the search for model matching from the inherent model topology. Beymer [Beymer94] also breaks down approaches into three major classes.

We propose summarising the face and profile analysis and recognition research relatively to the three classes of techniques: featural, model-based and holistic. Refer to Table 2.

	Featural	**Model-based**	**Holistic**
Frontal	Point localisation Curve localisation	MPEG-4, DLA, MBASIC	Eigenfaces isodensities
Profile	Fiducial marks	MBASIC	PF descriptors

Table 2: The distinction of frontal and profile face recognition approaches

b) Featural

Frontal feature approaches mainly concern the localisation of parts or points related to the eyes, mouth, nose and hairline or jawline, for inter-distance measurements, size estimation [Kaya72, Nixon85, Craw87, Kamel94, Brunelli95] or graph matching of feature points [Manjunath92, Wiskott96].

Feature analysis of head profiles mainly consists in the localisation of fiducial marks like the eyebrow, the nose tip, the lips and the chin [Galton10, Harmon78, Harmon81], as well as the extraction of features such as distances or curvature measures [Beumier95].

c) Model-based

Model-based methods for face analysis address the adaptation of a generic model of a face or its parts either by deformation of templates [Yuille89, HuangC92, Reinders92] or by localisation of facial features [Huang93, Akimoto93].
They are mainly used in image coding (MBASIC - Model Based Image Coding) for low bit-rate transmission [Aizawa95] and expression analysis [Choi91]. The MPEG-4 standard has contributed to an important research activity in the field [Pardas99, Ahn99].

In face recognition, *Elastic Graph Matching* [Wiskott96] or *Dynamic Link Architecture* [Kotro97] relate to a method implementing the model as a grid of nodes. Each node is described by local properties such as Gabor filter responses. Matching faces means comparing the properties of corresponding nodes.

d) Holistic

Global frontal approaches concern correlation in the spatial domain (possibly with several sub-templates as in [Baron81, Brunelli93], with a reduced set of pixels as in [Lucas97], or with a view-based template-based recogniser in [Beymer94]), or decomposition in principal components to enhance the separability of images in the feature (or face) space ('Karhunen-Loève expansion' in [Kirby90], 'eigenfaces' in [Turk90], 'Discriminant Analysis' in [Etemad97], and 'Independent component analysis' in [Bartlett98]). Some artificial neural network developments, especially the early ones, also considered images of the whole face [Stonham86, Kohonen89, Kerin90]. Let us finally mention a global technique of frontal views detecting and analysing isodensity lines [Nakamura91].

Global analysis of head profiles has been performed by Fourier descriptors extraction [Aibara92], circular correlation, moment invariants [Kaufman76], and profile distance estimation thanks to the Chamfer transform [Pigeon97].

e) Other classification aspects

As a second dimension for classification, we can distinguish between static and sequential analysis. A static approach considers one image for analysis. A sequential analysis considers a sequence of images based on which issues such as the best selection of images for comparison (concerning point of view for instance) or temporal analysis [Beumier97] can be tackled.

Still another way of classifying face recognition methods considers how image variation due to pose or expression changes is handled [Etemad97]. Several references (templates or feature sets) can be stored in the database with different facial orientation or expression. Matching can be realised through deformable templates, possibly with the help of a 3D model for other pose estimation. The variations can be incorporated in the feature extraction procedure.

Other aspects concerning face recognition such as the possible applications, the human ability and performance evaluation are presented in the survey of Chellappa *et al.* [Chellappa95].

1.6 3D analysis

1.6.1 Motivations

Full ***facial surface analysis*** is the natural continuation of the profile work. The use of the geometrical information, independent of lighting conditions, explains the success of the profile approach. Full 3D information will bring more clues, such as in the cheeks and forehead, independently of the pose adopted by the subject. A complete normalisation relative to rotation is achievable. Real distance measurements can be obtained.

The necessary task of ***face localisation*** is also easier with range information. Firstly, the face is normally the surface closer to the acquisition system, with a priori known dimensions. Secondly, the nose stands out of the face, providing a clear and precise reference at the centre of the face. Range information also facilitates the separation of the face from the background or from other faces.

But ***grey-level information*** is not to be underestimated. It perfectly complements geometrical data: in regions with low surface curvature, 3D acquisition is precise and in general little grey information is found. Regions characterised by large grey-level variations are often problematic for 3D acquisition, either due to local surface changes (nose, mouth, chin) or interference with the acquisition principle (eyebrow or beard for a structured light system). Grey-level clues can be used as another modality to improve 3D recognition. This will take advantage of 3D normalisation regarding pose (translation, rotation and scale). A reflectance model can also be used to normalise lighting conditions. In addition, grey-level can participate in geometrical matching by providing useful grey-level reference points to guide the matching procedure.

Human beings integrate 3D information when they recognise faces. This can be seen in the importance of shading in the cheek and chin regions of frontal images for instance. They also have to deal with pose variation, a typical 3D problem for which a solution could be to have several prototypical models or reference faces in memory. This has been shown in the brain of macaque and has been used as a recognition paradigm for pose invariance [Beymer94]. However, the 3D model created from shading or moving frontal face is certainly imprecise and incomplete, what is experienced when we first see the profile of somebody we are used to seeing from the front (e.g. on television). The conclusion is that a computerised vision system will take advantage of real 3D measures to perform quantitative measures where the human being is only capable of qualitative judgments.

1.6.2 Overview of related works

The number of publications related to 3D face analysis is rather limited. Giving an overview thus consists in presenting specific works rather then trying to summarize the directions of research.

One of the first works on 3D face recognition is the one of Lapresté *et Al.* [Lapreste88]. He extracts the *profile plane* from the symmetry of convex/concave transition curves around the nose. This symmetry plane is used to extract the profile from which characteristic points are localised. Distances between these characteristic points are the features for recognition. This approach needs very high quality range information as it uses curvature estimation.

In [Lee90], Lee and Milios segment 3D facial images into convex regions. Each of these are mapped onto the Gaussian sphere, using the surface normal at each point of the region. Matching is based on a similarity measure between two convex regions and on relational constraints between pairs of matched regions. The Extended Gaussian Image approach used here, although addressing easily the invariance to scale and orientation, suffers from the inadequate sampling rate in the areas of high curvature. Here also, range data sufficiently fine are needed.

In [Gordon91, Gordon92, Gordon92b], Gordon describes the 3D face as a set of nose, eye and head features in terms of points, lines and regions. These descriptors are obtained from constraints imposed on depth and curvature values. Simple features based on distance and curvature values are derived from the descriptors. Once again, the success of the approach was conditioned by the very high quality of the range data.

In their work about the labelling of facial components from range data, Yacoob and Davis [Yacoob94] analyse the facial surface in terms of curvature to highlight high convexity regions. A connected component analysis elicits a set of regions that must then satisfy size and position criteria (the boundary of the face is discarded). Facial component labelling is based on the relative positions of components.

Achermann [Achermann97] has applied two recognition procedures, initially developed for intensity images, to facial range data. These methods are based on "eigenfaces" and "hidden Markov models". In [Achermann00], the same author proposes the classification of human faces based on the Hausdorff distance, measuring the similarity of facial surfaces either as point sets or voxel arrays.

In [Chua00], 3D faces are compared by the similarity of point signature. The signature of a 3D point is a curve related to the 3D surface in the vicinity of the point. The same local shape description has been used in [Wang02].

Chang & Al. [Chang03] report on the use of Principal Component Analysis ("PCA", "eigenfaces") for the recognition of 2D and 3D faces of the same population. Similar recognition performances are obtained.

In [Romdhani02], 2D faces are identified by fitting a 3D morphable model containing shape and texture descriptors based on PCA. This approach inverts the face image formation process and is not 3D face recognition.

An important research effort in the field of 3D face analysis concerns the study of moving faces for teleconference applications. Many such works concentrate on the adaptation of a 3D facial model [Reinders92, Akimoto93, Huang93, Saulnier94, Aizawa95].

Another research field where 3D has been investigated is expression. Although most expression analysis approaches consider 2D facial images (see [Samal92] for a survey on the recognition and analysis of faces and facial expressions), Suzuki [Suzuki95] proposes a method for the analysis and synthesis of expressions using range images.

In light of the limited amount of publications in 3D Face recognition, a considerable development is still expected, and given the interest in our 3D database, many research projects are currently carried out.

By the time of our first developments in 1997, the reported methods relied on high quality facial surfaces. Our own experimentation with curvature estimation proscribed the related approaches due to the limited quality of facial surfaces with our first acquisition prototype. We preferred a geometrical comparison approach that also allowed for visual feedback during development.

Chapter 2 3D Acquisition

2.1 Introduction

Our ***objective*** is to capture the human facial surface as a 3-dimensional object for comparison purposes. This implies quality (density and precision) and practical constraints (timing, cost, bulkiness).

The system must be ***precise*** enough to make face comparison reliable. An excessive resolution (below 1 mm) is heavy and probably irrelevant relatively to the natural elasticity of some facial parts and the differences between people. Absolute 3D positioning is not intended, as relative measures are sufficient. Reliable 3D information allows acquired data to be compared with 3D data captured by other systems.

The system must be ***quick*** enough to capture faces within short poses to avoid motion blur. It must also deliver facial surfaces in a couple of seconds to limit user latency. Its price should be kept minimal to remain competitive with existing commercialised solutions.

Comfort in usage is a key requirement to ensure that the user will properly collaborate to be recognised. This implies a limited time for capture, but also a natural pose or easy interactivity with the system.

The next section presents an overview of existing techniques for range acquisition, as well as commercial products. *Coded structured light* appears as the appropriate technique for range acquisition and is the topic of section 2.3, where the motivation, principle and component description are given. Section 2.4 details the calibration procedure, which is necessary to achieve accurate and reliable measurements. The image processing tasks necessary to extract 3D positions are explained in section 2.5. Section 2.6 describes how surface texture is estimated. Section 2.7 presents the characteristics of the three prototypes, with the evolution of developments in Appendix F, and basic summary in Table 3. The databases collected with the prototypes are described in section 2.8. Section 2.9 concludes the chapter.

Prototype:	**Prototype A**	**Prototype B**	**Prototype C**
Nickname	“B&W thickness”	“Colour striping”	“Stripe with dots”
Date	1999	2002	2003
Full description	Appendix C	This chapter	This chapter

Table 3: Prototype definition

2.2 Range acquisition

As presented in Figure 5, a high number of ***range acquisition techniques*** exist today. For reasons regarding safety for the human face and cost, let us concentrate on wave reflection techniques.

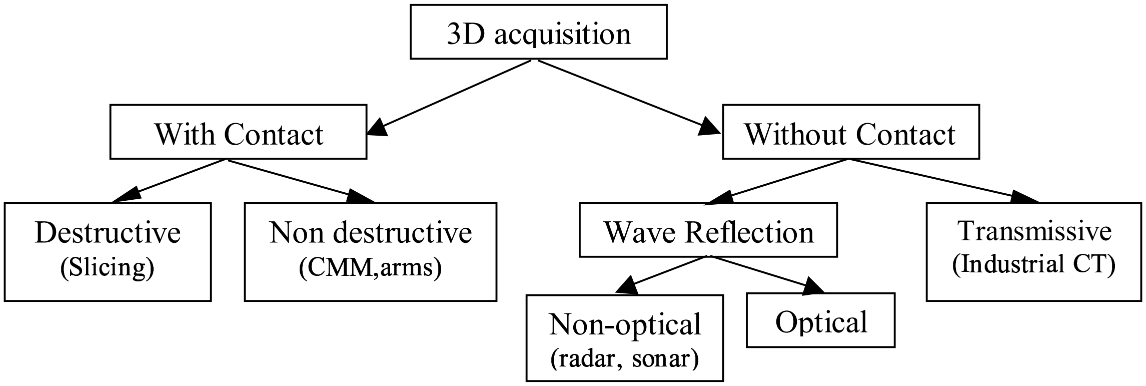


Figure 5: 3D acquisition techniques

[Jarvis83] presents a survey on range finding techniques based on wave reflection with a perspective on applicability and shortcomings in the context of computer vision. In a second paper [Jarvis93], he classifies the different methodologies based on wave reflection according to three criteria:

i) Passive versus Active

Passive methods use ambient light, while *active* range sensors project energy (light, ultrasonic, microwave) and are normally restricted to indoor applications.

ii) Image based versus Direct

Image based systems extract range information from image(s), while *direct* methods use other forms of physical properties (LASER time of flight, ultrasonic).

iii) Monocular versus Multiple view

In *monocular* systems, a single viewpoint is needed (range from focus, brightness or texture), while *multiple view* methods derive range information from disparities among images of corresponding areas.

Active	Direct	Image based
Triangulation	Simple Triangulation	Striped Lighting Grid coding Silhouette projection
Monocular	LASER time of flight Ultrasonic	Shape from shading

Passive	Direct	Image based
Triangulation		Stereo disparity Range from motion
Monocular		Texture gradient Range from focusing Range from attenuation

Table 4: Range sensing methodology classified according to [Jarvis93]

For automatic face recognition, we were looking for a range acquisition technique that is sufficiently precise, inexpensive and fast. Active solutions are attractive, as light projection is a way to reduce dependency on ambient conditions. Image based methods are normally low cost and compatible with the acquisition of grey-level or colour images for texture. Multiple-view solutions require additional hardware or acquisition time and rely on rather heavy algorithms to solve the correspondence problem. Based on this analysis, the ***structured light*** method appeared to be a good candidate. Further motivations for this choice are given in the next section.

Considering the high price of the 3D range sensors available on the market by the time of the first 3D experiments in 1995, we decided to develop ***our own 3D capture prototype*** A. We designed an opaque/transparent pattern suiting our low cost Black and White camera and digitiser. Later, we realised prototype B to take advantage of the high-quality digital cameras, and colour availability and to improve calibration. Texture recovery was introduced as design constraint to be able to capture 3D and colour in registration from the same image. Although evolving, the market did not yet offer low-cost 3D capture systems, so that we finally decided to realise a third prototype C based on black and white projection with a very interesting coding scheme which allows for good localisation and detection and which features texture capture possibilities.

Our current knowledge about ***commercialised 3D sensors*** is well summarised in the internet page of links collected by Cédric Laugerotte *http://student.ulb.ac.be/~claugero/PhD/3d_acquisition_systems.html* and based on *http://vr.isdale.com/3DScanners/3DScannerCompanies.html*.
Most of the systems are based on Laser projection and are rather expensive, providing a spatial resolution we do not really need. If we target face acquisition, only one system remains in this list ("Shape Snatcher" from Eyetronics). It captures 3D and texture in one shot. According to prices mentioned in *http://www.simple3d.com/faq.html* (7500 $), this system is low cost for a company, but expensive for research, considering that you only receive the slide and the software (camera and projector to be purchased independently).
Another interesting system was developed by the Turing Institute and sold as "C3D". It consists of two cameras that capture 3D surfaces by projecting small dots, simplifying the difficult "correspondence problem" of stereo. The system delivers 3D and texture.

More information about 3D capture, 3D measurement and 3D tracking devices is to be found in "Simple3D" at *http://wwwx.netheaven.com/~simple3d*.

2.3 Coded structured light

2.3.1 Motivations

Structured light acquisition systems use the projection of a known light pattern (like striped lighting or grid coding in Table 4) projected on the 3D scene and captured by a camera to recover 3D coordinates by triangulation. Compared to other 3D acquisition techniques as referred to in [Jarvis93], structured light features some ***interesting properties***:

1) In comparison with a classical camera, the ***additional cost is limited*** to a normal projector and its slide. Video projectors, although more expensive and cumbersome, can be used for dynamic projection (temporal series of patterns) or during the development cycle for rapid prototyping and testing.

2) Working with a normal camera allows to take advantage of the low cost and ***high speed*** of video hardware. A single image captured with the projected light pattern is sufficient to recover 3D information. This makes sequence capture for dynamic analysis (as needed by lip synchronisation with speech for instance), or integration of several captures for improved measurements (more precise and more complete) possible.

3) Switching the projected pattern on and off or measuring pixel values where there is no projected pattern is an economical way to acquire 3D geometry and ***texture*** information in close correspondence.

4) Active lighting renders the system less dependent on ambient ***light conditions***. As the projector direction is known from calibration, the surface reflectance characteristics (independent from lighting) can be estimated with a reflectance model.

The ***main disadvantage*** of most structured light systems is the necessity to solve the correspondence between an image position and a point of the projected pattern to apply triangulation. For those systems, the pattern must be coded ("**Coded** structured light") and image processing techniques must be developed to localise and decode the projected pattern in the image.

Other practical ***disadvantages*** of a structured light system are its relative bulkiness and its limited depth of focus due to the camera and projector lenses. Power consumption is also a matter of concern, as the projector lamp typically uses 150 W, which requires fan cooling. We finally used a flash lamp to solve these problems as the projector and camera diaphragms can be closed when the projected light is strong, increasing the depth of focus. Bulkiness has been addressed by keeping the minimum optical and mechanical elements of the projector, suppressing the fan cooling thanks to flash illumination.

A plethora of structured light systems have been implemented. Refer to a survey in [Battle98], a taxonomy based on the underlying assumptions about reflectance, space and time coherence in [Hall-Holt01], or an overview of scanners ordered by decreasing number of captured images [Zhang02]. As they principally differ in the type of patterns projected, some of them are reviewed in section 2.3.4.

2.3.2 Principle

The principle of structured light for 3D position estimation falls in the category of ***triangulation***. A scene point P is at the intersection of a line of sight of the camera and a line of projection of the projector. The third side of the triangle, the so-called **baseline** (D), joins the focal points of the projector and the camera. We can see the structured light system as a stereo head, one of whose two cameras is replaced by a projector. Unlike stereo, no correspondence problem (finding matching pairs of points in both images) has to be solved, but the projected pattern must be localised and, in most systems, decoded. This is advantageous for faces that contain few remarkable or precisely located regions for matching but large, rather smooth areas which do not affect the visibility of the projected pattern.

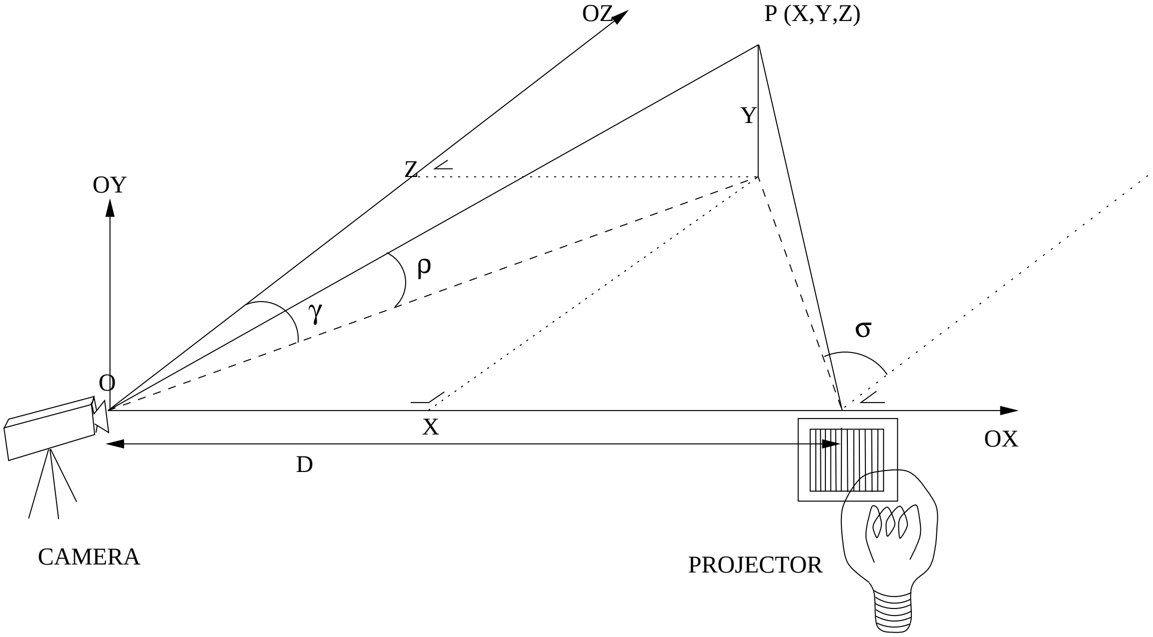


Figure 6: Structured light principle

Mathematically, the problem consists in determining the 3D coordinates X, Y and Z of a point P from its camera image coordinates (c,r) and projected element coordinates (s,t) from the projector. The solution can be a trigonometric or geometric formulation, as presented in section 2.4.2. It corresponds to the intersection of two lines of sight, one from the camera and one from the projector:

$$Mc * \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = a * \begin{pmatrix} c \\ r \\ 1 \end{pmatrix}$$

$$Mp * \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = b * \begin{pmatrix} s \\ t \\ 1 \end{pmatrix}$$

where a and b are two scale factors and Mc, Mp are camera and projector matrices obtained by calibration.
The knowledge of image position (c,r) and the projected element (s,t) leads to 4 equations for the three unknowns X, Y, Z. This is the reason why, in many implementations, projected elements are lines parallel to the camera vertical axis, so that only one projector coordinate 's' suffices. The solution is then the intersection of the camera line of sight with the projector plane of projected stripe. More generally, the fact that only one projector coordinate suffices is due to what is called the *epipolar constraint*: an image point (c,r) corresponds to a given point on a 3D line that emanates from a line (epipolar line) of the projector. One single coordinate in the projector plane identifies the projected point belonging to that epipolar line.

The main advantage of structured light is that a complete 3D scene surface can be captured with one image, from the localisation of (c,r) and stripe indices s in the same image. Moving objects can be captured with short camera exposure time.

To obtain precision and correctness of the captured data, the system has to be calibrated. This task, topic of section 2.4, has to be repeated if the camera or projector components or camera / projector arrangement is modified.

2.3.3 Light projection

There are mainly three types of projection that could be used for structured light: LASER projection, the traditional slide projector and the video projector.

a) LASER projection

LASER projection has the advantage of delivering an accurate and highly contrasted beam. However, this solution is more expensive and requires more specialised hardware [Ozeki86] to project the pattern, reasons possibly coupled with the common fear for the eyes (although not well founded with eye-safe LASERs). We have abandoned this possibility.

b) Slide projector

The ***slide projector*** is a simple and rather compact optical system, which projects light through the slide with projective perspective. This inexpensive solution is well suited to the realisation of a compact prototype, but requires the manufacturing of the slide.

The slide can be realised by photography on a ***slide film*** or by offprint on glass. Slide films are low cost, and dedicated machines exist to print a digital image on the film, with high accuracy in position, but apparently lower accuracy in chromaticity. Several slides were realised by companies in the photo business but the time response and quality suffered from the demand reduction in this industry (apparatus: Safir (Micheli), Genigraphics 8770 masterpiece).

The offprint of the pattern ***on glass*** is a more expensive and less classical solution that provides more mechanically rigid patterns, which are not affected by heat and are more contrasted. To our knowledge, classical offprints must be black patterns (no colour). Slides were realised according to the process used for masks of integrated circuits in microelectronics.

An advantage of slide projection is the possibility to project ***near infrared*** light, invisible to the human eye but detected by silicium CCD cameras (with no filter). This provides a discrete and comfortable solution that does not dazzle the individual but which is only applicable to black and white slide designs.

c) Video projection

Video projection is the de facto solution for prototyping and dynamic projection. The pattern can be changed at will to test or refine pattern designs with no additional cost or time latency. Unfortunately, video projectors are very expensive and cumbersome. They also suffer from a small depth of focus (typically 10 cm at 1 meter) at working distances we considered. But they are emerging on the market for home cinema at significantly reduced cost.

To sum up, video projection was used for slide pattern design, test and acquisition. Slides were realised to build up compact and inexpensive prototypes.

2.3.4 Projected patterns

a) The triple role for projected patterns

In the scheme presented in Figure 6, the projected pattern must fulfil two roles: point ***localisation*** and pattern item ***labelling***. These two roles allow for the 3D localisation of an illuminated point thanks to triangulation. We further constrained the slide design to include the possibility to recover fair or good scene ***texture*** (intensity and/or colour), finally providing a system capable of capturing the geometric and radiometric (reflectance) properties of the surface in a single image.

This section first presents a ***review*** of the structured light techniques described in the literature, concentrating on the differences in the projected pattern, according to its above-mentioned triple role. Then the most interesting of our ***original designs*** are described. Three of them are the patterns retained for the three realised prototypes.

b) Review

Many structured light systems have been described in the literature since the mid 80's [Jalkio85, Sato86]. A survey is presented in [Battle98]. They mainly differ in the type of patterns projected. As suggested in [Hall-Holt01], structured light methods can be characterised by the underlying assumptions made about the reflectance and the coherence in space and time of the scene.

The ***reflectance assumption*** supposes that the scene does not modify the colours so that the projected pattern can use colours [Davies96, Boyer87]. A range finder using monochromatic light is less sensitive to spectral modifications. Those systems are usually adequate for colour neutral scenes.

The ***spatial coherence assumption*** assumes that projector locations and camera pixels follow the same ordering, allowing for a labelling technique based on the distribution into several neighbouring elements of the pattern. The requirement of a local coherence in the labelling [Boyer87, Vuylsteke90, Beumier03a] reduces the coherence constraint. In the case of global coherence [Proesmans97b], only a single visible and connected surface can be acquired as discontinuities preclude labelling.

The ***temporal coherence assumption*** is required by systems that apply a sequence of patterns to a scene that is supposed to be static [Sato86, Morano98, Hall-Holt01, Rocchini01]. In [Bitner76], projected dots are switched on and off over time. Several variations have been realised such as black and white [Gärtner96], grey-level [Horn99] or colour [Caspi96] stripes. [Zhang02] presents some structured light systems along the continuum between multiple and single pattern projection.

Based on those assumptions, we rejected the *temporal coherence* which we found too restrictive for face acquisition in hardware design. Even for collaborating people, one second without facial movement is not natural.
The *reflectance assumption* is generally met with most faces as the colour content is rather poor. However, colours such as cyan and yellow may be reflected weakly in comparison with others. Black and white projection always brings more contrast for better detection and localisation.
The *spatial coherence assumption* holds in most of the facial area, thanks to the smoothness and convexity of most of the facial surface.

Let us now review systems according to the triple role of the projected pattern. Methods based on sequential pattern projection are not considered here.

b1) Localisation

Projected elements for point localisation are usually lines, line crossings, patches and corners. All these elements can be precisely detected thanks to the number of points involved (along a line, line intersection, points of the patch, crossing at corners).
Many techniques use parallel stripes (line of a given width) as basic element. This offers a continuous 3D description along the stripe and requires a 1-Dimensional labelling [Zhang02].

b2) Labelling

Many techniques are available to identify the projected elements. However, they should not disturb localisation or decrease precision. Due to the important number of elements (typically 100 or 200 along a direction), the identity is usually distributed either spatially among several neighbouring elements, or temporally in several images. Each piece of identity is usually coded in a colour [Boyer87, Carter90] or according to intensity [Carrihill85], or size such as line thickness [Beumier99b]. In [Maruyama93], the stripes (slits) are cut randomly to offer vertical segments of variable size to solve the correspondence problem.

With colour coding, colour neutrality of the scene is generally assumed [Boyer87, Morano98, Salvi98]. This constraint was relaxed in [Caspi98], where colour labelling is adapted to the scene to achieve robust and discriminative coding. Unfortunately, to achieve this, one ambient and one uniform white illumination images are acquired, which require the scene to be static between the two acquisitions. In the rainbow range finder [Tajima90], colour neutrality assumption of the scene is also relaxed by using a rainbow of monochromatic light projection, but changes of wavelength by reflection introduce noise. In [Forster01], the prototype for the European IST project HISCORE “High Speed 3D and Colour Interface to the Real World”, consists of two approaches, depending on the scene colour content: if the projected colours are not modified too much, colour gradients are sufficient for stripe labelling, otherwise a second image with white illumination is captured to compensate for the colour present in the scene.

We would like to refer to the original work of Carrihill [Carrihill85] who labels points by the intensity ratio between two images: one captured with spatially monotonic illumination and the other with constant illumination. The intensity ratio allows for range estimation at the cost of two image captures, requiring a static scene.

Another interesting solution is the development of [Proesmans97] where, under the assumption of orthographic projection, coding is not needed. It requires, however, a complete detection of successive projected lines, targeting the acquisition of a connected surface.

b3) Texture

The simplest way to capture intensity or colour information consists in acquiring a second image of the scene, preferably from the same camera and point of view. This supposes that no motion occurred since the 3D capture [Carrihill85, Forster01: approach 2].
When the texture quality is not critical, the image captured for 3D can be the basis of texture measures, if the projected pattern is compensated for in the image [Forster01: approach1]. If the desired texture resolution is not high, average texture values can be sufficient. Otherwise texture measure between projected pattern elements can be used, possibly interpolating or correcting values affected by pattern projection [Proesmans96]. Value correction is particularly interesting with colour pattern projection for which the R, G, B image fields are not affected all together [Beumier03]. In the specific case of face capture, as the colour content is poor, one R, G, B field can be reserved for texture while the others serve localisation and labelling (subsection c3) below).

c) Original designs

In this section, we describe the most interesting patterns we designed and we discuss their ability to fulfil the triple role.

The literature and our first experiments gave us confidence that parallel lines represent a very good projection pattern. Its continuity in one direction allows for optimal point density, detection continuity and subpixel localisation. In this presentation, a ***stripe*** is a vertical line of a given width. It has a colour, possibly only black or white and may be interrupted by colour changes. In the slide descriptions that follow, black and white are synonyms for opaque and transparent, respectively.

c1) Black and white thickness (Prototype A)

Our first development considered black stripes of varying width, either thick or thin, on a white background.

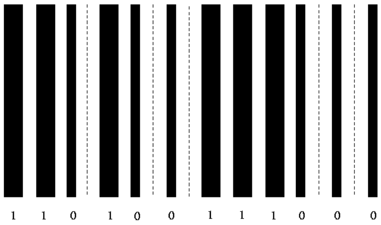


Figure 7: Part of the 'B & W thickness' slide

The left edge of black stripes is used for ***localisation***. These edges are regularly spaced as a result of the constant horizontal spacing.

Labelling is performed in the thickness of consecutive stripes. Each (vertical) stripe uses a horizontal space of w pixels: the first 1 (thin) or 2 (thick) thirds are black and the rest is white (see Figure 7). This way, the ratio of black and white pixels of a stripe is theoretically either ½ or 2, affording a good separation between the thin and thick classes. Theoretically, 7 or 8 consecutive stripes are necessary to uniquely identify 134 or 263 stripes. In practice, only 10 or 12 successive stripes can ensure correct labelling given errors of thickness estimation or stripe detection.

The ***label sequence*** was obtained by an original and simple algorithm described in Appendix B. The bit sequence is constructed by adding one bit at a time, ensuring that the n-bit (n=7 or 8 here) word is not used yet. When both words (ending with 0 or 1) have already been used, a sequence is stopped. A complete sequence using 2^n words is possible by inserting the partial sequences. It is 2^n+n-1 bit long. Several solutions exist, depending on the order of sequence insertion and on the starting word. The complete sequence is cyclical.

Texture was not addressed and is poor as half the space is projected with dark stripes hiding most of the texture.

We used glass etching for the ***realisation*** of the slide, as performed by the Katholieke Universiteit Leuven (Belgium) in the microelectronic department. We designed one slide with 134 stripes (stripe width is either 90 or 180 µm) and one slide with 200 stripes (stripe width is either 60 or 120 µm).

c2) Colour C6WC6 (Prototype B)

With the arrival of high quality digital colour cameras at a reasonable price, we tried to address the acquisition of texture, and reduce the number of stripes for labelling. Keeping the good localisation ability of vertical stripes, we distributed labelling in the colour of successive stripes separated by white spaces (see Figure 8).

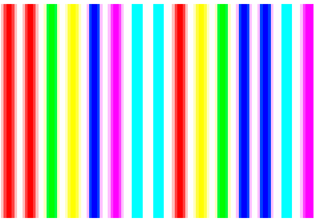


Figure 8: Part of the 'C6WC6' slide

Adopting equal horizontal size for colour stripes and white spaces, horizontal edges at colour stripe borders offer ***localisation*** cues, as much as twice the number of colour stripes, and regularly distributed in the horizontal direction. In comparison with the previous slide design, this implies three times more extracted points for the same size of the smallest stripe width.

Colour code helps in designing more compact ***labelling***. Among the 8 possibilities of minimal or maximal intensity for the R, G and B colour fields, we rejected Black which has a badly defined colour hue, that largely interferes with texture, and White, used as stripe separator. We thus kept 6 colours (Red, Yellow, Green, Cyan, Blue, Magenta) well distributed and separated in the hue space of colours (Figure 9).

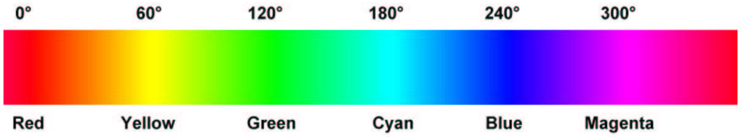


Figure 9: hue values and related colours

The complete ***label sequence*** is obtained as the concatenation of three sequences of unique pairs of colours among the 6 possibilities. A first sequence of unique pairs (6x6 = 36 pairs long) is obtained as described in Appendix B. The next two sequences consist of a transformation of the first one by a rotation in the R, G, B fields of each colour. Each colour pair occurs three times in the complete sequence but only a few colour triplets occur more than once. The whole sequence is 3 x 36 = 108 stripe long, providing 216 borders used for localisation.

The colour design of the slide also provided the opportunity to address ***texture*** capture from the same image. Half of the slide is white, while colour stripes have at least one colour field (R, G or B) with full projection intensity. Intensity contrast from surface texture should remain present in the image, even where colour is projected. On the contrary, colour items of the surface could be modified by the projected colour stripes. However, a face contains only a few colour items (eyes and lips) and hair surface capture is not intended.

The slide ***realisation*** was handled by ‘Business Slides, Brussels’ which utilised the “Genigraphics 8770 masterpiece” system to scan a Fuji colour slide film. The film was developed with standard settings and the slide was stuck on a thin plastic support for mechanical rigidity.

c3) RtexGlocBlab

This slide has been designed to give each field R, G, B one of the three roles.

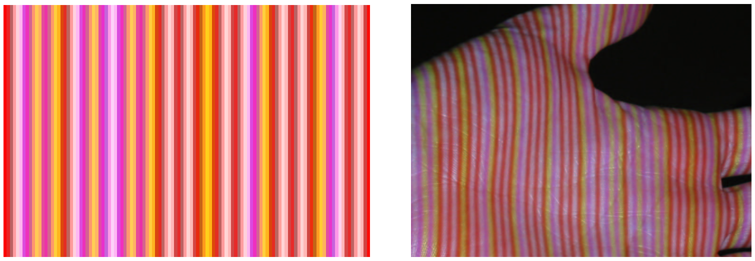


Figure 10: Part of ‘RtexGlocBlab’ slide

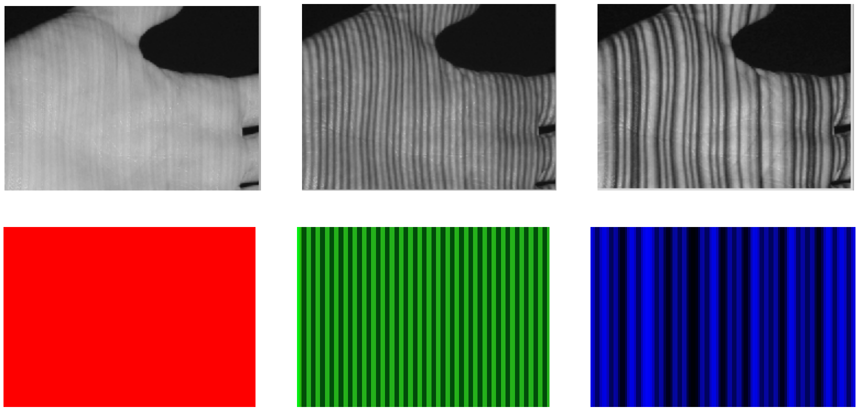


Figure 11: Red, Green and Blue fields of Figure 10

The image green field (G) is used for ***localisation***, for its higher resolution with the typical 1-CCD colour camera with BAYER filter (twice as many Green pixels than Red or Blue pixels). The green pattern is a periodic alternation of minimal and maximal intensities (G=0 and 255), constant on the stripe width.

The red field (R) is reserved for ***texture*** and has no pattern. This is a high intensity channel due to high reflectance of faces in red.

The remaining blue channel is used for ***labelling***, each stripe width coding one bit (B=0 or 255). The code density is not high, requiring theoretically 8 stripes (to identify one of 263 different stripes), but three times higher than slide c1) with the same minimum stripe width.

The slide has two major ***drawbacks***. First, it is rather specific to faces or skin, using red for texture, and generally produces poor results with strong coloured surfaces. Secondly, the so-called channel cross-talk (interference between R, G and B channels) affects the image quality in each field, especially with small stripe width, limiting the stripe density.

c4) Triangles

The interest of the slide design with triangles lies in the precise localisation of corners.

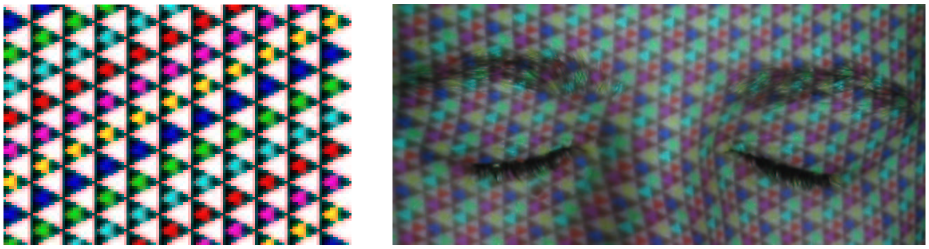


Figure 12: Part of 'Triangle slide' and projection on face

Localisation takes advantage of the three sides of the triangles to get a robust and precise corner detection. The distribution of triangle corners is evenly spread over the surface.
Labelling is coded in neighbouring colour triangles. (Figure 12 presents the idea but is not a real implementation of labelling).
Texture is based on the white and colour triangles to estimate the underlying intensity or colour information.

The major ***drawback*** of this slide design is the difficulty to realise a dense projection of triangles to get a high resolution of extracted points.

c5) Black stripes with colour dots

In order to maximise the stripe visibility of the projected pattern, we decided to use black and white with high contrast for ***localisation*** purposes (Figure 13).

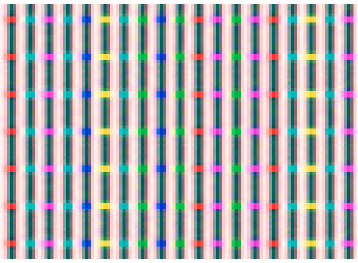


Figure 13 Slide with colour dots

The pattern consists of black stripes with colour dots for ***labelling***. Each stripe has constant colour dots distributed regularly along the stripe and small enough not to disturb stripe localisation. Vertical regularity allows robustness in dot detection and colour estimation.

The white space between stripes is the basis for ***texture*** measurement. RGB levels around the stripes are compensated to reach levels similar to those in the surrounding white space regions.

The major ***drawback*** of this slide is the sensitivity of detection and hue estimation of colour dots, especially with colour or dark texture.

c6) Black and white stripes with dots (Prototype C)

To circumvent the disadvantages of the previous slide design regarding colour dot detection and labelling, we designed another labelling method based on vertical dot positioning. The solution is very convenient as no colour is needed anymore, allowing for better contrast and slide offprint on glass. The solution is also not at all specific to faces and can be used for any colour surface.

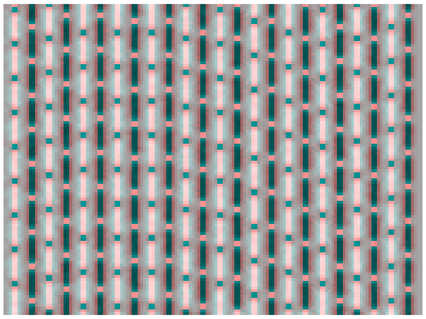


Figure 14: Part of the 'Black and White stripes with dots'

Localisation is achieved by alternative black and white vertical stripe detection. The periodicity and consensus in the three channels R, G, B ensure robust and precise detection. Stripes as thin as four or five pixels are well detected in images of faces due to the coherence in R, G, B fields (the colour specificity of the BAYER mask, Figure 15, has less influence).

The space between the stripes is filled with grey for texture. This also makes possible to keep the stripe width to a value optimal for visibility and localisation accuracy (around 4 or 5 pixels with our camera). Dots used for labelling are small and sparse. They do not greatly influence stripe detection and localisation.

In a structured light solution as presented in Figure 6 (section 2.3), a horizontal line (parallel to the baseline) is only scarcely deformed by depth. The vertical distance between horizontal lines should not vary much, especially when considering small areas for which depth and distortion variations are limited.

Labelling is performed in the vertical offset of dots in consecutive stripes. Each stripe is dotted with squares of opposite intensity for maximal visibility. Dots are repeated regularly in vertical direction with a fixed distance. This allows for filtering during dot detection and ensures a consistent offset along stripes.

The ***label sequence***, detailed in Appendix B, is based on a series of 6-class values where two consecutive values occur once. The series is 37 elements long. The values represent one of the 6 possible vertical offsets of dots between two consecutive stripes. Uniqueness is achieved when taking account of three successive stripes (two dot offsets).

The whole sequence consists of the triple repetition of the 6-class series. Ambiguity between the three repetitions will be resolved through the horizontal image position. As stripes are alternatively black and white, the series can be twice as long from repetition, leading to unique pairs of values when including the stripe colour (black or white). Triple repetition leads to approximately 3 x (2 x 36) = 216 elements for stripe labelling. Details are given in Appendix B.

The grey spaces between stripes provide the regions for ***texture*** estimation. Colour disturbances due to stripes are compensated for by the neighbouring grey regions (see subsection 2.6.3). Also, grey spaces induce smoother transitions between white and black, implying a more linear behaviour allowing for a better compensation. A slide pattern containing grey elements can be realised with black and white dots by dithering.

The slide is compact thanks to thin stripes, small dots and a compact labelling. Dots allow for labelling in the vertical direction. This could lead to 2-D localisation for better projector calibration, possibly compensating for projector radial distortion.

2.3.5 Camera

The camera conditions most of the parameters of the acquisition system. First, it is the entry point for acquisition from the PC, thanks to, for instance, a USB connector. Secondly, it limits, together with the projector, the field of view, the depth of field, the density and the precision of acquired points. Thirdly, it conditions, with the projected pattern, the general quality of the image (contrast, colour) and of the extracted texture.

For many years and until the end of the last century, typical hardware for image acquisition consisted in a 768x576 pixel Black & White camera connected by an analogue video cable to a digitisation board hosted by the computer. The total chain of acquisition suffered from limited pixel accuracy of the CCD, and Analogue / Digital conversion and signal synchronisation artefacts (jitter). We initially used a Panasonic WV-606BL Black & White camera connected to a Matrox acquisition board on the PCI bus.

In contrast, current ***digital cameras*** have much higher resolution, with colour and a simple digital connection for transfer to the computer. We benefit from the high quality of the all-digital. Resolution is typically 3 or 4 millions of pixels, however affected to either Red, Green or Blue, according to the BAYER mask. The only restriction for the time being with common cameras is the transfer time (several seconds), although solutions are already available (USB2, firewire).

G R G R G R G R G R G R G R
B G B G B G B G B G B G B G
G R G R G R G R G R G R G R
B G B G B G B G B G B G B G

Figure 15: BAYER mask for 1-CCD colour camera

Looking for a digital camera at a reasonable price, we finally decided on the ‘Canon G2’. It has a very good evaluation (in http://www.dpreview.com) for radiometric (colour) and spatial quality and can be driven by the computer (for the main camera selections and capture trigger). This last feature is necessary to reuse the settings of the last calibration because the camera automatically resets after power off, or after a timeout of inactivity. The highest image resolution is 2272x1704 of R, G or B pixels. An external flash can be connected to and triggered by the camera.

An important optical parameter of a camera is its focal length, which can be adapted either with variable lens or by lens replacement. The focal length and the image sensor size determine the field of view at a given distance of work. Short focal length lenses see a larger area at a given distance, implying less distance resolution per pixel.
The main limitation with common lenses is to work at short distances due to the difficulty to focus at short distance and to capture a large area and also because a short focal length lens usually suffers from greater distortion.

2.3.6 Geometry

The overall geometry of the acquisition system depends on the characteristics of the camera, the projector and their accommodation.

The ***field of view*** of 3D acquisition is the intersection of the field of view of the camera and the projector. Each 3D point must be seen by the camera and illuminated by the projector. The field of view is adapted by relative camera and projector placement and magnification and focus due to the lenses. A distance of work is first chosen (e.g. 1m), usually to fit the scene completely with available lenses (focal length).

The 3D ***depth of field*** is the common interval of sharp projection and image capture. Each lens has to be tuned ('focus') according to the desired distance of work. The volume of acquisition is the product of 3D field of view and depth of focus. Short distances of work encounter the difficulty to focus due to the limited depth of field.

The camera and projector must be placed at some distance ('***baseline***') to achieve depth precision. If the baseline is large, precision from triangulation is high but more parts are occluded, being either not illuminated or not seen by the camera. Short baselines imply lower precision and are limited by the minimal distance required to set the camera and projector aside. The baseline is chosen to solve the visibility and accuracy compromise.

With the objective of acquiring faces, we set the following geometric requirements:

Field of view	50x37cm
Depth of field	30 cm
Working distance	100 cm
Baseline	10 cm

Figure 16: Geometric requirements for face capture prototypes

They were established through long-standing experience with common and low-cost hardware (camera, projector, lenses) to obtain satisfying resolution. They are adapted to a normal seated position of a cooperative person. A seated position reduces position variations and height differences between people.

The field of view ratio comes from the usual image ratio of camera and projector (3/4 or 2/3). The baseline affects the desired depth resolution and visibility and is dependent on the camera resolution, distance of work and mechanical possibility to set the camera and projector close. We intend to keep it below 10 cm, for prototype compactness, larger 3D coverage (fewer hidden parts) and limited stripe deformation caused by depth for easier image processing.

2.4 Calibration

2.4.1 Introduction

Calibration is the determination of the system parameters in order to make the 3D measurement as accurate as possible. In a structured light acquisition system, these parameters concern the camera sensor and optics, the projected pattern and the projector optics, and the camera / projector arrangement.

The calibration approach consists in acquiring with the system images of a ***reference object*** of known characteristics. Reference points are detected in each image. The structured light principle allows deriving 3D coordinates of image points or 2D projections of 3D points in the image plane. An optimisation algorithm will adapt the parameters to minimise the error between coordinates of points from the reference object and coordinates of the same points measured by the acquisition system.

The success and quality of the approach depend on the:

- Quality of the model which is used to approximate the acquisition system;
- Quality of the reference object, in terms of the design (type of marker, density) and of the physical realisation (precision, material);
- Detection and localisation of reference points in the image;
- Minimum search algorithm looking for the best set of parameters.

The following sections develop each of these criteria in depth. Appendix E describes a graphical interface that helps realise the last two criteria.

2.4.2 Model

We developed two formulations to derive 3D positions from image measurements. The initial one, based on angular parameterisation, and derived from [Boyer87] was used in our prototypes A and B. The second formulation, used for prototype C and based on projective geometry, fits the parameter minimum search (section 2.4.5c)) developed by Lavest [Lavest98], which appeared to be the state-of-the-art solution for robustness against the calibration object.

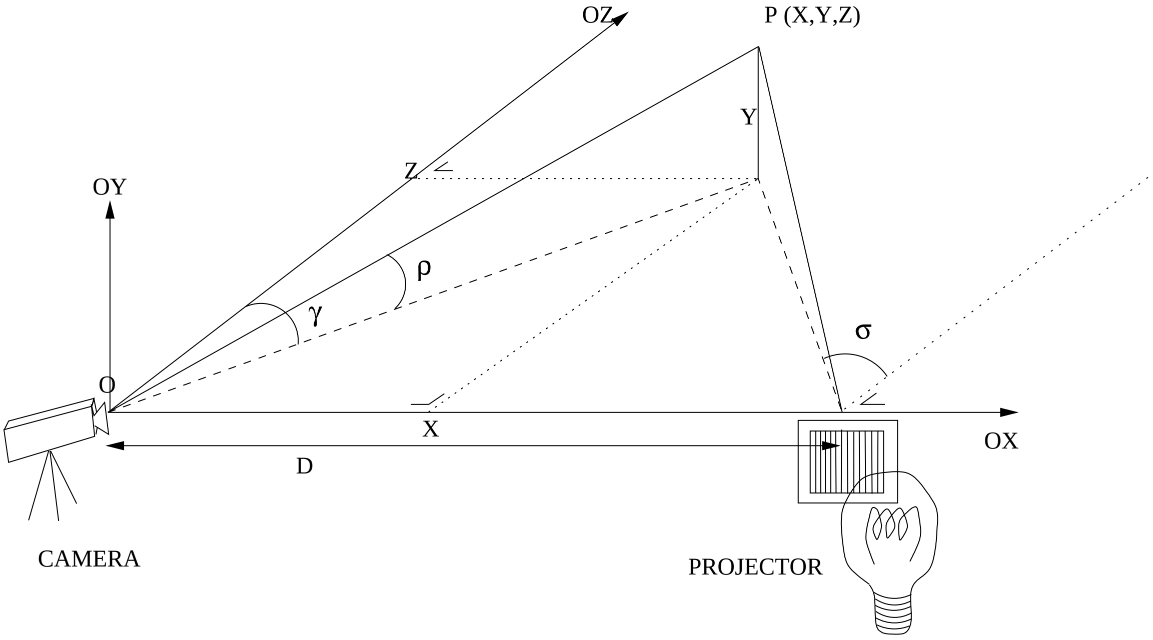


Figure 17: Axis system linked to camera and projector for angular parameterisation in structured light.

a) Angular parameterisation

We follow the modelling used in [Boyer87], depicted in Figure 17, and assuming the camera and projector optical axis are coplanar and perpendicular to the camera CCD columns and to the stripes. We can express the 3D position (X,Y,Z) linked to the prototype axis system relative to the camera and projector parameters. Let 'O' (0,0,0), the centre of the prototype axis system, be the focal point of the camera model; let OX be the direction joining 'O' to the projector focal point; let OY be oriented perpendicularly to OX and in the camera image plane; and OZ perpendicular to OX and OY, we have:

$$\left.\begin{aligned} Z &= D / (\tan(\gamma) + \tan(\sigma)) \\ X &= Z * \tan(\gamma) \\ Y &= Z * \tan(\rho) / \cos(\gamma) \end{aligned}\right\} \textbf{(Equations 1)}$$

With

$$\gamma = \gamma_0 + \mathrm{atan}(\, x / f_{cx} \,)$$
$$\rho = \rho_0 + \mathrm{atan}(\, y / f_{cy} \,)$$
$$\sigma = \sigma_0 + \mathrm{atan}(\, s / f_{px} \,)$$

and

$$x = c - c_0$$
$$y = r - r_0$$
$$s = stripe - s_0$$

Where

D is the distance between the camera and the projector focal point,
ρ, γ, σ are angles as depicted on Figure 17,
ρ_0, γ_0 identify the direction of the camera optical axis,
σ_0 identifies the horizontal direction of the projector optical axis,
f_{cx} (f_{cy}) is the "horizontal (vertical) camera focal length" in pixel unit,
f_{px} is the "horizontal projector focal length" in stripe unit,
c, r are horizontal and vertical coordinates relative to the upper left image corner,
x, y are image coordinates relative to the image centre (c_0, r_0),
stripe is the stripe index relative to the leftmost stripe of the pattern,
and s is the stripe index relative to the slide centre (s_0).

The "horizontal camera focal length" refers to the focal length of the camera model divided by the pixel size in the horizontal direction, leading to a measure in pixel unit. As the pixel size may be different in horizontal and vertical directions, two 'focal length' are introduced: f_{cx} and f_{cy}. The same naming convention has been used for f_{px}, expressed in stripe unit.

The 10 parameters (D, ρ_0, γ_0, σ_0, f_{cx}, f_{cy}, f_{px}, c_0, r_0, s_0) present in the model are given initial values and are obtained by an optimisation procedure, which is the topic of section 2.4.5.

We refined the model to take optical distortion and rotation into account. The values of x, y, and s are modified as follows. First, the possible rotation of a real implementation between the camera vertical axis and stripe projection orientation is accounted for by angle φ :

$$x' = (c - c_0) * \cos(\varphi) + (r - r_0) * \sin(\varphi)$$
$$y' = -(r - r_0) * \sin(\varphi) + (c - c_0) * \cos(\varphi)$$

Then the camera distortion is compensated for by parameter K_{dc}, following the hypothesis of an isotropic (radial) deformation:

$$x'' = x' * (1 + K_{dc} * dist^2)$$
$$y'' = y' * (1 + K_{dc} * dist^2)$$

where *dist* is the distance to the image centre:

$$dist^2 = (c - c_0)^2 + (r - r_0)^2$$

Projection distortion was considered, but not retained, as only one projection dimension is measured (stripe index) and the main distortion effect is bi-dimensional (radial).
The two additional parameters (φ, K_{dc}) raise the number of parameters to 12. These refinements were included in the second prototype B.

b) Projective geometry approach

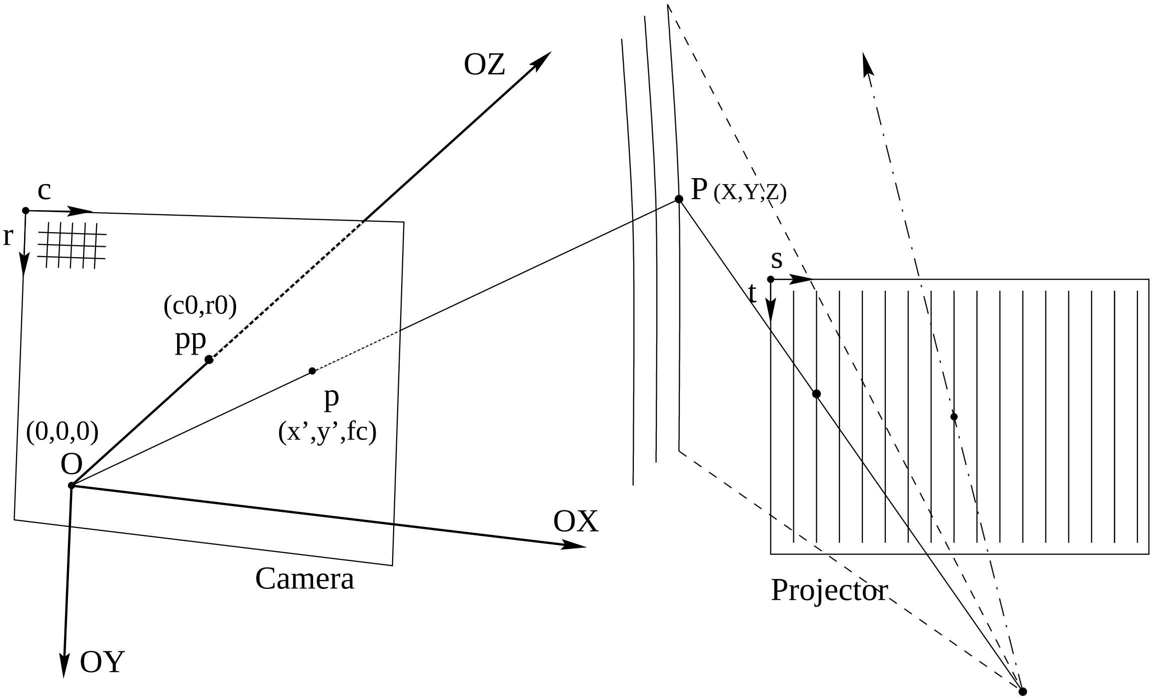


Figure 18: Axis system linked to camera and projector for the projective geometry solution in structured light.

There is a simpler way to find the 3D coordinates (X,Y,Z) of a point P from the image position (*x,y*) and the identity of the pattern element (*s*). P is the intersection of the line of sight of the camera, defined by the focal point O and the image point p, and the line of projection, passing through the projector focal point and pattern element. In the case of stripe projection, we have to find the geometrical intersection of a line (of sight) with a plane (of projected light). This is optimally computed when the 3D axis system is centred at the camera focal point 'O' with the X and Y axes parallel respectively to the pixel column (*c*) and row (*r*) axis of the image plane, while Z is given by the direction of the optical axis, perpendicular to the image plane.

An image point (*c,r*) gives the image coordinates (*x',y'*) relatively to the principal point (c_0, r_0) and according to a distortion compensation function DistortionFct

x' = DistortionFct(c-c_0, r-r_0)

y' = DistortionFct(c-c_0, r-r_0)

specifying directly the vector Op= (x',y',f_c) where f_c is the focal length of the camera, i.e. the distance separating the image plane from O (0,0,0).

The planes of the projected stripes have been obtained during calibration, in the convenient form:

$$a_s * X + b_s * Y + c_s * Z + d_s = 0$$

where the normalised vector (a_s,b_s,c_s) gives the direction normal to the plane corresponding to s so that d_s, the independent term, gives the (signed) distance of (X,Y,Z) to that plane. From the alignment of O, p and P (in the plane), and proportional rule, the 3D point P corresponding to (x',y') and s (stripe number) is obtained thanks to:

$(X,Y,Z) \equiv OP = k * Op$

with $k = d(O,P) / d(O,p)$ and

$d(O,P) = d$,

as P belongs to plane and O = (0,0,0)

$d(O,p) = d(O,P) - d(p,P)$

$= d_s - (a_s*x' + b_s*y' + c_s*f_c + d_s) = -a_s*x' - b_s*y' - c_s*f_c$

with d(P,Q) denoting the distance between P and Q.

We finally arrive at the following equations:

$$\left.\begin{aligned} X &= d_s*x' / (-a_s*x' - b_s*y' - c_s*f_c) \\ Y &= d_s*y' / (-a_s*x' - b_s*y' - c_s*f_c) \\ Z &= d_s*fc / (-a_s*x' - b_s*y' - c_s*f_c) \end{aligned}\right\} \quad \textbf{(Equations 2)}$$

In the case of a 2D pattern projection (instead of 1D striping), a second variable (‘t’) exists in the image plane of projection, and OP (X,Y,Z) is found as the intersection of two lines.

The parameters present in this model refer to the camera intrinsic values (c_0, r_0, f_c, DistortionFct) and the plane coefficients integrating projector intrinsic values and system parameters. Those values are determined by the parameter optimisation procedure detailed in 2.4.5c).

2.4.3 Reference object

The objective of the reference object is to provide reference points of known geometry so that system parameters can be accurately determined.

a) Design constraints

Several constraints have to be observed to design a correct reference object:

- the object and the pattern must be precise enough to satisfy calibration accuracy expectations;
- the density of reference points must be high enough to deliver sufficient data for precise calibration but must be limited to keep a good localisation of the points;
- reference patterns must lead to correct and precise localisation;
- in the case of structured light, the patterns of the reference object should not interfere with the projected light pattern.

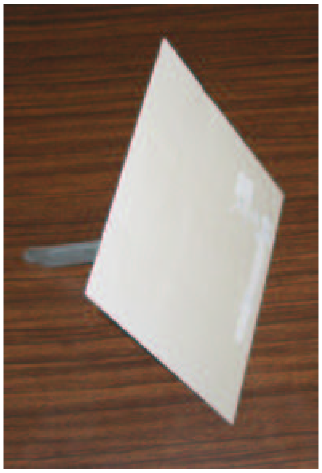
b) Review

Camera calibration for 3D capture considers either planar (2D) or volumetric (3D) objects. The advantage of volumetric objects is to allow calibration in one presentation [Proesmans96], but they are more difficult to realise with precision. Their precision is also more difficult to assess or requires more sophisticated hardware. On the other hand, high-resolution 2D objects are realised very easily.

The ***pattern*** to use for reference must be precise and visible. Visibility is not a concern because the number of reference points (typically 100) is relatively low in comparison with image resolution. Localisation precision is critical, because it conditions the overall precision. However, the number of points plays in favour of precision increase thanks to least mean squares techniques, for instance. Two types of pattern are often used: crossing lines [Guisser92] and circles [Lavest98, Redert02]. Line intersections give precise reference points, if lines are correctly detected and localised. Sufficiently large circles ensure good detection and provide high precision thanks to the numerous points if an elliptic model is used to fit the circle under affine transformation due to point of view changes.

c) Selected designs

Based on the considerations presented above, we decided to realise planar objects.



Prototype A considered as reference object a planar white square (Figure 19), 15 cm large, whose reference points are its 4 corners. The very small number of reference points per image was clearly the weak point of the calibration procedure, but results were sufficiently good to concentrate on other aspects of the project (face acquisition and recognition).

Figure 19: Calibration object (Prototype A)

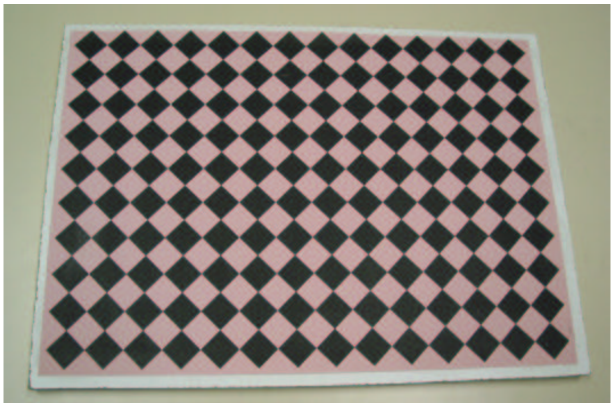


Figure 20: Calibration object (Prototype B)

A considerable effort during Prototype B development was devoted to calibration. The calibration object consists of a planar chessboard (Figure

20), printed in A3 (42 x 29.7 cm) and stuck on a rigid plate. It nearly covers the whole field of view and provides one reference point each 3 cm in both directions. These reference points are the square corners separating two white squares horizontally. The chessboard offers good contrast for edge detection and the white squares were finally painted in pink (close to skin colour) to offer an intensity range close to facial images with the same camera settings. To avoid interferences between the projected stripes (nearly vertical) and the reference squares, the chessboard was printed diagonally. As presented in section 2.4.4, reference points are the intersection of detected square edges.

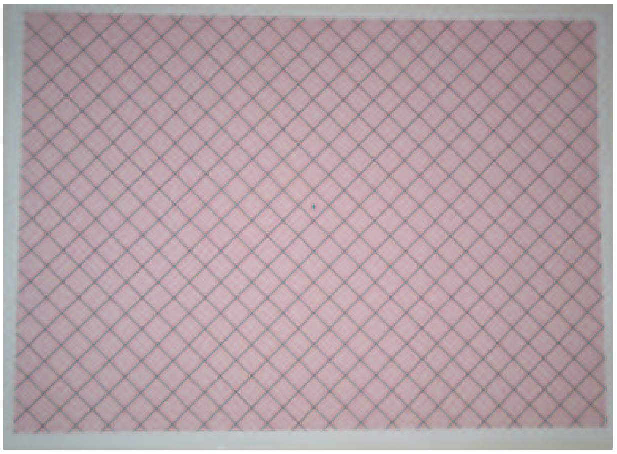


Figure 21: Calibration object (Prototype C)

For increased visibility of the projected stripes and building on Prototype B developments, the Prototype C reference object consists of a grid of dark diagonal lines on a pink background. The A2 size (59.4 x 42 cm) ensures complete coverage of the field of view, even for slanted or tilted presentations. The planarity constraint of the underlying plate has been relaxed by a robust estimation of camera intrinsic parameters (see section 2.4.5). For consistent labelling of reference points among different images of the object, the central square is marked with a small vertical segment.

2.4.4 Reference point extraction

The reference point extraction necessitates image analysis tasks to deliver corner image positions with stripe labels. Symbolic corner coordinates follow the regular corner arrangement of the reference object. Image measurements and symbolic coordinates are used by the parameter optimisation algorithm to determine model parameters.

Prototype A is a simple implementation based on the detection of the four corners of the presented square object (Figure 19). Stripe labelling is globally solved on the square to derive decimal values at the four corners. Due to the limited accuracy of using only four points, no further details are given here. See Appendix C.

To simplify processing of prototypes B and C, two reasonable assumptions were made. First, stripes are projected nearly vertically on the image, following the adopted arrangement of our structured light system. Secondly, square edges (chessboard of Prototype B) or grid lines (Prototype C) of the reference object are diagonally oriented in the captured images. This rule, easily observed by the user during capture, leads to clear square edge detection thanks to a vertical gradient computation which is not influenced by the nearly vertical stripes. In what follows, the term 'corner' will be used indistinctly for a square corner of the chessboard of prototype B or for the intersection of grid lines of the grid reference object of prototype C.

a) Corner localisation

Around an initial position, we look for vertical grey intensity changes. In the case of prototype B, square edges are localised as maximal vertical gradients. For prototype C, vertical local minima, normally corresponding to grid lines, are searched for. The detected points rarely arise from projected stripes because these are nearly vertical in the image.

Each detected point is then classified as belonging to a rising edge/line or falling edge/line. A line is fitted to each category of points. The intersection of the two lines gives a precise estimation of the corner position with subpixel accuracy.

A first corner 'C1' is looked for near the image centre. From its precise position obtained as explained in the previous paragraph, the above ('C2') and right ('C3') corners are searched for respectively in the vertical and horizontal directions. From their precise positions, they form with C1 the basis from which to find all the corners of the grid.

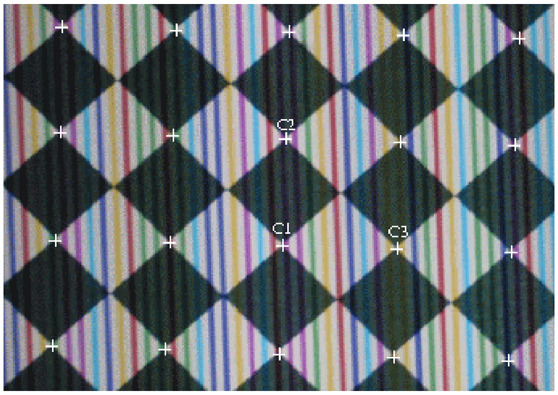


Figure 22: Corner localisation in partial image (Prototype B)

For prototype B with the chessboard object, two types of corners exist, based on the colour (black/white) of the surrounding squares. We retained the corners that separate horizontally two white squares. This ensures the visibility of projected stripes on a horizontal line passing through the corner, allowing for the determination of the stripe index at the corner position.

b) Stripe labelling

Around each corner, a few stripes (typically 5 on each side) are detected and labelled to derive the decimal stripe index at the corner position.

From a horizontal line passing through the corner, stripe detection is performed with vertical following of edges or local minima (Figure 23). Vertical average of horizontal positions x along the stripes allows for noise reduction and subpixel accuracy. The stripe intersections with the horizontal line are ordered in increasing x values.

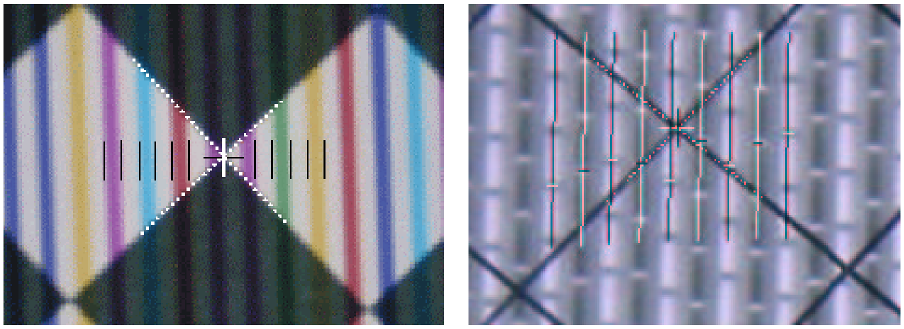


Figure 23: Stripe localisation and identification around a corner for prototypes B and C

For prototype B, the colour of each stripe is determined by comparison of the R, G and B levels with the neighbouring white spaces. For prototype C, dots along the stripes are localised and their vertical positions are compared between consecutive stripes to derive the stripe offset. The stripe labels are obtained by stripe colour or offset matching with the reference sequence as explained in section 2.5.3.

The decimal stripe index at the corner horizontal position is obtained from the stripe intersections and index, supposing a local linear dependence between stripe indexing and horizontal position. Each pair of detected stripes with position and index gives a decimal index value at the corner x position by linear interpolation. For precision, the pair of stripes considered must be at least three stripe widths apart. All the decimal index values are finally filtered by choosing the median value.

c) Stripe indexing versus x position

In general, the interval between consecutive stripes (Δs_x), rather independent from the distance to the object (which affects in an opposite way the width of the stripes on the object and the size of an object in the image), depends on the surface orientation. For a planar object, the stripes appear regularly spaced in the image so that we can approximate the relation between the stripe positions and x position in the image by:

$$x = (s + s_1) * \Delta s_x \ \{ + y / \Delta s_y \}$$

where

s is the stripe index obtained from the distribution of stripe labels of a few neighbouring stripes

s_1 is the constant of the linear assumption

Δs_x is the horizontal increment in pixels between consecutive stripes, supposed to be constant

Δs_y is the correction factor with vertical position in the image when stripes are not vertical in the image

An analysis of vertical edges (horizontal large transitions) in the centre of the image allows a direct evaluation of the stripe interval Δs_X. The determination of stripe indices allows to estimate the constant s_1 in the above relation. If the stripes are not vertical enough in the image, the approximation can be refined with the introduction of a third term taking the vertical image position y into account.

This general relation will help to ensure consistency in the stripe indexing all over the image.

d) Symbolic indexing

Symbolic indexing allows identifying each corner in the space of the reference object.

One corner is used as reference (0 0). It can be the one associated with the marked square. Thanks to symbolic indexing, and assuming a planar object, the theoretical relative 3D positions of each corner is known:

$$X = i*\mathit{dist}$$
$$Y = j*\mathit{dist}$$
$$Z = 0$$

with (i,j) the symbolic coordinates, and *dist* the grid step (distance between corners).

The grid of corners can be roughly derived from three corners defining two basis vectors. The central corner, labelled (0 0), the right corner, labelled (1 0) and the top corner, labelled (0 1).

2.4.5 Parameter optimisation

The estimation of parameters has been largely improved during the development of the different prototypes. They share so little that they are presented separately.

Parameter estimation necessitates a model, data and a criterion for optimisation. Please refer to section 2.4.2 for the model and equations. Section 2.4.4 presents how data is acquired. This section describes the initialisation of parameter values, and how parameter optimisation proceeds to minimise an error measure.

a) Full search from initial values (Prototype A)

The simple ***approach*** of Prototype A consists in estimating initial parameter values and in adapting them by a full search in limited parameter ranges to reduce a 3D error based on distances and planarity of the four detected corners of the square calibration object.

a1) Initial parameter values

The 10 parameters present in the model (angular parameterisation, 2.4.2) are listed in Table 5.

Name	Description	Initial value
D	Camera – projector focal point distance	Ruler [mm]
γ_0	Camera axis angle with OZ in OXZ plane	Rough measure [rad]
ρ_0	Camera optical axis angle with OXZ plane	Supposed 0 [rad]
σ_0	Projector optical axis angle with OZ	Rough measure [rad]
f_{cx}	Horizontal camera focal length	(fc * IMGwidth) / CCDwidth [pix]
f_{cy}	Vertical camera focal length	(fc * IMGheight) / CCDheight [pix]
f_{px}	Horizontal projector focal length	(fp * Nstrp) / slideWidth [strp]
c_0	Image centre abscissa	ImgW/2 [pix]
r_0	Image centre ordinate	ImgH/2 [pix]
s_0	Central Stripe	Nstrp/2

Table 5: Parameter definition and initial values for Prototype A

D, γ_0 and σ_0 receive values from rough measurements of distances between the camera, the projector and an object point illuminated by a central stripe and projected near the image centre (see Appendix C).

Parameter ρ_0 is supposed 0. Its value is normally small by system construction. A change of this parameter mainly implies a translation of all 3D coordinates.

Parameters f_{cx} and f_{cy} depend on camera focal length and CCD size according to the given formula. They are expressed in pixel. If one of these measures is not known, Appendix C details how to get approximations from the camera field of view and image size.

Parameter f_{px} depends on the projector focal length and slide size in a similar way as f_{cx} does.

Parameters c_0 and r_0 relate to the image centre and s_0 to the slide centre.

a2) Parameter optimisation

The initial parameter values are refined so that the distances measured between the four corners are as correct as possible. 3D corner positions are obtained thanks to Equation 1 for current parameter values. For each calibration image, six corner distances are computed from 3D corner positions: four distances correspond to square sides and two distances are diagonals (divided by √2). The error measure to be minimised consists of a distance error and a planarity error. The distance error is the mean deviation of the distances separating square corners with 150 mm (distances corresponding to diagonals are divided by √2). The planarity error is the distance of the fourth corner to the plane defined by the first three corners. The two errors were combined in one measure in mm by the expression

Error = 10*distance error + planarity error.

Among the 10 parameters, only 4 were optimised. Small changes of ρ_0, c_0, r_0 and s_0 mainly lead to a translation, not affecting the error measure. D gives a scale factor and can be estimated by the residual mean value if we change the error measure to compute the standard deviation. As $\gamma_0+\sigma_0$ constant mainly induces a translation, σ_0 was tuned with same increments as γ_0.

The optimisation algorithm is four nested loops of the independent parameters γ_0, f_{cx}, f_{cy}, f_{px}, considering initial ranges of 20 % of the parameter values. For each parameter value, the error consisting of corner distances and planarity is evaluated. The parameter values leading to the minimal error at one iteration are used as central values for the next iteration, considering ranges divided by 2. The process ends after 10 iterations.

The solution is automatic and fairly accurate. Five correct images normally already bring a clear optimum. The major limitation comes from the reduced set of corners. The bad localisation of one of these corners hampers the use of the related image. This is the reason why we considered more images and retained the five images leading to lowest errors.

a3) Results

Results are presented in Appendix C, in the form of estimated distances between corners and planarity. The level of errors is about 1 % in distance measurement and in planarity for the calibration object, when the worst images are rejected.

b) Global optimum with simplex (Prototype B)

For prototype B, the parameter initialisation is similar to that of prototype A. The error measure exploits the regularity and planarity of the grid of corners of the chessboard. Parameter optimisation is based on the simplex algorithm.

b1) Initial parameter values

Compared to prototype A, the model contains two additional parameters: the angle φ and the distortion constant K_{dc}, both initialised to 0.

Name	Description	Initial value
φ	Camera angle around optical axis	0 [rad]
K_{dc}	1st order radial distortion factor	0

Figure 24: Definition and initial values of additional parameters of Prototype B

b2) Parameter optimisation

The ***error criterion*** is based on the specificity and knowledge of the set of reference points (planar grid of corners) for calibration. Mathematically, this is nicely implemented when considering pairs of (non parallel) vectors that join corners. The planarity of the expected grid of corners should tend to align the vector products perpendicularly to the plane, with a magnitude dependent on the vector size in symbolic coordinates only, as imposed by the regular spacing of the grid.

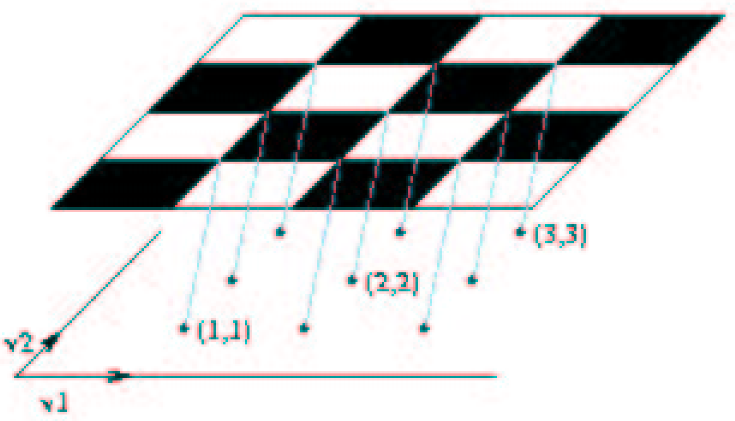


Figure 25: error vectors between theoretical and estimated positions

Practically, two basis vectors v1, v2 were estimated as the mean vectors (although other estimators robust against outliers could be better) of the vectors which join corners along symbolic axes and normalised by the symbolic distances. Each corner detected in the image brings an error vector that joins the theoretical position based on v1, v2 and symbolic coordinates ('symbolic corner') to the corner position obtained from image analysis and current parameter values ('real corner'). The adopted ***global error*** is the standard deviation of the error vectors, corresponding to the dispersion of vectors that should ideally be identical. The mean vector of error vectors has no use. This is an average translation between the real plane and the plane reconstructed from symbolic coordinates and v1, v2.

The ***optimisation procedure*** tries to minimise the global error, sum of the distances between real and symbolic corners, for several images. It is not directly concerned with the global position (translation, orientation) and distance of the reference object. Each image contributes to the error independently from the others. No relative position or orientation between images is taken into account so that the reference object can be presented freely to the camera. However, the visibility and quality (focus) of the stripes should be checked for correct processing, and the reference object should occupy the field of view as much as possible to better estimate the parameters.

The ***optimisation algorithm*** is based on the Simplex algorithm, as implemented in Numerical Recipes [NumRecipes]. For N parameters, N+1 initial points in the space of N parameters must be supplied. We derived N initial points from variations in one component of each of the initial parameter values with some specific parameter increment. The last initial point has all its components modified with opposite increments. This set surrounds the initial parameter values in the space of N parameters.

As for approach 2.4.5a), certain parameters are dependent and are better frozen to start optimisation, speed up processing and avoid artificial solutions. The residual error is not much affected by these unused parameters. The general idea is here again to avoid translation of the solution that does not affect the error criteria based on relative positioning.

b3) Results

This calibration procedure was used for the BIOMET campaign 02 (see section 2.8.3). 12 calibration images were captured during three different days. Blurring obliged us to drop two of them. The direct use of the remaining 10 images resulted in apparently fair results. However, face reconstructions with the obtained parameters indicated a severe distortion and a close look at the parameters revealed a doubtful value of the focal length.

In order to identify possible origins of the problem, we estimated the value of the focal length thanks to another calibration program (see 2.4.5c), Camera calibration). We set the obtained number as initial value and froze the corresponding parameter in the graphical interface (see Appendix E). The residual error after optimisation was about 50 % higher than when the focal length was tuned but the resulting parameters allowed us to derive qualitatively correct facial 3D reconstruction.

We arrived to the conclusion that the poor estimation of the focal length was due to the similarity in position of the reference object among the calibration images. We decided to separate the two problems of camera and projector calibration (next section).

Other tests of calibration were conducted with Prototype B. They indicated an average error of about 1 mm. This value is dependent on the quality (especially the planarity) of the reference object. It appeared that the results and the convergence of optimisation are very good when a large depth of field is considered. This is possible with Prototype B because the important light power of the flash and the rather thick stripes allow an important depth of focus.

c) Separate camera calibration (Prototype C)

The parameter optimisation approach used for prototype C is based on the method developed by Lavest [Lavest98]. It was considered to avoid the problems experienced with prototype B due to the simultaneous optimisation of camera, projector and system parameters that led to large variations in the estimation of intrinsic parameters.

Prototype C has also received a new modelling (projective geometry, section 2.4.2b)), in accordance with Lavest's work. The structure light 3D axis system is centred at the focal point of the camera model with X, Y, Z axes respectively in the direction of the sensor rows, sensor columns and the camera optical axis. Equations giving the 3D coordinates of an image point lit by a given stripe are also lighter (compare Equations 2 relative to Equations 1).

c1) Mathematical model

Lavest, in his work about geometric camera calibration [Lavest98], considers the simultaneous estimation of intrinsic and extrinsic parameters as well as reference point coordinates.

As detailed in Appendix D, the pinhole model for the perspective projection leads to the colinearity equations:

$$\begin{cases} \varepsilon_x = P(\boldsymbol{\Phi}) - c \\ \varepsilon_y = Q(\boldsymbol{\Phi}) - r \end{cases}$$

with

$\varepsilon_x, \varepsilon_y$, the estimation errors, in pixel, along the x and y directions;

$\boldsymbol{\Phi} = [f_x, f_y, c_0, r_0, a1, a2, a3, p1, p2, T_x, T_y, T_z, \alpha, \beta, \gamma]^T$, the vector parameter;

c, r, the column and row image coordinates of reference points.

$\boldsymbol{\Phi}$ must be estimated to minimise the sum of ($\varepsilon_x*\varepsilon_x + \varepsilon_y*\varepsilon_y$) over all reference points.

c2) Problem solving

Writing $P(\boldsymbol{\Phi}) - c$ and $Q(\boldsymbol{\Phi}) - r$ collectively in a vector $\mathbf{V}(\boldsymbol{\Phi})$, the problem is to minimise $\mathbf{V}(\boldsymbol{\Phi})$ by a non linear optimisation procedure. The classical method is to linearise $V(\boldsymbol{\Phi})$ around an initial value $\boldsymbol{\Phi o}$:

$$\mathbf{V}(\boldsymbol{\Phi}) = \mathbf{V}(\boldsymbol{\Phi o}) - (d\mathbf{V}/d\Phi_i)*\Delta\Phi_i$$

and to find the correction $\Delta\Phi i$ to bring to the parameters in order to minimise $\mathbf{V}(\boldsymbol{\Phi})$.

We can rewrite $\mathbf{V}$ as $\mathbf{V} = \mathbf{L} - \mathbf{A}\ \Delta\boldsymbol{\Phi}$, with $\mathbf{L}$ the current value and $\mathbf{A}$ the matrix of derivatives around $\boldsymbol{\Phi o}$.

Introducing $\mathbf{W}$, the matrix of measure weighting (accounting for parameter correlation and measure precision), we have to minimise $\mathbf{V}^T\mathbf{W}\mathbf{V}$ which leads to

$$\Delta\Phi = (A^T WA)^{-1}(A^T WL)$$

Each calibration image, containing n reference points, will give 2n relations between the parameters Φ_i. For m calibration images, this amounts to 2mn equations. The unknowns are 9 intrinsic parameters (f_x, f_y, c_0, r_0, a1, a2, a3, p1, p2) and 6 extrinsic parameters (T_x, T_y, T_z, α, β, γ) for each image, leading to 9+6m parameters.

Camera calibration

The position of each detected reference point of each presentation of the reference object (of known symbolic indices) is predicted into the camera sensor coordinates thanks to the model and compared to detected image positions. As presented in the previous section, optimisation consists in adapting camera intrinsic parameters and rotation - translation matrices to minimise the global error.
The optimisation criterion is typically expressed in pixel. The loop of optimisation ends when a target value has been reached or when no significant improvement occurs anymore. Then points with large errors are discarded if their error is three times the standard deviation of all points.

Camera auto-calibration

So far, the calibration results depend on the quality of the coordinates of the reference points. In our case, for a planar object, the critical point is the planarity assessment of the object. Taking advantage of the considerable redundancy in the system of equations (2mn equations for 9+6m parameters), we can incorporate the reference point coordinates (x,y,z) as additional parameters so that we reach 9 + 6m + 3n parameters. For example, for 10 calibration images of 50 points each, we have 1000 equations for 219 parameters.

Auto-calibration does not only deliver more precise parameters. It also relaxes the requirements on the calibration object. To take advantage of this freedom, the reference points must be consistently labelled among images.

System calibration

Once the camera has been calibrated, we dispose of:

- the camera intrinsic parameters;
- the localisation (roto-translation matrix) for each presentation of the reference object;
- the refined position of reference points in the reference object (from auto-calibration).

Since the optimisation criterion considers distances in the camera sensor coordinate system, the solution needs the projection of 3D positions of the reference points, applying the roto-translation matrices to the refined position of the reference points. Consequently, the 3D position of each reference point in each presentation where it has been detected is available.

System calibration concerns camera calibration and the two remaining parameter types: projector intrinsic and camera/projector extrinsic parameters. We preferred however a global approach, estimating the planes of stripe projection in the camera axis system. We followed the projective geometry approach (section 2.4.2b)), considering 3D point localisation as the intersection of a line of sight from the camera and a projected stripe plane (Equations 2).

The estimation of the ***projected planes*** is based on the 3D position of the reference points estimated during camera calibration and the decimal stripe estimation performed during image analysis. Since plane estimation must rely on a sufficient number of 3D positions, we enlarge the set of available points. Each reference point of each presentation image gives rise to four left and right 'synthetic' neighbouring points thanks to interpolation with respectively left and right immediate reference points (Figure 26). 3D coordinates of the synthetic points are obtained by interpolation at the position of integer stripe values, supposing local linearity (supported by the weak distortion). Considering a mean value of 10 images with 50 reference points, each providing eight synthetic points with integer stripe values, 4000 synthetic points are available for 200 projection planes (stripes), providing on the average for 20 points per plane. Practically, not all stripes are present in images and central stripes possess more reference points, leading to more than 20 synthetic points per plane.

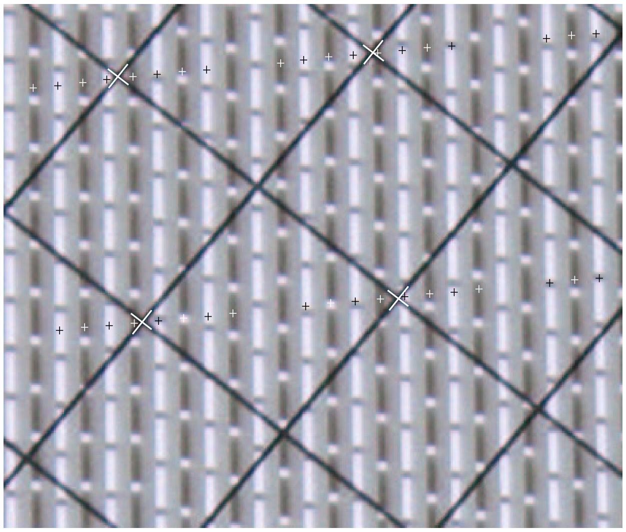


Figure 26: Synthetic points obtained by interpolation

To determine a ***plane from belonging points***, the best numerical approach is to consider the plane normal (*a, b, c* coefficients, normalised to 1.0 as unit vector)

$$a*\mathrm{x} + b*\mathrm{y} + c*\mathrm{z} = d$$

and to compute *d*, the distance to the plane from each point (x,y,z). As *a* is large (planes are nearly vertical (y) and mainly along z), we adapt *b* and *c* (with the constraint for '*a*' of keeping a unit vector) to minimise the variance of *d* for all points. Due to the monotonic variation of the variance with values *b* and *c* over a large range of values, a successive approximation in decreasing increments works perfectly, leading to a fast and accurate implementation. The values *a*, *b* and *c* for minimal variance are used as plane normal coefficients and the corresponding mean value is used for *d*.

Individual plane equations must follow a ***global constraint***. Theoretically, all planes cross along a line passing through the projector focal point, with a constant increment of the plane normal due to the regular spacing of stripes. In practice, the projector suffers from distortion, compromising the planar assumption of stripes and their regular spacing. Nonetheless, typical distortion magnitudes are moderated and the planar model helps deriving planes for which there were no or few points collected.

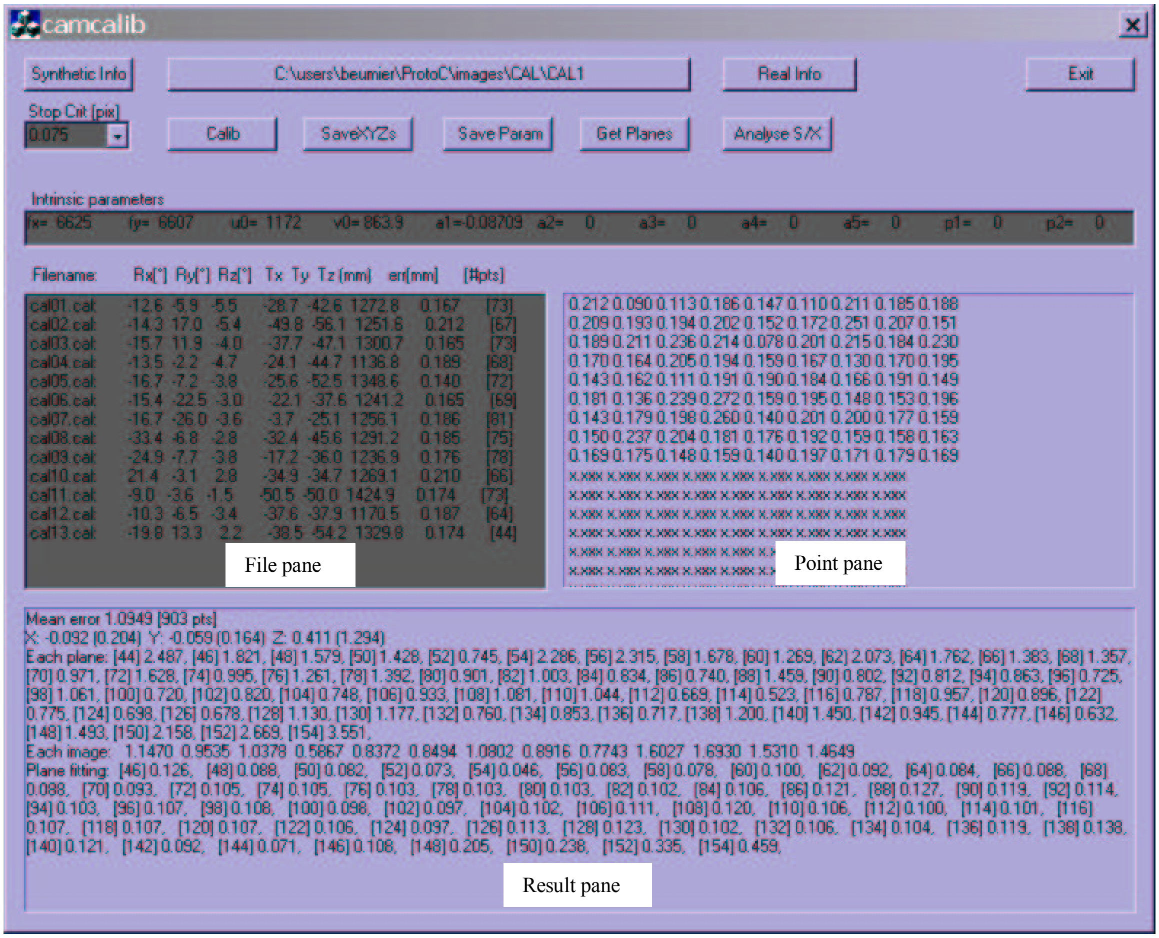
c3) Results

The calibration procedure for Prototype C first considers the camera calibration. The error criterion, expressed in pixel, allows verifying the quality of the calibration data (subsection 2.4.4). In the case that the residual error is high, the reference point localisation and labelling (stripe index) should be checked.

The error measure we used for camera calibration is the standard deviation of the distance in pixel between the projection of 3D reference points onto the image plane and the localisation of the corresponding points in the captured image. In our experiments, the residual error after optimisation ranges from 0.1 to 0.2 pixel. This value is dependent on the quality of the camera, the correctness of the reference point extraction procedure and the adequacy of the mathematical model. The influence of the quality of the reference pattern has been minimised thanks to the auto calibration approach (see 2.4.5 c2).

As a second and final calibration step, the system calibration identifies the planes of all the projected stripes. To estimate the error in 3D localisation, we compared the theoretical 3D positions known from the reference object with the 3D coordinates obtained by the model of subsection 2.4.2b) (Projective geometry) and dependent on the estimated planes. This comparison leads for each point to an error expressed as a distance in mm. In order to analyse the distribution of errors with the projected stripes, an error per plane is computed as the average of the errors related to the points close to each plane. Another way to present the errors is to display the average error of the points of each image capture. This allows for discarding images responsible for large contribution to the errors and which mostly correspond to large rotations of the calibration object. Finally, the average error of each reference point is presented as an array of errors in mm. This gives the possibility to analyse the influence of a specific point of the reference object.

The calibration procedure was tested on a few sets of calibration data. A few tests were carried out on the data gathered for prototype B, for comparison purposes. Results were improved when the calibration object was not planar or when the object only occupied a reduced volume of the capture space. Figure 27 shows the graphical interface developed for the calibration of prototype C, in the status of having completed the optimisation for the calibration of data captured by the prototype C.



Stop Crit [pix] 0.075

Intrinsic parameters

```
fx= 6625   fy= 6607   u0= 1172   v0= 863.9   a1=-0.08709  a2=  0   a3=  0   a4=  0   a5=  0   p1=  0   p2=  0
```

```
Filename:   Rx[°] Ry[°] Rz[°]  Tx  Ty  Tz (mm)  err[mm]  [#pts]
cal01.cal   -12.6  -5.9  -5.5   -28.7 -42.6 1272.8  0.167  [73]
cal02.cal   -14.3  17.0  -5.4   -49.8 -56.1 1251.6  0.212  [67]
cal03.cal   -15.7  11.9  -4.0   -37.7 -47.1 1300.7  0.165  [73]
cal04.cal   -13.5  -2.2  -4.7   -24.1 -44.7 1136.8  0.189  [68]
cal05.cal   -16.7  -7.2  -3.8   -25.6 -52.5 1348.6  0.140  [72]
cal06.cal   -15.4 -22.5  -3.0   -22.1 -37.6 1241.2  0.165  [69]
cal07.cal   -16.7 -26.0  -3.6    -3.7 -25.1 1256.1  0.186  [81]
cal08.cal   -33.4  -6.8  -2.8   -32.4 -45.6 1291.2  0.185  [75]
cal09.cal   -24.9  -7.7  -3.8   -17.2 -36.0 1236.9  0.176  [78]
cal10.cal    21.4  -3.1   2.8   -34.9 -34.7 1269.1  0.210  [66]
cal11.cal    -9.0  -3.6  -1.5   -50.5 -50.0 1424.9  0.174  [73]
cal12.cal   -10.3  -6.5  -3.4   -37.6 -37.9 1170.5  0.187  [64]
cal13.cal   -19.8  13.3   2.2   -38.5 -54.2 1329.8  0.174  [44]
```

```
0.212 0.090 0.113 0.186 0.147 0.110 0.211 0.185 0.188
0.209 0.193 0.194 0.202 0.152 0.172 0.251 0.207 0.151
0.189 0.211 0.236 0.214 0.078 0.201 0.215 0.184 0.230
0.170 0.164 0.205 0.194 0.159 0.167 0.130 0.170 0.195
0.143 0.162 0.111 0.191 0.190 0.184 0.166 0.191 0.149
0.181 0.136 0.239 0.272 0.159 0.195 0.148 0.153 0.196
0.143 0.179 0.198 0.260 0.140 0.201 0.200 0.177 0.159
0.150 0.237 0.204 0.181 0.176 0.192 0.159 0.158 0.163
0.169 0.175 0.148 0.159 0.140 0.197 0.171 0.179 0.169
x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx
x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx
x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx
x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx x.xxx
x.xxx x.xxx x.xxx x.xxx x.xxx x.
x.xxx x.xxx x.xxx x.xxx x.xxx x.
```

```
Mean error 1.0949 [903 pts]
X: -0.092 (0.204) Y: -0.059 (0.164) Z: 0.411 (1.294)
Each plane: [44] 2.487, [46] 1.821, [48] 1.579, [50] 1.428, [52] 0.745, [54] 2.286, [56] 2.315, [58] 1.678, [60] 1.269, [62] 2.073, [64] 1.762, [66] 1.383, [68] 1.357,
[70] 0.971, [72] 1.628, [74] 0.995, [76] 1.261, [78] 1.392, [80] 0.901, [82] 1.003, [84] 0.834, [86] 0.740, [88] 1.459, [90] 0.802, [92] 0.812, [94] 0.863, [96] 0.725,
[98] 1.061, [100] 0.720, [102] 0.820, [104] 0.748, [106] 0.933, [108] 1.081, [110] 1.044, [112] 0.669, [114] 0.523, [116] 0.787, [118] 0.957, [120] 0.896, [122]
0.775, [124] 0.698, [126] 0.678, [128] 1.130, [130] 1.177, [132] 0.760, [134] 0.853, [136] 0.717, [138] 1.200, [140] 1.450, [142] 0.945, [144] 0.777, [146] 0.632,
[148] 1.493, [150] 2.158, [152] 2.669, [154] 3.551,
Each image:  1.1470  0.9535  1.0378  0.5867  0.8372  0.8494  1.0802  0.8916  0.7743  1.6027  1.6930  1.5310  1.4649
Plane fitting:  [46] 0.126,  [48] 0.088,  [50] 0.082,  [52] 0.073,  [54] 0.046,  [56] 0.083,  [58] 0.078,  [60] 0.100,  [62] 0.092,  [64] 0.084,  [66] 0.088,  [68]
0.088,  [70] 0.093,  [72] 0.105,  [74] 0.105,  [76] 0.103,  [78] 0.103,  [80] 0.103,  [82] 0.102,  [84] 0.106,  [86] 0.121,  [88] 0.127,  [90] 0.119,  [92] 0.114,
[94] 0.103,  [96] 0.107,  [98] 0.108,  [100] 0.098,  [102] 0.097,  [104] 0.102,  [106] 0.111,  [108] 0.120,  [110] 0.106,  [112] 0.100,  [114] 0.101,  [116]
0.107,  [118] 0.107,  [120] 0.107,  [122] 0.106,  [124] 0.097,  [126] 0.113,  [128] 0.123,  [130] 0.102,  [132] 0.106,  [134] 0.104,  [136] 0.119,  [138] 0.138,
[140] 0.121,  [142] 0.092,  [144] 0.071,  [146] 0.108,  [148] 0.205,  [150] 0.238,  [152] 0.335,  [154] 0.459,
```

Figure 27: Calibration interface for prototype C and results after optimisation

The interface lists the values of the camera intrinsic parameters. In the current experiment, only one distortion parameter (a1) has been used. The graphical interface continuously displays the errors when looking for the best set of parameters. Several presentations of the errors are displayed. In the ‘File pane’, the information about each image is given: the filename, the values of the extrinsic parameters (rotation in degrees and translation in millimetres), the average error in pixel and the number of points considered. In the “Point pane”, the error in pixel averaged over all the files is given for each point of the reference object. In the “Result pane”, errors are specified in millimetre.

The “Mean error” is the average of the 3D error for all the points that were kept during optimisation. Here the value of 1.09 mm is precise enough, considering that the reference object was placed at distances of more than 1.1 m. The second line gives the X, Y and Z part of the error, with the standard deviation between parentheses. As expected, the depth error (along Z) is much larger than the X and Y errors. Then the distribution of the 3D errors of points is grouped into the closest plane corresponding to a stripe index (indicated in brackets). We see that the error is lower in the centre of the pattern (stripe index around 100), what is advantageous for the capture of objects placed near the image centre, which is a normal situation for face. Finally, the average 3D error (in mm) per image is given.

2.5 3D extraction

This section describes the different steps needed to deliver 3D points extracted from images with structured light illumination.

The stripes are first localised with high pass filtering (section 2.5.1). A label is then given (section 2.5.2) according to a given property along the stripes like thickness, colour or position of dots. A unique stripe index (section 2.5.3) is derived for each stripe from the distribution of stripe labels among a few neighbours. Several filters (section 2.5.4) are then applied to improve the correctness of stripe indexing. Image position and stripe index are converted into 3D positions (section 2.5.5) that are stored in a file according to several possible 3D file formats (section 2.5.6).

2.5.1 Stripe detection

Basically two different techniques for stripe localisation have been developed. The first one detects the ***stripe borders***, localised by maximal values of intensity gradients (derivatives). The second one localises the ***stripe centres***, from the local minima or maxima obtained by the zero crossings of the image intensity first derivative.

As the three prototypes utilise vertical stripes, the ***derivative is implemented*** as a digital ***horizontal difference*** of intensity points separated by a given number of pixels (see below). Depending on the stripe pattern, the intensity can be the grey value (black & white camera), a colour field (R, G, B) or a combination of colour fields (see below).

a) A reference size

To a certain extent, depending on surface orientation, the ***stripe width in pixel*** is rather constant in images of a given camera/projector arrangement, leading to a practical size unit for size-dependent image processing tasks. This property can be expected because the camera and projector optical axes do not deviate much (a tenth of degrees) to keep an important area of capture (receiving projected light and seen by the camera). Also, because the projector-to-camera distance is short (10 cm) compared to the distance to the scene (1m), the enlargement of projected stripe width with distance is compensated in image size (in pixel) by the camera perspective projection.

The apparent stripe width in the image mainly depends on the surface orientation. For the facial surface, the largest variation in stripe width is to be found in areas around the nose and the throat.

b) Noise reduction

Several techniques have been employed for ***noise reduction*** in order to increase stripe detection quality.

Robustness of stripe detection can be increased with a ***vertical lowpass*** filter. The vertical nature of the stripes is consolidated with vertical averaging while noise is reduced. The positive effect is that stripes are better detected and localised thanks to the continuous nature of the stripes, allowing for subpixel accuracy. Large depth discontinuities are not adequate for lowpass filtering, but these regions are usually problematic for 3D capture. Lowpass filtering has been applied with a kernel size equivalent to the stripe width.

The regular spacing of stripes allows for ***geometrical filtering*** of the detected points. In regions where the surface is smooth, the horizontal distribution of stripes remains rather regular, as the slide pattern is. Stripe border or centre detection benefits from this property, keeping only one detected point in a horizontal window whose size is based on the stripe width.

The gradient estimation, based on the intensity difference at two points, gives results largely dependent on the ***horizontal distance*** separating the two points. For small distances, the estimated gradient is limited and subject to noise, leading to many false detections. Too large distances include more than the transition, leading to inaccurate localisation. An appropriate distance, given below for each prototype, improves the quality of the detected points by reducing the influence of noise.

c) Maximal gradients

Maximal gradients highlight the ***stripe borders***. They can provide more points (two edges per stripe if both sides are useable).

Edge localisation is more sensitive to noise, especially in the case of colour striping with transitions only in one of the colour fields (R, G or B).

The ***horizontal distance*** separating pixels of the grey-level difference (gradient) must be large enough to encompass most of the transition, for good precision and lower sensitivity to noise (isolated or parasite gradients), but small enough to keep localisation precision. This value is not critical and must roughly follow the optical properties materialised by the point spread function (combined effects of the camera and the projector).

d) Minima and maxima

Extreme (minima and or maxima) values of image intensities highlight ***stripe centres***, precisely localised from the zero crossings of the derivative if the stripes are thin. If the stripes are wide, the approximation of a local planar surface might not be valid. In this case, the stripe centre may not be at equal distance from the stripe edges, and imprecision could arise.

The ***horizontal distance*** separating pixels of the difference (gradient) must be large enough to filter out (spurious) local sign inversion of the gradient due to noise. A distance equal to the stripe width is convenient. Distances larger than the stripe width will integrate the influence of neighbouring stripes and reduce accuracy.

e) Prototype A

The light pattern of prototype A (Figure 7) possesses edges for localisation on the left of each opaque stripe. These are detected as maximal gradients (Figure 28). With the black and white camera we used, the gradient was computed as the grey-level difference of two points separated horizontally by 2 pixels. The grey-level transition at stripe edges indeed spreads over a few pixels due to the important contrast of the stripes.

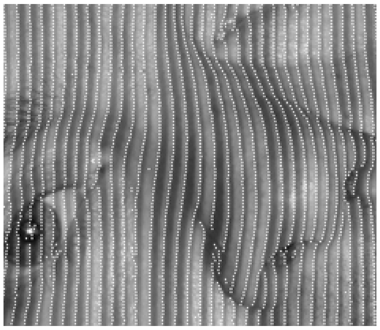


Figure 28: Left edge detection (Prototype A)

f) Prototype B

The colour pattern of prototype B was designed to deliver two points per colour stripe. These correspond to the stripe borders and are detected as maximal gradients. With the high resolution of the digital camera we used, the gradient was estimated as the intensity difference in each colour field for two points separated horizontally by 6 pixels. The colour gradients in the R, G and B fields with sufficient contrast are used for localisation and give clues about the colour of the stripe.

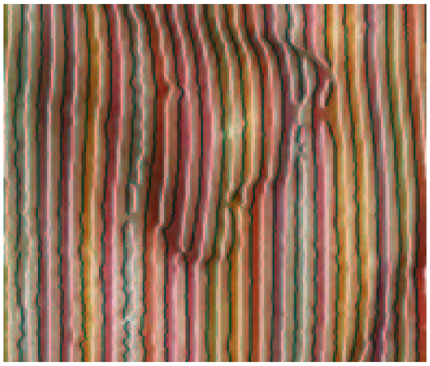


Figure 29: Stripe edge detection (Prototype B)

For the database acquired in the BIOMET project [BIOMET], the image quality concerning contrast and colour definition was unfortunately too weak to detect colour stripe edges. We successfully detected stripe centres with minima localisation in the R, G and B fields (Figure 30). This does reduce by two the intended horizontal density of extracted points.

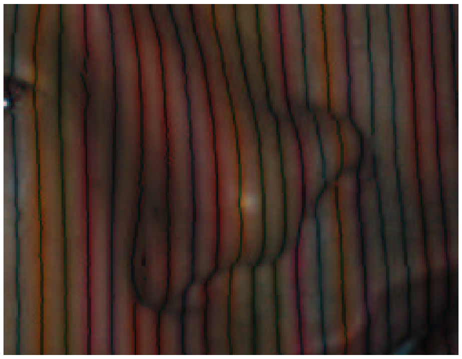


Figure 30: Stripe centre detection (Prototype B)

g) Prototype C

The black and white pattern of Prototype C (slide of subsection 2.3.4c6), Figure 14) alternatively provides maxima and minima thin stripes. Stripe centre localisation is achieved with zero detection of the gradient for two points distant of 6 pixels (Figure 31).

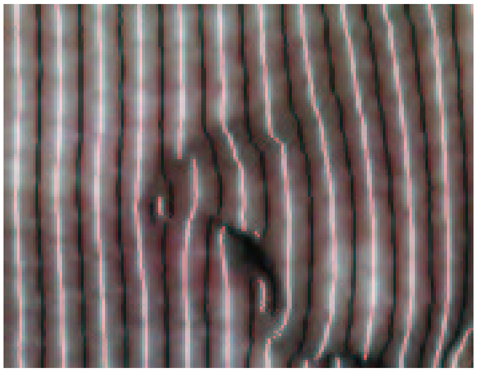


Figure 31: Stripe centre detection (Prototype C)

Vertical averaging also helps in the present case to detect stripes continuously, by filling the gaps due to the dots on the stripes.

2.5.2 Stripe labelling

The objective of this section is to detail image processing tasks to perform ***stripe labelling***. In order to achieve triangulation for 3D estimation, the stripes localised in the image must be identified. The identification of each stripe is distributed in several neighbouring stripe labels. A stripe label is the belonging to a class dependent on some stripe property. Prototype A, B and C have respectively used the stripe thickness, the stripe colour and the vertical position of dots to label the stripes.

a) Prototype A

Prototype A projects stripes with two ***thickness*** possibilities (thin or thick) according to a predefined binary sequence. Each opaque stripe belongs to an opaque/transparent pair of constant width. The opaque part is either half or double as wide as the transparent part. Thickness evaluation consists in comparing the thickness of the dark and bright parts of the opaque/transparent pairs.

To estimate thickness, we subtract from each pixel grey level ('Original grey levels' in Figure 32) the average value of grey levels ('Lowpass filtered grey levels' in Figure 32) on the horizontal as far as five stripe pairs away from the considered pixel. Pixel values after subtraction are summed along the stripe pair. A negative sum indicates a thick (dark) stripe while a positive sum accounts for a thin (bright) stripe. See Appendix C.

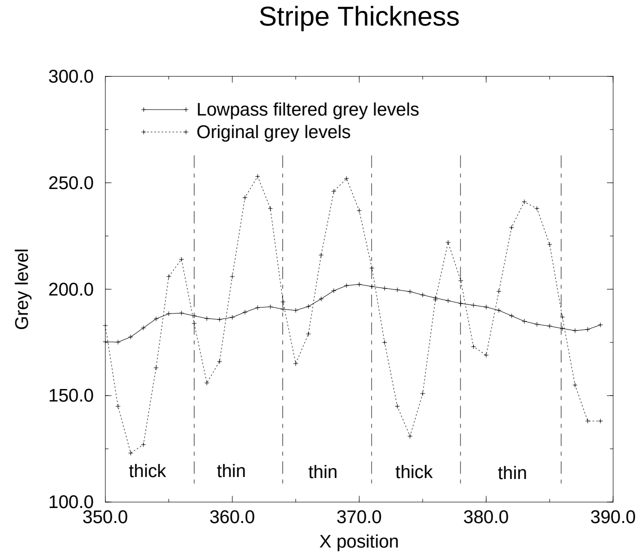


Figure 32: Stripe thickness estimation

Although thickness estimation is based on an adaptive threshold (local average), errors may occur. These are mainly due to local texture influences or saturation. To improve robustness, individual estimations on horizontal lines can be merged (vertically) along the stripes.

b) Prototype B

Prototype B projects ***colour stripes*** separated by white stripes. Each colour stripe must be labelled as one of the six possible colours. A direct estimation of the colour based on the R, G, B gradients, computed for stripe edge localisation, was investigated but led to the difficulty to set thresholds. Also, the fields R, G and B cannot be compared relatively to each other, as their ranges might be much different.

We preferred to estimate a colour hue, angular representation of the colour, providing more continuous values than the discretisation into one of the six possible colours.
We introduce the ***relative colour hue***, based on the R, G, and B gradients according to:

```
//Let dR, dG, dB be the gradients in R, G, B channels
// the basic hue value depends on the dominant field
max = max{dR,dG,dB}
if (dR ==max) hue = 0; if (dG == max) hue = 120; if (dB == max) hue = 240;
// Refinement from the two weakest fields, proportionally
min = min{dR,dG,dB}
mid = middle{dR,dG,dB}
ratio = (mid-min) / (max-min)
hue = hue + (60*ratio)
// Invert hue and cast to [0..360[ (in 'degrees')
hue = (hue + 180) modulo 360
```

This follows the definition of the hue except that colour gradients are considered, which necessitates an inversion (+ 180). As the gradients relate to the difference between colour and white stripes, the hue estimation is less dependent on the colour of the underlying surface.

The hue value can be converted into the closest basic colours (see Figure 9). We preferred to keep the more precise hue values for matching colours.

c) Prototype C

Prototype C is somewhat particular for stripe label estimation because the stripe property is not distributed along the stripe but at some vertical positions. As a consequence, the stripes are better first connected before they can be followed to look for discontinuities corresponding to label dots.

Label dots are detected as stripe discontinuities. To reduce false detections that can result from depth or texture discontinuities, dots are localised by pair, with an imposed vertical distance separating each pair. This distance is obtained during calibration.

Detected dot positions of a stripe are then compared to neighbouring stripe dot positions to obtain a vertical offset that can be classified into one of the six labels. But again, we prefer to keep the vertical offset value before classification in order to compare labels with the reference sequence with greater precision.

2.5.3 Stripe identification

Stripe identification consists in determining the index of each stripe so that triangulation can be applied from the knowledge of image position and stripe identity. The index of a stripe is obtained by comparison of a few neighbouring stripe labels with the reference sequence of labels.

Two opposite methods have been envisaged. The first one initially creates a ***look up table*** with all possible label n-uples and corresponding stripe index. Each series of n labels of neighbouring stripes collected from the image can be directly converted into the stripe index, as stored in the Look Up Table. This method is very fast but expects a correct label estimation. It can be refined by considering larger series of labels (as long as the lookup table size remains reasonable), with the possibility of giving an impossibility status to table entries of impossible label combinations. In that respect, a Gray code can help recovering labelling errors thanks to redundant coding.

The second method consists in looking for the ***best match*** between the series of n labels and the reference sequence. The method is more flexible, allowing for any kind of distance measure. In particular, the distance does not have to be estimated with labels but can use the image measurement. For instance, the hue measurement of colour stripes with prototype B can be compared with theoretical hues of the colour sequence. This avoids the error prone classification into colour classes.

The lookup table approach requires high quality labelling. Otherwise, longer label sequences must be considered, which is not possible with reasonably long tables. On the contrary, best match between sequences is well adapted to longer sequences. Facial surfaces present smooth areas normally delivering long sequences of correct labels.

Let us unify the output of the stripe labelling process to [0..1], corresponding for prototype A to theoretical values 0.33 (thin) and 0.66 (thick); prototype B to 0.08 (red), 0.25, 0.42, 059, 0.76, 0.92 (magenta); prototype C to 0.08, … 0.92 (6 classes).
A set of n measured labels is compared to all sub-sequences of length n of the reference sequence. We used as score to be minimised the average value of the absolute differences between the measure and the theoretical value of label outputs.

The value of the length n is not critical. Larger than the minimum for uniqueness, and some units more for robustness, we took the opportunity of large smooth regions in the face to set n large (prototype A: 12; B: 10; C: 5).

Normally, thanks to statistical properties, the larger the sequence the more reliable the results. This holds true as long as no depth discontinuity is encountered. To handle small discontinuities present in facial surfaces, one possibility is the ‘Dynamic Space Warping’, similar to the well-known ‘Dynamic Time Warping’ techniques used for speech recognition. The idea is to match the longest sequence, accepting discontinuities. We preferred to find numerous smaller, possibly erroneous matches that will be filtered later.

Matching returns a stripe index for each vertical position along the stripe. Each stripe receives as index the maximum occurrence of the returned indices.

2.5.4 Filtering

Several sources of error may arise from the previous algorithms, leading to incorrect stripe identification. Most of these errors can be ***filtered*** out when referring to a facial surface model containing rather smooth surfaces of connected components.

A first filter is applied to remove the ***isolated points***. A point that does not possess enough neighbours is not expected to belong to the face. Points are vertically connected into stripes if they have little horizontal distance and if they share a similar label. Stripes are rejected if they are short or if they do not have many horizontal neighbours.

A second filter ensures the global ***coherence of the face***, suppressing other objects or background. On the facial surface, depth discontinuities are limited in number and amplitude. As the face is supposed to be central in the image and to represent the majority of pixels, the relation between the stripe index and the horizontal position of the facial stripes is roughly linear. If we account for depth discontinuities of one centimetre on the face, stripe indices may deviate from the linear approximation up to a few stripe indices (typically 3 to 5).

Finally, constantly referring to a facial surface, a last filter ensures the ***smoothness of the face*** by reducing local depth discontinuities due to deep holes or bumps. A lowpass filter with uniform kernel of 10x10 mm is applied. A 2D median or nearest neighbour filter (see [Hoffman87]) would probably perform better, smoothing the facial surface, while keeping most of the details.

2.5.5 3D conversion

The formulae given in section 2.4.2 converts the image position (*c,r*) and stripe index *s* into the 3D coordinates X,Y,Z.

Every detected point with stripe identification *s* that was not filtered is part of the extracted surface.

2.5.6 File storage

Several formats are available for 3D storage to file.

We first developed our proprietary format '.xyz' which efficiently stores X, Y, Z values on 2-byte integers in a binary file with a 0.1 mm resolution. Each stripe is considered successively, storing in increasing Y (increasing *r* in the image). This format is compact (6 bytes per point) and well adapted to stripe intersection with planes thanks to the point ordering. Refer to Appendix G for details.

For display purposes, especially for rendering textures and light, standard formats like .obj and .wrl have been used. For .obj, a list of vertices consisting of the retained points is first stored in ASCII, listing one vertex per line. Then triangular patches ('facets') are specified with 3 indices referring to the vertex list. This format supports point, wire frame and solid display with occlusion and shading. For .wrl, additionally to the vertex and facet lists, a texture list refers to relative coordinates in a specified texture image that will be used for texture mapping at display.

2.6 Texture estimation

With the objective to acquire 3D geometry and texture rapidly and in perfect correspondence, we designed slides with special care for texture measures (except prototype A).

2.6.1 Prototype A

In the first ***prototype A***, about 50 % of the area is projected with full light. By the time of development, we were not concerned with texture recovery, and the slide design concentrated on stripe localisation and labelling. For face recognition, a very low resolution grey measure was obtained by averaging over several stripes, giving rough but useful cues about eyebrows, eyes, nose or mouth position.

2.6.2 Prototype B

The slide of ***prototype B*** is better suited for texture colour estimation. 50 % of the slide is transparent for white light projection, while the colour stripes filter one (25 %) or two (25 %) out of the three colour fields R, G and B. The recovery of texture is based on the compensation of absent colour fields in the projected light from directly surrounding white stripes. The position and theoretical colour of the stripes are known following the stripe analysis presented in section 2.5.

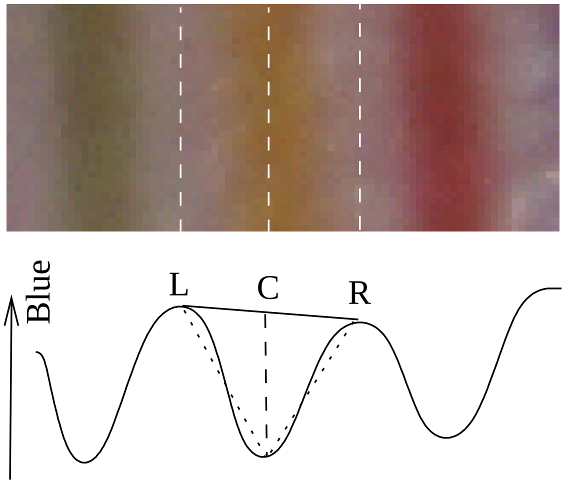


Figure 33: Texture compensation by linear segments

The influence of a stripe extends beyond its visible edges (Figure 33), so that neighbouring white stripes also have to be corrected. For each absent colour field of a colour stripe C, the intensity values of that field at the horizontal central positions of the left (white) L, the right (white) R and the current (colour) C stripes are evaluated with a average of three pixels in each direction for better estimation of the intensity. The real intensity values between L and R are added a value to obtain linearity between L and R, assuming linearity between L and C, and C and R (roof model). In reality, the transitions are ‘S’ shaped, so that over and/or under estimation of the correction apply to some of the corrected values.

Although not perfect, this simple model already smoothes out much of colour striping. The residual colour traces does not affect the recognition of the face and is not visible in black and white (intensity image).

The whole picture (Figure 34 right) was finally lowpass filtered (uniform filter with size equivalent to a half stripe width) to reduce the above-mentioned artefact. The reduction in quality is limited and should therefore not negatively affect applications like face recognition or low-resolution rendering.

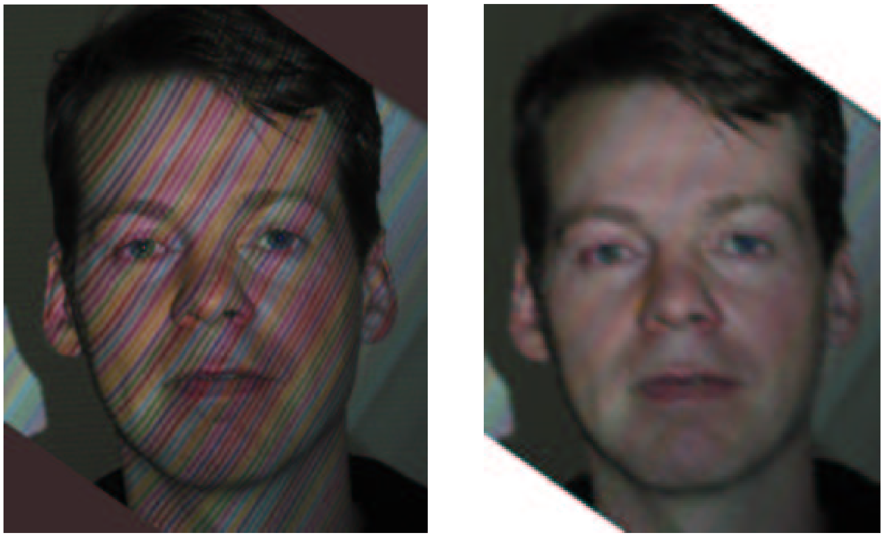


Figure 34: Texture compensation

2.6.3 Prototype C

The projection for ***prototype C*** contains constant (grey) illumination and thin stripes (slits). The intensity profile across the stripes (horizontally) is rather linear. The stripe analysis has led to the determination of the stripe centres. Stripe compensation is carried out by a procedure similar to the one developed for prototype B.

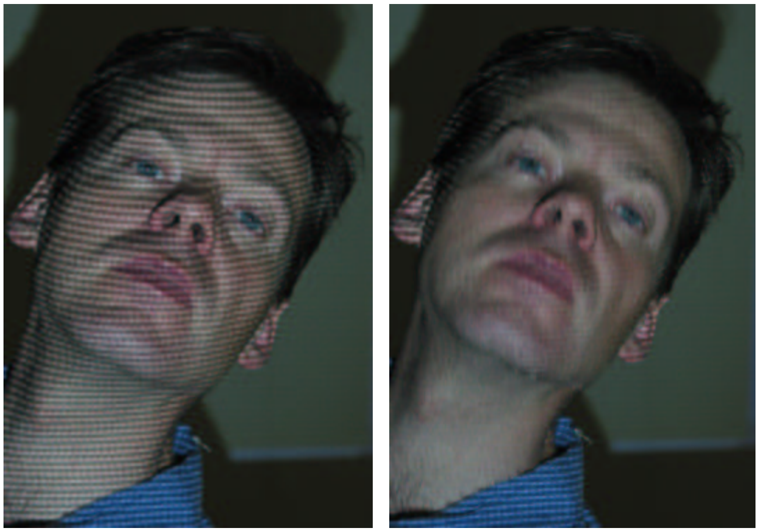


Figure 35: Texture compensation for prototype C

The advantage for prototype C is that no colour artefacts are visible in the compensated image.

2.7 Prototype characteristics

The successive developments have led to three major prototypes named “B&W thickness” (prototype A), “Colour striping” (prototype B) and “Stripe with dots” (prototype C), in chronological order.

Their main features are summarised in the next table. They followed an evolution detailed in Appendix F. Refer to the previous sections for deeper technical explanations.

Prototype:	**Prototype A**	**Prototype B**	**Prototype C**
Nickname	“B&W thickness”	“Colour striping”	“Stripe with dots”
Date	1999	End 2002	End 2003
Full description	Appendix C	This chapter	This chapter
Camera			
Type	Panasonic WV-BL600	Canon G2 / colour	Canon G2 / colour
Image size	768x576 (digitiser)	2272x1704	2272x1704
Texture	No	Yes	Yes
Projection			
Type	Projector 150 W	Flash lamp	Video projector
Slide	B&W print on glass	Colour film slide	BMP file
Pattern	Vertical lines	Vertical lines	Vertical lines
Code	Line thickness (2)	Line colour (6)	Dots on line
Calibration			
Object	One large square	Chessboard	Grid of black lines
Optimisation	Semi auto	Simplex	Gradient descent
3D			
Point density on face	50x70	60x80	70x100
Database	2 x 120pers / 3shots	1 x 81pers / 6 shots	None
Output format	Proprietary	‘.obj’	.wrl (VRML), .obj

Table 6: Main characteristics of the three acquisition prototypes

2.8 Database

2.8.1 Introduction

A database first serves the purpose of quality assessment of the acquisition system. A ***qualitative evaluation***, often called validation, can be performed from the visual appearance, mentioning the typical problems encountered (see subsection 2.8.2c)). In the case of faces, these relate mainly to the hair and beard, iris, eyebrows and nostril regions. These problems are very important as they largely affect the visual appearance and the recognition performance. They commonly overpass the acquisition precision limitation and may fool the conclusions about the 3D recognition potential derived from the comparison experiments.

A second role of the database is to provide data for the recognition experiments that may lead to a ***quantitative evaluation***, such as the recognition rate. The number of individuals, the ability to represent a given population and the variability of each individual and of the shooting conditions are important criteria to take into account when defining the database requirements. The storage requirement of a database is probably the most important design criterion as it limits the variation of other criteria in order to keep a database of manageable size. We concentrated the variability of our databases in the number of individuals (around 100) and in the variability of poses (3 or 6 different viewpoints). Special subject's attitudes or decoration were not considered as people do cooperate in our scenario. We think that the lightening conditions were kept similar from capture to capture thanks to the projector illumination.

Although three prototypes have been developed, only the first two ones were used to collect a database. Prototype A collected 120 people (RMA, Brussels) during two sessions separated by two months and a third session one year later with 100 of the 120 individuals. Prototype B collected 81 persons (ENST, Paris) during the third campaign of the BIOMET project. Both databases captured people in the close neighbourhood for later availability during additional sessions. Only the author is common to the two databases. Both databases aimed at providing 3D data for face recognition experiments.

2.8.2 The 3D_RMA database (Prototype A)

The ***3D_RMA*** database is the 3D facial part of data collected at the SIC (Signal & Image Centre) of the Royal Military Academy in Belgium, called the ***SIC_DB*** database [Beumier99c]. Refer to Appendix G to learn how to get the 3D_RMA database.

Although the SIC_DB database initiative arose from the desire to test our structured light acquisition prototype and the need to dispose of sufficient data for 3D recognition experiments, we took the opportunity to capture speech, frontal and profile images for other person recognition activities of the department.

As detailed in [Beumier99c], the speech data consists of French utterances of imposed digit sequences and sentences, and 20 seconds of free speech, to allow testing the different kinds of speaker verification methods (text-dependent, prompted text, text-independent, verbal information). Two frontal images and one or two profile images per session are supplied, with a blue background to facilitate face segmentation.

Nearly at the same period, the multi-modal database XM2VTSDB was acquired [XM2VTSDB98], as a by-product of the M2VTS consortium, containing 4 sessions of 2 shots of 295 people, including voice records, frontal and profile images. A high quality 3D shot was acquired during the third session with a stereo-based 3D camera developed by the Turing Institute (http://xm2vtsdb.ee.surrey.ac.uk).

a) Population

To be worth the effort, we imposed a minimum of 100 people to be present in the database. On the one hand, we decided to go beyond the common size (30-40) of our previous experiments. On the other hand, we tried to keep a reasonable size to be able to have many sessions. Indeed, we selected people among the academic population likely to stay in our reach for several years. This will allow to make long term studies about facial and voice characteristics.

This database is intended for use in recognition experiments in cooperative situations. The person wants to be recognised and obeys to given guidelines. Pictures are taken in the sitting attitude to reduce position variations and microphone distance. People are asked to gaze in a few directions.

Two sessions were first acquired: one in November 97 and the second in January 98. They both contain the same 120 individuals. Only a few of them were not correctly represented in some shots. 80 persons are students from the Military Academy. The remaining 40 people come from the teaching and research staff. From this population, only 14 women were present. Hair is most of the time short, there are few black and no Asian people. A third session was captured in May 99, consisting of 100 individuals from the 120 original ones.

b) 3D shots

Each session contains four images per person related to 3D capture. The first image was shot without stripe projection, to give a texture image of a frontal pose. This image was rather dark as ambient light was used to have balanced illumination.
The second image was taken directly after, with the projector switched on.
The third and fourth images (with projection) were then shot, with limited (up to 15°) left/right and up/down head rotations of the individuals.

Each session has thus three striped shots and one grey shot in near correspondence with the first striped shot. Remember that these images were taken with a 40° inclination of the camera/projector head (see Figure 77) to reduce grey-level interference of eyebrows, mouth and nostrils with stripes. An example is given in Figure 36.

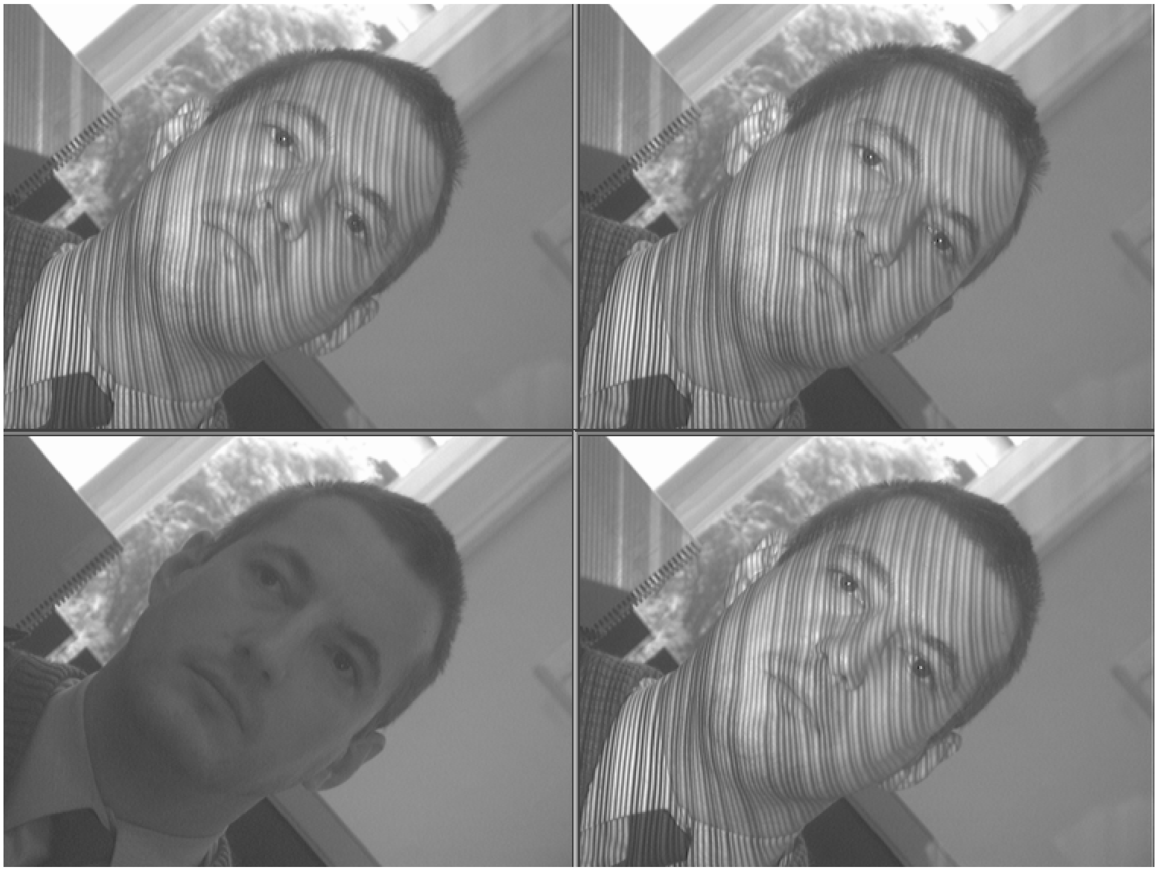


Figure 36: Samples of grey and striped images from the SIC_DB database

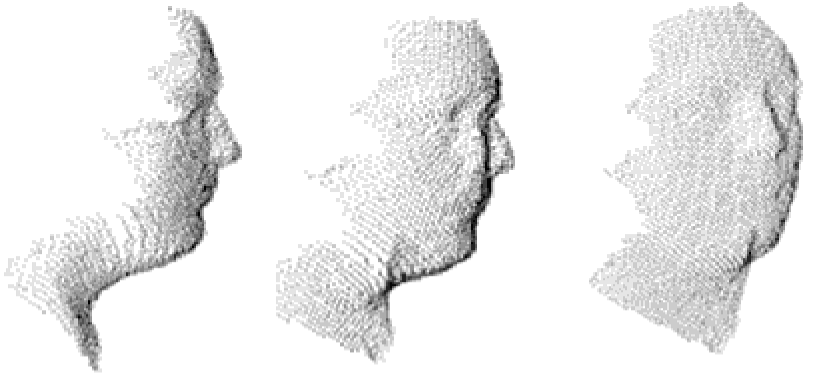


Figure 37: 3D reconstructions from top right image of Figure 36

The third session was acquired with major modifications concerning the camera and projector arrangement. To reduce the size of the 3D camera/projector head, the camera and projector were brought nearer, unfortunately with an inversion in their relative positions due to the cumbersome fan cooler of the projector. The scanning order of the data points had to be adapted to be compatible with prior sessions. More importantly, because of the reduction of the distance camera/projector, the depth (along Z) precision is reduced, what appeared clearly visible in the data.

c) Qualitative evaluation

In order to improve the acquisition system and the acquisition procedure and to evaluate the face recognition potential of the 3D comparison approach, we analysed thoroughly the origins of acquisition errors. Major defects were identified and each of the 720 shots (2 sessions, 120 people, 3 shot each) was classified according to their defects. The following two tables give a summary of major problems, leading to important visual defects, affecting the 3D acquisition quality and compromising the correct extraction or the correct comparison of some of the profiles. For example, a moustache, a beard or the darkness of the eye usually impairs the visibility or detection of the projected stripes. The same holds with blurred images or when the person wears glasses, but in those cases, an appropriate action can be taken to get a correct image. The remaining problems relate to the difficulty to capture 3D by light projection in areas where the surface is not smooth (‘Chin’, ‘Nose’ or large ‘Rotation’), possibly with parts hidden to the camera or not receiving projected light.

3D acquisition defects	Beard	Blur	Cheek	Chin	Eye	Glasses	Moustache	Nose	Rotation	Rejections
Session1	6	3	3	2	3	3	2	2	2	**22**
Session2	5	1			2			3		**11**

Table 7: Number of 3D acquisition defects and people rejections for 3D acquisition

In Table 7 and Table 8, 'Rejections' gives the number of persons who could not participate to 3D comparison experiments because at least one of their 3D shots was judged inappropriate for comparison due to strong acquisition deficiencies. For session1, 22 people were rejected, corresponding to 26 images out of 360 captures. If we attribute inadequate images due to blurring or glasses as independent from the acquisition procedure, the acquisition success rate achieves 94 % of the captures.

Table 8 gives the origins and distribution of defects encountered during profile extraction. Here again, the figures give the number of people involved, as soon as one of their 3D shots had a problem. This table considers the shots that were not rejected during 3D acquisition. For instance, session1 has 98 people accepted for 3D comparison, but 19 were rejected due to wrong profile extraction, leading to 79 people available for comparison. The noise, referring here to incorrect 3D information due to erroneous stripe localisation or labelling, often prevents from extracting correct profiles. Wrong 3D surface capture in the nose region is the second main cause of problems.

Profile extraction	Moustache	Noise	Nose	Optimisation	Rotation	Rejections
Session1		10	6	3	2	**19**
Session2	1		1			**1**

Table 8: Number of 3D acquisition defects and rejections for profile extraction

As it appears from the difference between session1 and session2, the second session benefited from the experience of the first campaign in reducing wrong acquisitions. We also asked people to take their glass out (except for two persons).

d) Manual database

In order to estimate the influence of acquisition imperfections, a set of striped images were manually processed to avoid wrong stripe localisation or labelling. An interactive program was written. Clicking on a stripe (assigning at the same time its thick/thin property), the program tracks the stripe by following the local minimum in the privileged vertical direction. The user can specify the end of a stripe (which is not guessed by the program) or can cut out the ending when it does not follow the stripe. Partial stripe following is allowed.

At any time, the labelling process can be launched. For this, the tracked stripes with their thick/thin property are compared with the reference to find a match. In the case there is no match, the user knows some stripes were incorrectly assigned the thick/thin property or were incorrectly tracked.

At any time, the points along the tracked stripes and their label can be stored to a file. They will be later translated into xyz coordinates thanks to the calibration data.

Because this manual procedure is time consuming, we limited the supervised extraction to the first 30 alphabetically ordered people of the database, leading to 180 image manipulations for session1 and session2. The corresponding database is called the ***manual DB***. The part of the automatic database consisting of the same 30 people is called the ***auto DB***. Recognition experiments on 30 people of the ***auto DB*** gave similar results than the ones carried on the whole automatic database.

Following rejections proposed by the qualitative evaluation of subsection c), tests conducted with the manual database discarded three persons of session1 and one different person of session2.

e) Distribution

By reception of a signed agreement, the 3D_RMA database is made freely available for research purposes (refer to http://www.sic.rma.ac.be/~beumier/DB/3d_rma.html or Appendix G). The grey-level images are however not accessible as they represent sensitive information about life privacy of the individual who participated freely to the database collection.

Although it has been four years that the database is available and referenced in a publication, we only have since February 2003 an idea of the interest for the database: about one request per month, from all over the world. Indeed, the database is not publicly available from internet anymore since 2003. The signed agreement has become a mandatory condition to access the database from our ftp site.

2.8.3 The BIOMET database (Prototype B)

a) The BIOMET project

The BIOMET project [BIOMET], running in the years 2001 and 2002 aimed at bringing together the competences of the French GET schools active in the field of security through person identification from biometry with modalities such as signature, face, fingerprint, hand shape and speech. More precisely, the BIOMET objectives were:

- To build up a multi-modal database (of about 100 persons);
- To implement verification systems for each modality;
- To search and analyse one or several fusion strategies;
- To let distribute the database (ELDA: European Language Distribution Agency).

During the project, three acquisition campaigns were carried out at ENST, Paris, separated by three and five months. 81 subjects were present in all three sessions (50% male), with age ranging from 20 to 65 years. Each subject spent 20 minutes:

- To present his/her left hand to a A4 scanner;
- To draw genuine and impostor signatures (dynamic);
- To put her/his middle and index fingers onto a capacitive and an optical fingerprint devices;
- To present his/her face in front of a proprietary (INT, Paris) infra-red camera;
- To pronounce phrases and rotate her/his head while being captured by a digital camera (image sequence and speech);
- To stand in front of the 3D acquisition system for a few shots.

Details about the protocol and data collected are to be found in [Salicetti03].

We participated with 3D acquisition to the second and third campaigns.

For the second campaign, we used the enhanced prototype A, consisting of the black and white pattern coded in thickness, the canon G2 camera and the flash. Unfortunately, the canon G2 triggers the flash remotely (cable) without duration control, and focusing was randomly good, resulting in many blurred images. No 3D reconstruction has been carried out for that campaign.

Prototype B was developed in order to be less sensitive to focus by introducing larger, colour stripes. It was also the opportunity to address texture and 3D acquisition from the same image. Illumination was this time again not perfect, from the impossibility for the camera to adapt the flash intensity in remote connection. Nevertheless, we adapted the algorithm to detect the stripe centres (and not the edges as intended).

b) 3D shots

The third campaign of the BIOMET database captured 81 persons. For 3D captures, people sat on a chair and were asked to gaze in six different directions (frontal, left, right, up, down and frontal), although limiting the deviation from a frontal pose. Six shots are normally available for each person.

The quality of the 3D data is not as high as expected. The light projection was not uniformly distributed, implying contrast and colour weaknesses in some areas. The algorithm was adapted to localise the stripe centres, more visible than the edges, resulting in half the intended density of 3D points across the stripes.

Figure 38 shows a good quality striped image captured with prototype B with two 3D representations (frontal, profile) of the extracted 3D data.

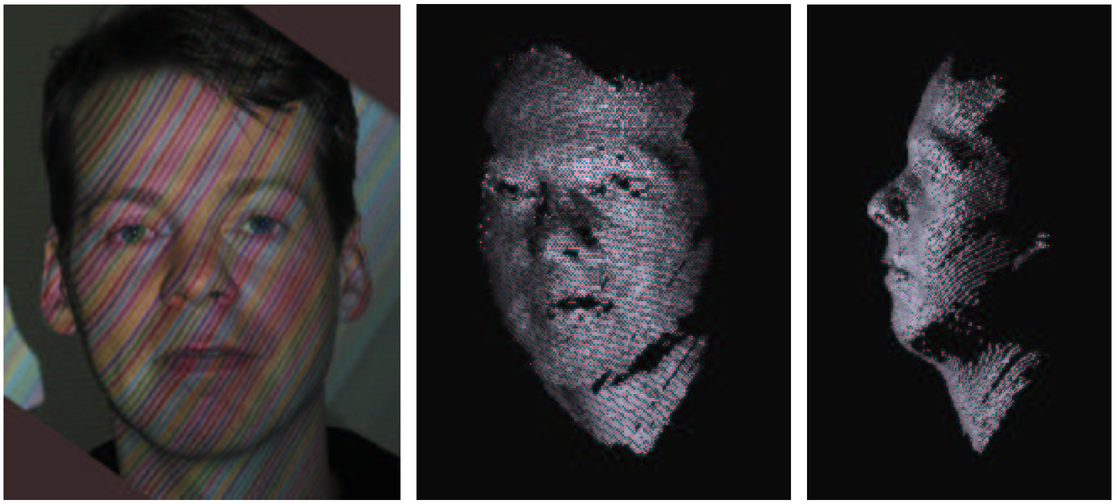


Figure 38: Striped image and 3D representations from extracted 3D data

Figure 39, left, shows the texture image obtained from the striped image of Figure 38, left. This colour information was mapped (Figure 39, right) on the frontal 3D representation of Figure 38.

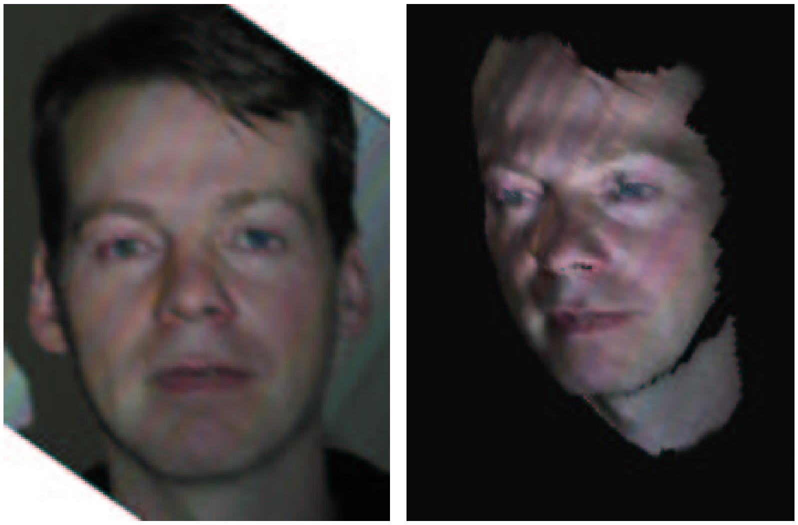


Figure 39: Texture extracted from Figure 38 and mapped onto the frontal 3D representation

c) Qualitative evaluation

The quality of the 3D shots of the third campaign has been evaluated by assigning all the reconstructed 3D representation into one of the following classes:

- Reject: the shot suffers from severe capture problems and is clearly inappropriate for recognition experiments;
- Fair: one or several problems may affect the surface but the shot is still valuable for recognition experiments;
- Good: no major defect impairs the visual quality or the possibility to use the shot for recognition.

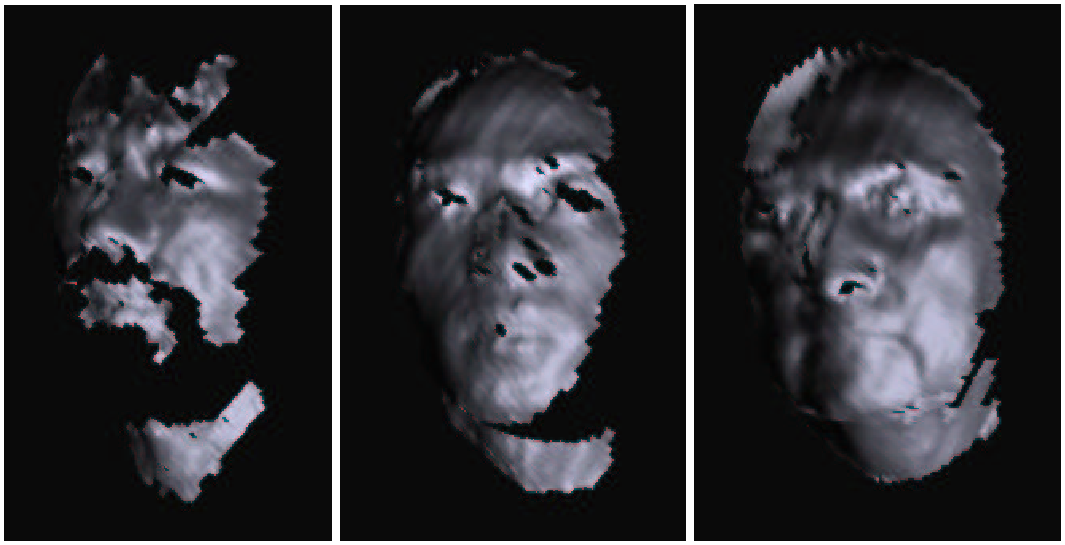


Figure 40: Examples of Reject, Fair and Good quality surfaces

The distribution of people between the three classes is given in Table 9.

Quality	Reject	Fair	Good
Number of persons	15	34	32
Percentage	18.5 %	42 %	39.5 %

Table 9: Distribution of people among the three quality classes of their 3D faces

Out of the 81 persons represented by five or six 3D shots in the database, 15 persons are subject to major capture deficiencies such as the presence of the beard and/or moustache (4), the wrong capture of the nose (2), the large rotation of many shots (4) and the limited contrast leading to inaccurate stripe detection (4).

The general conclusions concerning the quality of this 3D campaign is that the projected light was not carefully controlled by the camera, due to the flash, and that many images suffered from low contrast and weak colour definition. The influence on stripe labelling and texture extraction was limited but stripe localization was less accurate. On the one hand, the intended stripe edge detection had to be replaced by the stripe centre localization, reducing the surface point density by two. On the other hand, the actual stripe localization was less precise from the lack of contrast and introduced visible noise in reconstructed surfaces. Most people with bright skin were correctly captured.

d) Distribution

The distribution of the BIOMET database is still under discussion about legal issues concerning the life privacy of the participants. The database indeed contains very sensitive data such as fingerprints, signatures and face images. A permanent effort has however been undertaken to provide for the eventual distribution of the database, once the validation and evaluation phases have been completed. The distribution will be carried out by an appropriate organism that will take care about administrative and practical matters (ELDA: European Language Distribution Agency).

2.9 Conclusions

We presented the evolution of the ***developments*** we carried out for ***3D capture***, detailing the three realised prototypes. After justifying the structured light approach on the premise of cost, speed and hardware availability, we described the different elements constituting the prototypes. The specificity of a structured light system lies in the projected pattern. We proposed a triple criterion for pattern evaluation (localisation, labelling and texture). After those design considerations, calibration and 3D extraction were presented for the three prototypes.

As far as the ***light projection*** is concerned, and considering that a video projector is more appropriate for research purposes due to its bulkiness, light for projection is better produced by a ***flash lamp***. The important light power delivered by a flash lamp over a short period, allows for a very good contrast, even for small aperture and short duration, leading to large depth of focus and no motion blur. Flash lamps are compact and do not require cooling. The drawback of using a flash is the long latency time needed to recharge the system. Independently of the light source, and as long as a monochromatic pattern filters light, infra red projection is possible, ensuring discreteness and avoiding dazzling the user.

In light of the different ***projection patterns*** that were designed and analysed under the triple criterion (localisation, labelling and texture), ***parallel lines*** (stripes or slits) appear to be preferable. They are simple and continuous, leading to a high precision of the detected points. Colour can be an advantage for labelling and texture recovery, but we finally decided on black and white striping which ensures a better detection and precision and better quality of the consecutive processing (stripe labelling and texture compensation). In the specific case of stripe labelling by dots (prototype C), the pattern has 2D labelling capabilities (dot vertical position), allowing projector distortion compensation, with what is basically a 1-D stripe pattern, easy for localisation.

Thanks to higher quality at increasingly lower cost, ***digital cameras*** perfectly suit to high quality 3D capture. The current prototype still suffers from the low image transfer rate (about 10 sec) through the USB connector. We are waiting for the next generation of camera connectors that are already available. For a prototype including a flash lamp, the camera must be able to pilot an external flash for proper synchronisation and controlled lightening, which is not so common.

The work of ***calibration*** aims at ensuring the fidelity of the 3D estimation. This way, 3D surfaces can be compared from acquisition to acquisition or with other acquisition systems. However, for face comparison with a given prototype, 3D precision is not so critical in comparison with other factors which impact quality like stripe miss or labelling error. Also, precision errors affect 3D descriptions in a similar way for faces always presented in similar orientation, position and distance.

The calibration procedure is rather long (15 minutes) if we must deal with point selection but calibration is rarely needed and could be optimised. In particular, some parameters remain the same from calibration to calibration.

The sensitive part of the acquisition system based on structured light is the ***extraction and labelling*** of the stripes. The proper extraction of the stripes, based on their visibility, depends on:

- the quality of the facial surface. Dark skins and the presence of dark facial hair, bushy beard or moustache impair the success and quality of the extraction.
- The quality of the projected light. The advantage is given to the projection of a black and white pattern thanks to the stronger contrast. Colour stripe projection is appropriate for a compact labelling with texture recovery, but suffers from a limited contrast, especially for yellow and cyan stripes on skin.
- The visibility of the surface. Regions that are not reached by the projected light or not visible by the camera cannot be captured. The nose and the throat are typical regions that partly suffer from visibility. This is the reason why frontal poses are better, leading to an equilibrated distribution of the information. The facial coverage of the capture in one shot is acceptable for a frontal pose.

Labelling the stripes is necessary but not crucial for a correct capture of the face thanks to the large labelling redundancy and geometrical continuity of the stripes.

The ***qualitative analysis*** of the acquired ***databases*** showed the general good quality and potential of 3D face capture with structured light but also revealed the two types of problems related to stripe extraction. For a good quality projection and image capture, obtained through the control of the capture, typical errors are either due to the point of view (reducing the visibility of some facial parts) or the person’s characteristics (skin colour, presence of beard or spectacles). User cooperation helps in capturing 3D faces of high quality.

Texture recovery has also been addressed in the developed prototypes. The objective was to capture from the same shot the 3D and the intensity/colour information, in perfect correspondence. For face recognition, the texture quality does not have to be very high. A simple compensation scheme has been applied successfully to the striped images.

Practically, we paid a special attention to spatial precision (pattern localisation and 3D estimation) and resolution (density of extracted 3D points) with a system based on a simple light projector, avoiding the expensive video projector. Prototype B is compact, low cost and could capture a 3D facial surface in a couple of seconds if the image download from the camera was quicker. Prototype C is expected to deliver higher quality 3D surfaces with the same practical features. Both prototypes exceed the requirements of face capture for recognition followed during the first developments and are capable of capturing other 3D scenes.

Chapter 3 3D face recognition

3.1 Introduction

This chapter presents the approaches we followed to achieve recognition of 3D facial surfaces. Data and results mainly concern the 3D_RMA database (see subsection 2.8.2) captured with Prototype A. Only section 3.7 is devoted to the partial experiments carried out with the BIOMET 3D database (Prototype B).

The retained approach consists in ***matching the facial surfaces*** to be compared by adapting rotation and translation parameters. The 3D faces are captured in real dimensions so that no scale must be adapted. We consider the human face as a rigid body, which is acceptable for the cooperative scenario with neutral expression. The quality of the match is measured thanks to the average distance (with outliers rejection) separating corresponding geometrical profiles extracted by planar cuts of the facial surfaces. In a first experiment, the planar cuts consist in the intersections of parallel planes with the facial surface. In a second experiment, the adopted technique consists in taking advantage of the intrinsic symmetry of the face to extract central and lateral profiles automatically.

Texture comparison of the facial surface has been smoothly integrated into the second geometrical comparison approach thanks to the availability of grey-level information. Grey-level values are collected along the extracted central and lateral profiles and compensated for to be less sensitive to ambient illumination.

The ***combination*** of the geometrical and texture comparison scores is then considered to increase the recognition performance.

Results about recognition experiments are given in the related sections and are finally discussed in the light of the possible sources of error in section 3.6.

Experiments with the ***BIOMET 3D database*** are reported in section 3.7 to judge on the quality of the acquired data, in spite of the deficiencies affecting the BIOMET 3D campaign.

3.2 Test protocol

Tests and results are presented along this chapter. As already mentioned, they mainly concern prototype A and the related 3D_RMA database. Experiments with Prototype B (BIOMET) are localised in section 3.7 and have their own test protocol. The objective of recognition tests is to guess the potential discriminative power of the 3D face recognition approach and to highlight the possible deficiencies of one or another component of the recognition chain.

Only sessions 1 and 2 of the 3D_RMA database (section 2.8.2) have been used, each providing three shots for 3D description and one grey-level shot for 120 persons.

In order to reduce the errors due to acquisition and comparison problems impairing the recognition performances but not directly related to the face recognition potential, we considered two partial databases.

The first one, called ***auto DB,*** consists of the first 30 alphabetically ordered people, rejection made of three individuals in session 1, and one in session 2, due to strong acquisition deficiencies. A few comparison experiments were made to analyse the possible difference in recognition rates when passing from 120 to 30 people. Similar recognition rates were obtained.

The second database, called ***manual DB***, consists of the same population (27 persons in session1 and 29 in session2, 26 persons common in both), but with supervised 3D extraction to reduce acquisition errors, as explained in section 2.8.2d). The important effort required for manual processing prevented us to envisage a larger manual database.

Tests were carried out from comparisons intra and extra sessions. Intra session tests (session1 or session2) should reveal the influence of acquisition problems and the stability of the approach relatively to posture. Extra sessions tests (session1-session2) should show the possible stability of the comparison results over time. The results are presented either in the form of ROC curves or more synthetically with Equal Error Rates (see section 1.3.2).

3.3 3D Recognition

3.3.1 Introduction

Approaches for 3D recognition are either based on feature extraction and classification or on primitive comparison.

For ***feature extraction*** approaches, a set of properties based on distance, angle or curvature are derived from the 3D data and compared with the reference set of features. These features are selected for their discriminative potential between objects, their stability relative to a given object and their adequacy for software implementation.

In ***primitive comparison*** approaches, parts or whole of the surface is matched with the corresponding parts or whole of a reference surface. The matching procedure must consider translation, rotation and scale to account for the point of view dependence.

In both cases, the comparison issues a match distance on which the decision about acceptance or rejection of similarity will be based.

In the present field of application, discrimination is complicated by the similarity of faces. They all contain the same basic components (eye, nose, mouth, ...). Differences are to be found in component details (shape, size, colour) and configuration (distances and angles between components).

More specifically, methods for 3D object recognition consider 3D points, 3D contours, 3D surfaces or 3D volumetric primitives. A comparison is established and examples are referred to in [Dickinson93]. The advantage is given to more elaborated descriptions thanks to their reduced model and search complexity and their higher flexibility. One may also combine different approaches (see [Stein93] about fusing edges and surface descriptions to recognise 3D objects).

For ***3D facial analysis***, we first considered ***3D points localisation***, taking advantage of the very clear positioning of the nose tip and the nasion point (nose saddle between the eyes). However, other candidates are difficult to find as the facial surface mainly exhibits smooth patches with no reference point. Other interesting regions could be the mouth and the eyes but 3D data is subject to noise in these areas due to texture and motion.

3D contours have not been retained although the nose ridge is a very good candidate thanks to its visibility and ease of localisation. The jaw line delineation depends on the visibility of the throat. The eyebrows could not be localised due to the lack of resolution and precision of our acquisition system in those textured areas. A special case of use of contours concerns the work described in [Cartoux89] which extracts the central profile from facial symmetry to make comparison.

Although faces are similar and thus compatible with a common model, isolated ***3D primitives*** were not considered due to the noise of the data and the lack of clear delimitations of components. A global model could be used, but to get the small differences between faces, it should contain complicated descriptors with many parameters.

Some works related to 3D face recognition have considered ***transforms*** such as the ”Principal Component Analysis” (3D eigenfaces) [Achermann97], and surface curvature [Lee90, Gordon91]. The success of the approaches, as specified by the authors, however relied on the high quality of the 3D sensors used. We did not use curvatures due to important noise in the data.

We preferred to consider ***global 3D surface matching*** to compare faces. Indeed, extracting components would be difficult due to the imprecise delimitation, and matching would require a global consistency of the individual matches. To solve the correspondence problem between the surfaces to be compared, we used a set of parallel planes intersecting the face along profiles. The global 3D surface comparison is carried out by the comparison of those individual 2-D profiles.

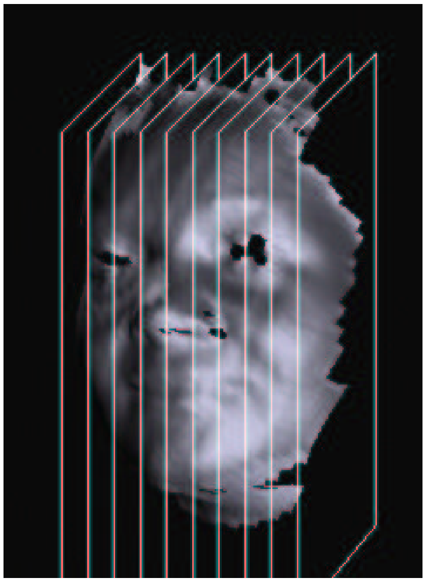


Figure 41: Intersecting the facial surface with parallel planes

Before detailing the surface matching approach, topic of section 3.3.4, the next two sections present preliminary studies about 3D face analysis. The first study is an attempt to directly analyse the striped images, without explicit 3D extraction, while the second study describes our experiments with 3D facial feature extraction. Section 3.3.5, about central and lateral profile comparison, is a particular implementation of surface matching, taking advantage of the symmetry of faces.

3.3.2 Analysis from striped images

One interesting 3D comparison approach is to get information directly from the 2D (striped) images in order to avoid or postpone the time consuming 3D conversion. Although studies were carried out in that direction by some researchers [Wang87, Chen93], coping with the influence of the viewpoint on the shape of the stripes seemed too difficult. Only the prominence of the nose enabled its localisation. However, nose localisation is crucial to detect faces. It allows concentrating the efforts (such as 3D conversion) around it.

Because the nose is the more prominent part of the face, it is easily detected by highlighting the points of each stripe closer to the camera. We can also look for the maxima of curvature, which are much larger for the nose than for other parts of the face.

Figure 42 shows the localisation of the leftmost point of each stripe, corresponding to the points closest to the camera. We can see that the nose is well and precisely localised. Taking into account the variation of stripe inter-distance with the surface orientation, we were able to localise the nose tip and the nasion point (black crosses on the image).

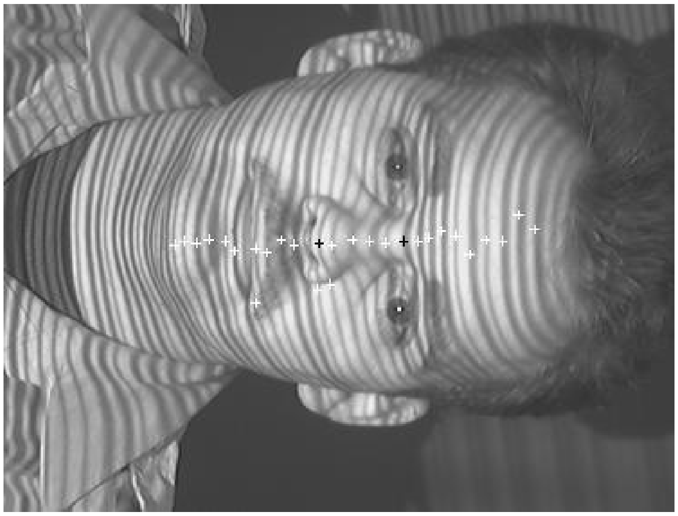


Figure 42: Nose localisation from depth

However, in our particular setting using diagonal striping, the point closest to the camera for each stripe is not always on the central profile. For instance, the nose ridge is not well positioned. See Figure 43.

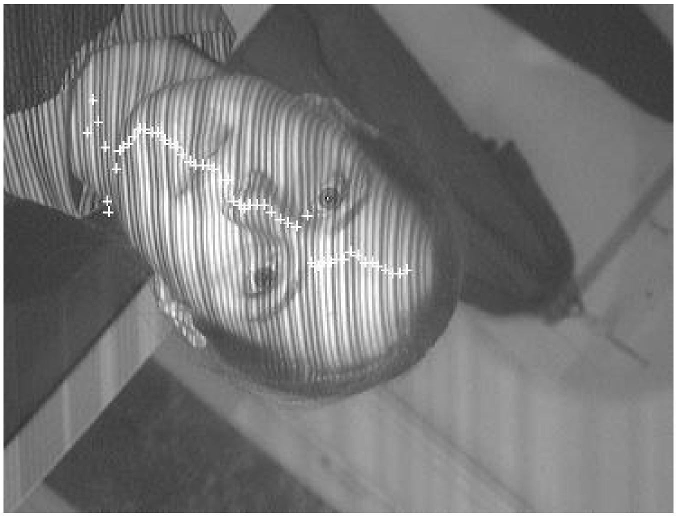


Figure 43: Nose localisation from depth in case of diagonal striping

3.3.3 Analysis from features

Work has been carried out in looking for discriminative (different among people) and reproducible (stable for a given person) features. The principal objective was to reduce the 3D data to a set of features easily and quickly compared.

We estimated the prominence of the nose relative to points of the cheeks located at a given distance (e.g. 7.5 cm) from the nose tip. This led to stable values for each person (variations less than 1 mm) with a span of more than 4 mm among 10 individuals considered in a preliminary study, before the 3D_RMA was captured (see Table 10).

Name	Prominence (mm)	Length (mm)
Alplum00	49.6	42.6
Alplum01	49.2	40.8
Alplum02	50.1	41.8
Basche00	53.5	52.5
Basche01	53.2	50.4
Basche02	53.5	53.6
Chbeum00	52.1	46.9
Chbeum01	51.9	42.8
Chbeum02	51.8	48.7
Dideme00	52.6	47.5
Dideme01	53.5	46.0
Dideme02	52.6	50.3
Jmmang00	53.3	45.7
Jmmang01	53.7	47.6
Jmmang02	53.8	45.9
Maache00	52.6	!34.5!
Maache01	52.9	51.2
Maache02	53.3	49.5
Marass00	50.5	45.3

Marass01	51.4	48.1
Marass02	51.2	44.2

Table 10: 3D Feature values for the nose

The nose length was also measured by localising the nose tip and the nose saddle (between the eyes). Look at the black crosses on Figure 42. Although this measure was less precise, it brought information thanks to the large variability of the nose length among individuals.

However, the nose seems to be the only facial part providing robust geometrical features for limited effort. Mouths and eyes are not rigid and may involve acquisition problems. Foreheads and chins, interesting rigid parts, do not clearly exhibit reference points for normalisation. We thus abandoned feature extraction and considered the global matching of the facial surface.

3.3.4 Global surface matching

a) Overview

Global 3D surface matching is carried out by first extracting parallel profiles from the two surfaces to be compared. Corresponding 2-D profiles are compared two by two to issue the mean distance that separates them. Outlier values due to spurious points or incompatible profiles are discarded. The average of the distances for all the profile pairs is the global error that has to be minimised. The minimisation is achieved by tuning the 6 parameters of orientation and translation since the 3D representations are normally correctly scaled. We treat the face as a rigid body, which is an acceptable assumption for cooperative people with neutral expression.

b) 2-D profiles

We first used parallel planes (Figure 41), separated by 1 cm, to extract vertical 2D profiles as shown in Figure 44. A set of maximum 15 planes is used around the nose tip localised as the most prominent point.

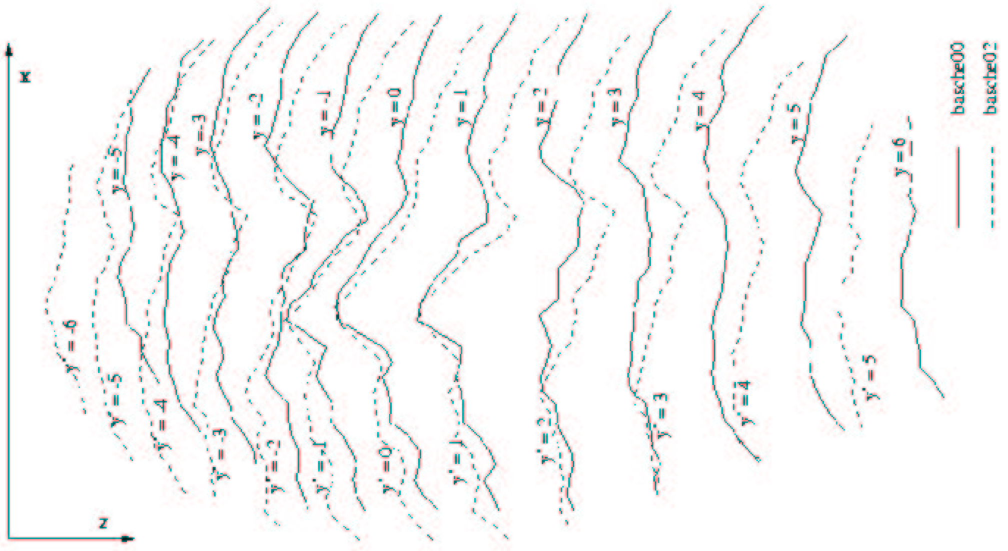


Figure 44: Profiles from two 3D representations with noses already in correspondence

Each 2D profile is obtained as the intersection of the corresponding plane with all the stripes. A profile consists at the maximum in as many points as there are stripes. In the conducted experiments related to the acquired databases, this amounts to around 50 points.

The implementation took advantage of the stripe oriented format of 3D data (the 3D points are stored as appearing along the stripes, stripe after stripe). Due to the stripe inclination on the face and the rather vertical direction of the planes, each plane normally intersects each stripe only once. Therefore, the intersection is quickly obtained by applying successive approximation based on the signed distance to the plane.

c) 3D Distance

The distance between 2-D profile pairs is measured as the area subtended between the two profiles to be compared divided by the arc length. The 3D distance is obtained as the average of the individual profile distances that are not too large.

The area between the curves is estimated by summing the area of triangles consisting of two successive points of one curve and one point of the other curve (1, 2, 1' and 1', 2', 2 in Figure 45). If the two successive points of a curve are too far from each other, an acquisition problem is suspected and the triangle is not considered.

The surface of each triangle is advantageously computed by a vector product (half the value of a⊗b or c⊗d in Figure 45, '⊗' being the vector product). This involves only subtractions, multiplications and additions (division by 2 can be done at the end, once for all) and the result has a sign depending on the orientation of the vectors.

When the curves cross each other, the vector products change sign. Near the cross, a⊗b and c⊗d have opposite sign so that their sum discards the excessive area patterned by a grid on Figure 45. The area computed by the sum is in that case underestimated but the contribution of this error to the whole profile distance is expected to be small. The double of the total area has been computed as the sum of the absolute values of a⊗b and c⊗d when they have same sign and a⊗b + c⊗d when they have opposite sign. Visual results of matches confirmed the approach.

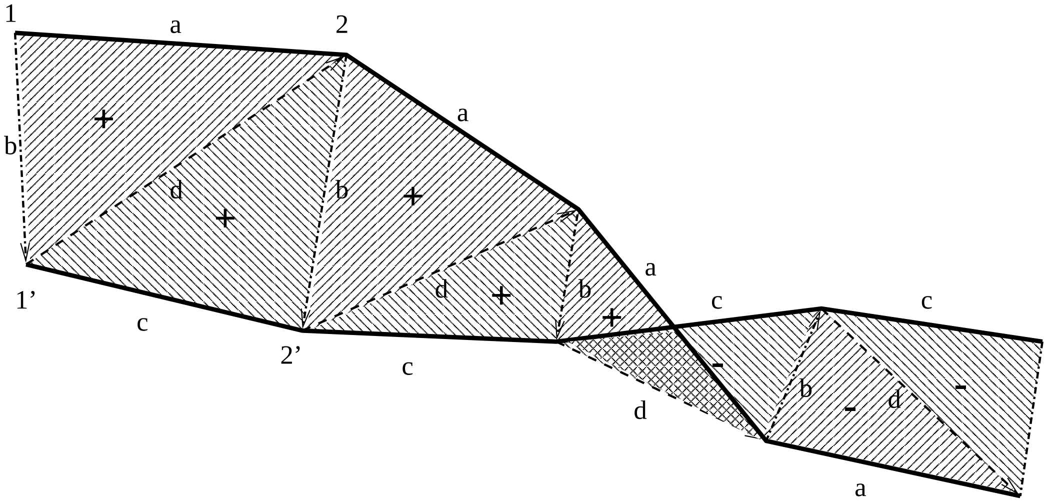


Figure 45: Surface estimation between two curves

d) Surface matching

Minimisation was first achieved by an automatic initial guess of the three translation parameters (thanks to the nose tip: see Figure 44) and the manual tuning of the 6 parameters (3 rotations and 3 translations) thanks to visual appreciation (see Figure 46). We successfully automatised the procedure by an Iterative Conditional Mode optimisation, which optimises each parameter, one after the other. The optimisation was organised in cycles of ICM, separated by a reduction of the span of each parameter. Although efficient, this procedure is slow (5 seconds on a Pentium 200 MHz), because the profiles have to be extracted for most of the parameter changes.

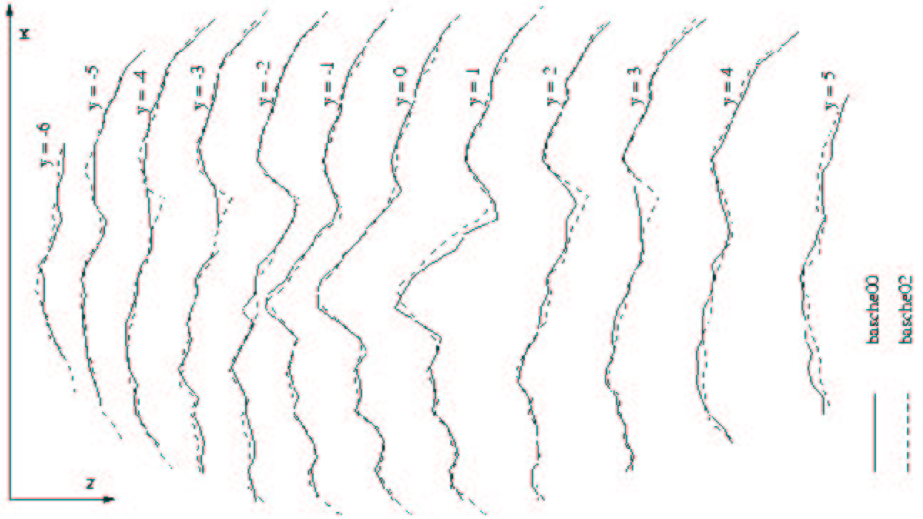


Figure 46: Profiles of the representations after surface matching

The method presented so far gave good results, but the automatic implementation often fell in local minima (compare the manual and automatic curves of Figure 47). Initial guesses of left/right and up/down rotations (based on the nose, cheeks and forehead) reduced bad optimisation due to local minima and led to a quicker solution by earlier confinement of the search space. An average speedup of 6 was achieved, what brought the comparison of two faces just below 1 second. These last automatic results, called *Improved Auto Optim* on Figure 47, are now closer to manual refinement.

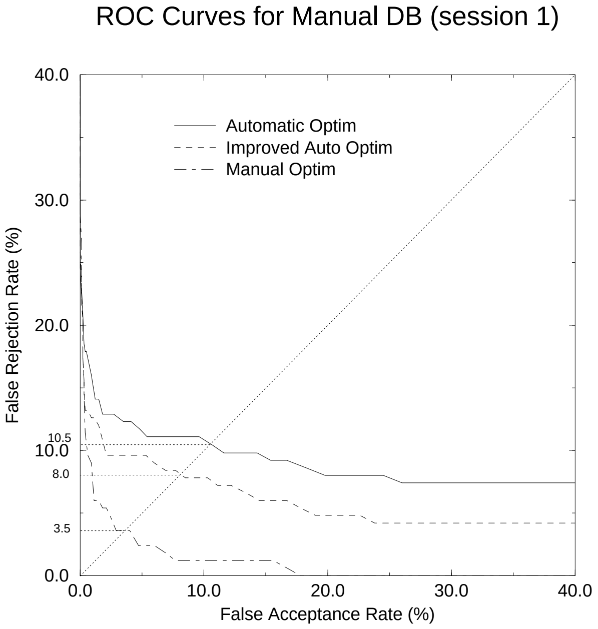


Figure 47: ROC curves of 3D surface matching for the manual database

Other tests of the improved automatic optimisation sometimes exhibited a smaller improvement because noise affected the rotation guesses, what may prevent a correct match due to the reduced search space.

e) Results

Table 11 lists the results in Equal Error Rate of the global surface matching approach for comparison of 3D representations within session 1, within session 2, and between sessions 1 and 2. Based on 27 persons in session1 and 29 in session2, 3 shots per person, intra session comparisons contain 3x26 = 78 client and (27x29-26)x3 = 2271 impostor tests. Inter session comparisons (session1-2) contain 234 client and 6813 impostor tests.

Several conclusions arise from this table. First, the gain of supervising the 3D acquisition ("***manual DB***") is not very important. This can be explained by the fact that supervision cannot solve major problems related to structured light acquisition such as bushy beard or glasses. Secondly, the difference between automatic and refined surface comparison is rather important, showing that many automatic comparisons were stuck in local minima. Finally, the rightmost column clearly shows a reduction in the performance due to the comparison of data acquired at different times. By careful inspection, we noticed that the second session was less upsetting, especially for young students, and many serious faces in session 1 are replaced by smiling faces in session 2.

	Session1	**Session2**	**Session1-2**
Auto DB, auto	9.0 %	9.0 %	13.0 %
Auto DB, refined	4.5 %	3.25 %	6.0 %
Manual DB, auto	8.0 %	7.0 %	9.5 %
Manual DB, refined	3.5 %	2.0 %	4.75 %

Table 11: EER of surface matching for the *auto DB* and *manual DB*, with (*refined*) and without (*auto*) manual refinement

In order to reduce the problem of local minima, we tried to normalise some of the parameters thanks to the facial symmetry of faces, as explained in the next section.

3.3.5 Central and lateral profiles

a) Central profiles

From the experience we gained in the field of profile recognition ([Beumier97]), the most practical solution to evaluate the quality of a 3D facial acquisition system is to analyse the central profile (vertical and passing through the nose). It is quickly extracted, directly appreciated by visual inspection, and easily compared by 2-Dimensional curve matching.

We started to extract manually a few central profiles to roughly estimate the quality of the 3D acquisition system. Although typical problems were present (nose broken, disturbances in beards or moustaches), the profile, acquired at the real scale, was worth being compared. We automated central profile extraction by looking for the profile with maximal protrusion (due to the nose) and which best symmetries left and right profiles extracted by planes parallel with the central profile and 3 cm far from it. This optimum search is quick as it only depends on three parameters (one translation and two rotations).

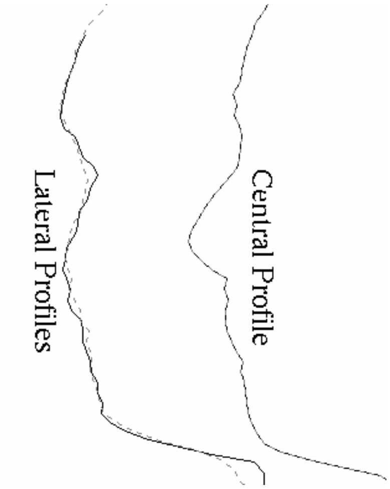


Figure 48: Central and Lateral profiles after intrinsic normalisation

The automatically extracted central profiles were compared in the angle space, by transforming each 2-Dimensional profile into the 1-Dimensional local slope values along this profile for two running points separated by 4 cm (Figure 49). A rotation of the profiles corresponds in the angle space to an offset of the slope values, so that measuring the standard deviation of slope differences is independent from rotation. Dealing with only one shift parameter (for instance to bring the noses of two profiles in correspondence), the 3-parameter search to match two profiles is replaced by a 1-Dimensional minimum search in the angle space.

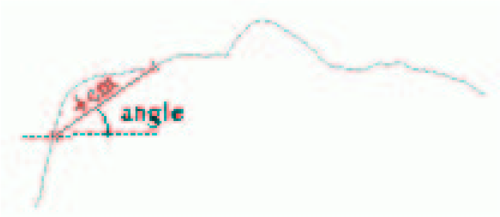


Figure 49: slope measurement along the profile

To summarise, the 6-Dimensional (3 rotations, 3 translations) optimisation problem is decomposed into a 3-Dimensional intrinsic normalisation based on the head symmetry and a 1-Dimensional profile comparison based on local slope comparison.

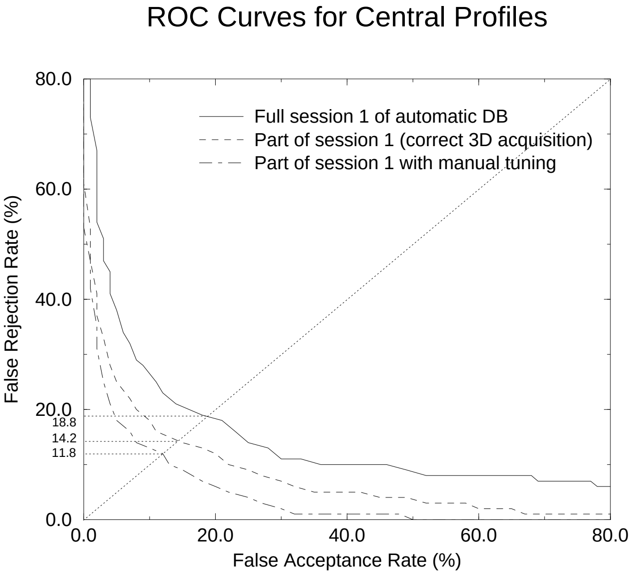


Figure 50: ROC curves of comparison of central profiles extracted automatically on the full *auto DB (120 persons)*, on the set of correct 3D representations (*79 persons*), and with manual tuning on this set

Figure 50 shows the results as ROC curves for the comparison of central profiles extracted automatically from the 3D representations of session 1 of the full (*automatic) DB*. The poor recognition rate (EER: 18.8 %) is partly explained by bad 3D representations of the (*automatic) DB* due to glasses or bushy beards. The same figure shows the results on a reduced database, including only people with profiles of sufficient visual quality (79 persons were kept from session 1) (EER: 14.2 %). Finally, manual tuning of some profiles that were clearly not well extracted, brought the recognition rate (EER: 11.8 %) to a performance similar to experiments carried out in our laboratory about 2D profile recognition from a single view of reference. More specifically, using the colour images of profile view of the same people (SIC_DB, see 2.8.2), and applying the algorithm for profile identification described in [Beumier97d] to compare session1 and session2 profile images, we achieved an EER of 10.5 %.

b) Lateral profiles

To include more 3D information, we analysed the lateral profiles (left and right profiles at 3 cm from the central profile) that were used to automatise the central profile extraction procedure thanks to the vertical symmetry assumption of the face. To bring robustness, these left and right curves were averaged to offer a mean lateral profile. The mean lateral profiles were compared, based on the local slope comparison described above. As seen on Figure 51, the lateral profile discrimination power (EER: 10.0 %) is similar to what was obtained with the central profile. Although more specific to individuals, the central profile often suffers from acquisition deficiencies in the nose region.

c) Combining central and lateral profiles

The fusion of comparison scores for central and lateral profiles by a simple average brought a clear advantage (EER: 6.2 %) as shown in Figure 51.

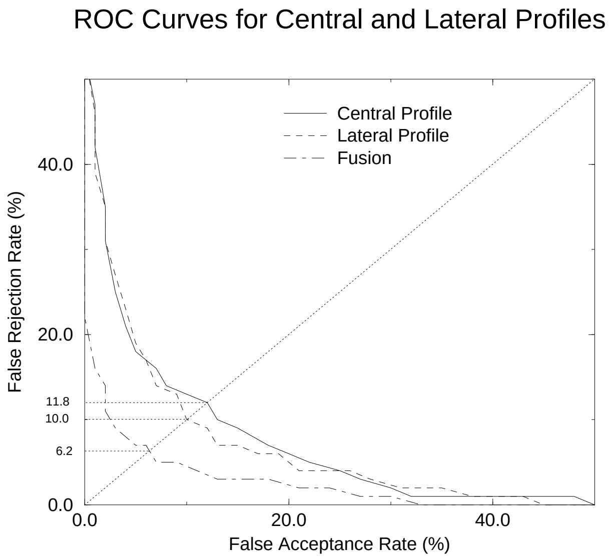


Figure 51: ROC curves for central profile comparison, lateral profile comparison and for the fusion of central and lateral profile comparison, all with manual tuning (session1, 79 persons)

Tests were then applied to session 2, where 108 persons were kept out of the 120 people, according to 2.8.2c). Similar figures were obtained. But comparing session 2 profiles relative to session 1 profiles revealed a clear reduction in the performances as shown in Figure 52, and attributed to clear changes of mouth and cheek shapes between the two sessions, mainly due to smiles.

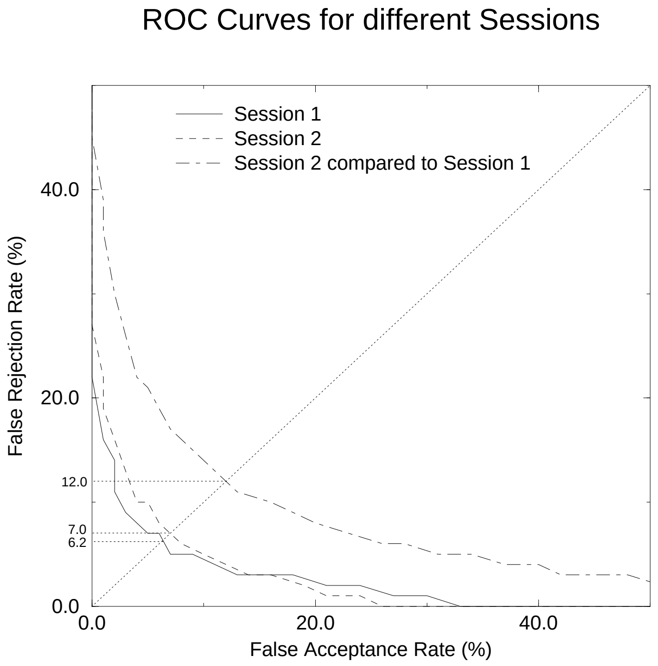


Figure 52: ROC curves for fusion of central and lateral profile comparison (with manual tuning), for session 1 (79 persons), session 2 (108 persons) and session 2 compared to session 1.

d) Results

With the same objective of trying to identify the influences of acquisition and optimisation errors, the central and lateral profiles approach was tested for the two database sessions, considering the *auto DB* and *manual DB*, with and without manual refinement during comparison. The same number of client and impostor tests as for 3.3.4e) are involved. Results are summarised in Table 12.

	Session1	**Session2**	**Session1-2**
Auto DB, auto	7.25 %	7.75 %	9.0 %
Auto DB, refined	6.25 %	7.0 %	9.5 %
Manual DB, auto	4.75 %	6.75 %	7.25 %
Manual DB, refined	2.25 %	3.75 %	6.75 %

Table 12: EER of the central and lateral profiles method for the *auto DB* and *manual DB*, with (*refined*) and without (*auto*) manual refinement

Results are here more homogeneous among sessions. The important role of the central profile (to detect the central and lateral profiles) makes the recognition chain (acquisition and comparison) probably more sensitive to correct 3D acquisition around the nose and chin. Typical problems due to the nostrils, mouth or throat will hardly be compensated by supervised acquisition (auto DB) or supervised comparison (refined). It seems however that when acquisition and comparison are both supervised, the gain is effective and the figure for intra session (session1-2) comparison highlights again the natural 3D differences between session1 and session2.

3.3.6 Comparison of the methods

Global surface matching seems to suffer from local minima, as it can be seen from the differences between ***auto*** and ***refined*** rows of Table 11. On the contrary, Table 12 shows the importance of acquisition errors on the performance of the central/lateral profiles approach (differences between ***auto DB*** and ***manual DB*** rows). Comparing those two tables, global surface matching (4.75 %) has a higher discrimination potential (when 'refined') than central/lateral profiles (6.75 %).

The main advantages of the central/lateral profiles method are its speed and low storage requirements. The intrinsic normalisation based on vertical facial symmetry takes 0.5 second (Pentium 200 MHz) and delivers two profiles of a few hundred bytes. These can be matched in 1 ms with each reference profile of the database (extracted off-line).

The global surface matching algorithm takes a mean time of 0.8 second to compare two 3D representations. This precludes decision methods based on ranking the whole database or performing sequence analysis. 3D representations are about 25 Kbyte large.

For verification applications, where only one person of the database is to be checked, both methods satisfy a time constraint of three seconds and only consume a few hundred Kbyte during execution.

To conclude on the 3D face analysis approaches considered in this thesis, and referring to the tests, the global surface matching approach integrates more information in a consistent way, providing robustness against 3D acquisition errors. On the contrary, the central and lateral profiles approach takes advantage of the normalisation based on the head symmetry to compact the information and offer a rapid solution for matching.

Due to the available computer power by the time of development (1998), and targeting real-time implementation, the preference was given to the central and lateral profiles approach. This approach also allows for a convenient inclusion of grey-level analysis, as described in the next section.

3.4 Grey-level comparison

3.4.1 Motivations

A grey-level analysis complements well the 3D processing steps performed so far.

First, the 3D acquisition by structured light is sensitive to the texture information of the underlying surface, typically introducing disturbances in eyes, nostril, mouth and facial hairs. A possible improvement is to avoid 3D extraction in such regions.

Secondly, 3D matching has been carried out with the sole 3D geometrical information. Noise or local minima may prevent from reaching the optimal solution. The inclusion of intensity clues during matching could speed up comparison, reduce local minima by giving more accurate initial conditions and increase the quality of the match by combining grey and 3D comparison.

Finally, facial 3D and texture clues are likely to be weakly correlated, so that their combination should enhance performances. 3D matching mainly concerns areas of the forehead, cheeks and chin, where grey information is weak. On the contrary, grey-level features are related to parts where 3D sensing is difficult or inaccurate. A grey-level analysis can also incorporate facial hairs localisation, and skin, eye or facial hairs colour.

We considered in this thesis the inclusion of grey-level clues to improve the recognition performances, by an a posteriori combination with 3D comparison scores.

3.4.2 Grey-level measurement

One way to get intensity values in registration with 3D data is to read between the stripes of the striped image. This requires no extra image acquisition or storage and 3D are in perfect alignment with grey-level values. Proesmans [Proesmans96] proposed a nice method to do this, although our projected stripes (prototype A) are too thick to obtain results of a similar quality.

Nevertheless, striped images involve important light reflections due to the directionality of the projected beam. Also, the stripes largely influence grey levels in their neighbourhood. Before trying to compensate for these effects, we preferred to compare grey-level images obtained by switching the projector off (Figure 53). It seems that ambient light was sufficiently isotropic in grey-level images to neglect corrections for reflections.

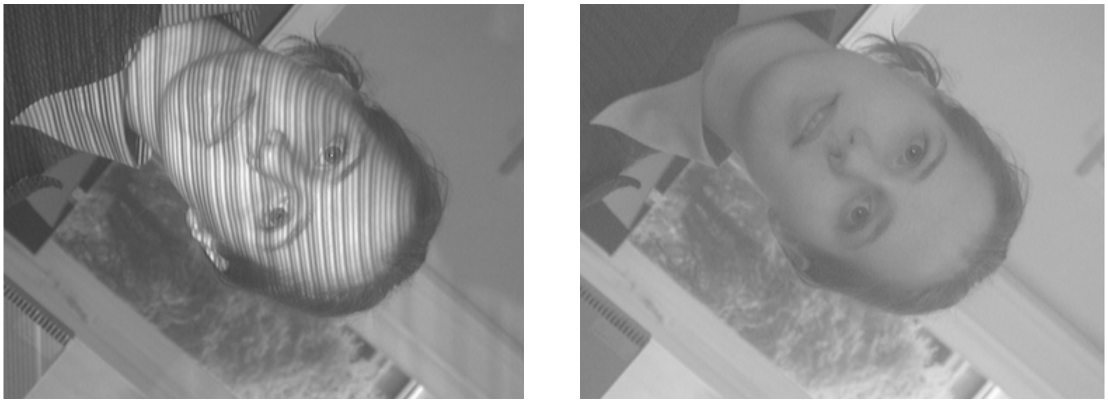


Figure 53: Corresponding (a) striped and (b) grey images from the 3D_RMA database

3.4.3 Geometry compensation

We first decided to get a 2-D profile of grey-level values along the central line (passing through the nose). Because distances between points of a 2-D image depend on pose, we used the 3D coordinates of the automatically extracted central profile (see section 3.3.5) to derive a Euclidian indexing of points along the profile. The definition of this profile (intersection with the vertical plane of symmetry) also calls for a better reproduction than a straight line on a 2-D image.

Since the grey image is expected to be registered with the striped image, striped image coordinates of the profile points were used to get intensity values from the grey image. To reduce the influence of noise and to extract more information from the mouth, nostrils, eyes and eyebrows, for each point of the profile, we averaged grey values in a direction perpendicular to the central profile (up to 4 cm on each side).

To better describe the face, we later extracted grey values from the lateral profiles obtained during the 3D analysis (see section 3.3.5). We then reduced the stripe width for average of the central profile to about 2.5 cm and we adopted the same width for the lateral profiles (see Figure 54). These left and right grey profiles were summed to increase robustness.

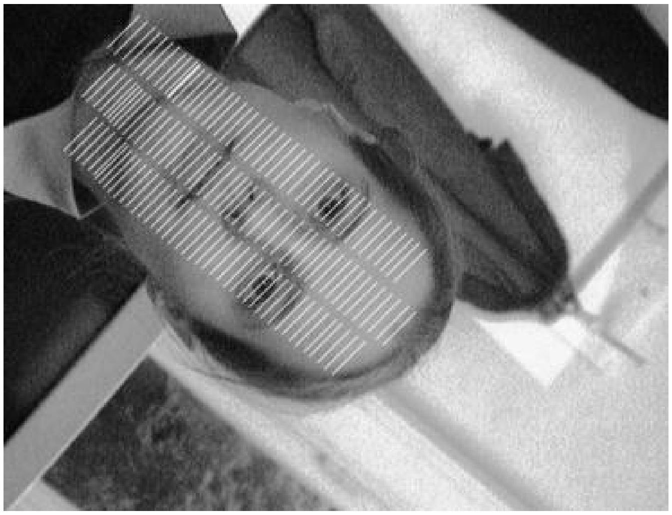


Figure 54: Extents of grey-level average perpendicularly to the central and lateral profiles

3.4.4 Grey-level compensation

Absolute grey-level values are not invariant. They depend on illumination.

A first compensation was achieved by considering the local difference of grey levels along the profiles. This reduces the dependence of grey measures on ambient light or skin variation (e.g. due to sun exposure) and reinforces the importance of the position of grey-level features relatively to their intensity values.

Another compensation should take into account the influence of the local surface orientation on the grey level. From the 3D description and a reflectance model, one is able to derive the albedo, a surface characteristic independent of viewpoint and illumination, and better suited to comparison. However, the grey images had a rather diffused illumination so that grey levels were used without correction.

3.4.5 Grey profiles

The grey-level extraction and compensation process delivers two 2D curves indexed by the signed distance to the nose reference point (negative towards chin, positive towards forehead), giving the local difference (along the profile) of average intensities (across the profile).

Figure 55 shows the compensated grey profiles of two persons for the two sessions. The left part of the figure is related to the central profile and the right part corresponds to the mean of the right and left lateral grey-level profiles.

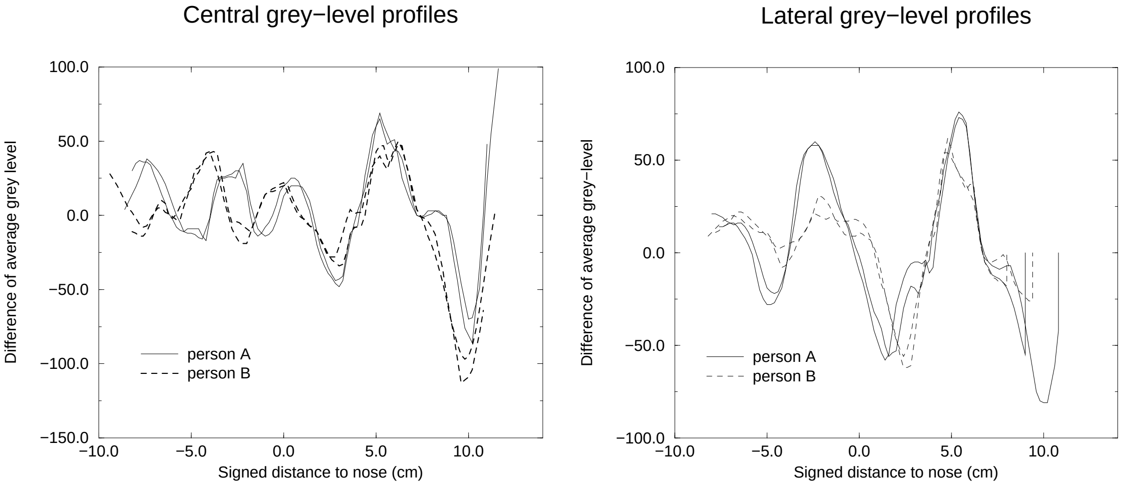


Figure 55: Central and lateral grey-level profiles of 2 persons, sessions 1 and 2

3.4.6 Grey profile comparison

The geometry and grey-level compensations allow for the direct comparison of extracted grey profiles. However, the nose point used as reference may suffer from imprecision so that several shifts (-1cm..+1cm) between profiles were considered.

For each shift, the histogram of the difference of corresponding profile values is computed. For all shifts, the minimum value of the mean of the 95 % lowest bins of the histogram is retained as the distance measure. Getting rid of the 5 % highest differences is a way to eliminate non-representative values.

3.4.7 Using striped images

Thanks to the average perpendicularly to the profiles and diagonally to the stripes, grey-level measurements from striped images are not so much influenced by the stripes. This simplifies the integration since a single striped image delivers 3D and intensity data. It also allowed us to conduct more tests thanks to the three striped shots available per person.

The only practical implication was the necessity to increase the width of average (from 2.5 to 3.5 cm) to better smooth the stripes out.

3.4.8 Results

We only discuss here the results (Table 13, ‘Grey Ctr’) related to the central grey profile, session2 compared to session1. Similar results hold for the lateral profiles (Table 13, ‘Grey Lat’).

EER (%)	3D Ctr	3D Lat	Grey Ctr	Grey Lat	Fusion
Shot 1 (grey)	12	8	9.5	16.5	1.2
Shot 1 (striped)	12	8	12	17.5	2.8
3 shots	14	12	16	20.5	7.2
3 shots (fusion)	10	9	15	17	4.1
Temporal fusion	8	7	9	16.5	1.4

Table 13: EER from 3D and Grey analysis for the central and lateral profiles, and fusion

In a first experiment, the grey-level image ("**Shot 1 (grey)**") in correspondence with the first 3D shot has been used for grey-level measurement. The Equal Error Rate of about 9% benefits from the similarity of the pose (frontal for shot 1) and the rather uniform lightening of ambient light.

In a second experiment ("**Shot 1 (striped)**"), the striped image of 3D shot 1 has been used to measure grey levels (see section 3.4.7). Thanks to the balanced influence of the stripes for a wide average, the recognition rate was not impaired too much (EER = 12.0 %). In those first two experiments, the figures (especially EER) have to be taken with care due to the reduced number of client tests (staircase aspect of the curves relatively to False Rejection).

Including more client and impostor tests is possible from the availability of the striped images relative to 3D shots 2 and 3, leading to 78 client and 2271 impostor tests. The recognition rate drops (EER=16%) due to the larger variability in pose among shots implying grey-level changes due to reflection of the directional projected light.

Disposing of three shots for session1 (reference) in the database allows to combine the comparison scores ("**3 shots (fusion)**") belonging to the same person, in order to increase robustness, taking advantage of the larger representation of the person in the database. The improvement in recognition was not very high (EER=15%) with our data.

Finally, the three individual tests (session2) may be combined as well to reduce the error rate. We call this process "**Temporal fusion**", because it consists in capturing and fusing more data over time when the person is tested.

The fusion considered in this section (for "3 Shots (fusion)" and "Temporal fusion") has been implemented by a simple average of score values to show possible improvements before envisaging other fusion techniques. The next section investigates an adaptive linear combination, but in the specific case of combining the 3D and grey, central and lateral profiles.

3.5 Fusing 3D and grey experts

The main objective of the grey-level analysis presented in the previous section is to access additional information to be combined with the 3D analysis in order to improve the recognition performance and the robustness of the system. Although the boolean decision of the 3D and grey experts could have been used ("hard fusion"), we preferred to combine the scores ("soft fusion") since thresholding may hide part of the information.

We linearly combined the scores of the four experts:

score = k1*3DCtr + k2*3Dlat + k3*GreyCtr + k4*GreyLat

where 3DCtr and 3Dlat are the scores of the geometrical experts and GreyCtr and GreyLat are the scores of the grey experts for the central and lateral profiles.
The coefficients of this linear combination are estimated using the Fisher method [Fisher36] that looks for the hyperplane that best separates the client and impostor scores. In his approach, Fisher approximates the client and impostor distributions of each expert as normal distributions, including expert dependencies expressed in a covariance matrix. Coefficients k1,…,k4 are determined to minimise false acceptance for a given false rejection rate. They are obtained analytically, avoiding local minima. The solution is however dependent on the validity of the normal distribution hypothesis. A part of client and impostor claims is used to train the fusion engine, while the rest is used to evaluate the operational false acceptance and false rejection rates. The ROC curve drawn for the optimal coefficients represents the false rejection versus false acceptance rates for different values of the decision threshold.

3.5.1 Results

We applied the Fisher method to linearly combine the two grey and two 3D experts. The results are presented in the form of ROC curves (Figure 56) for different cases of grey analysis. Grey measurements either come from grey images (“Shot 1 (grey)”) or striped images (“Shot 1 (striped)”, “3 Shots”, “3 shots (fusion)” or “Temporal fusion”). The EER after combination (Table 13, ‘Fusion’) shows a clear advantage of fusion in recognition performance.

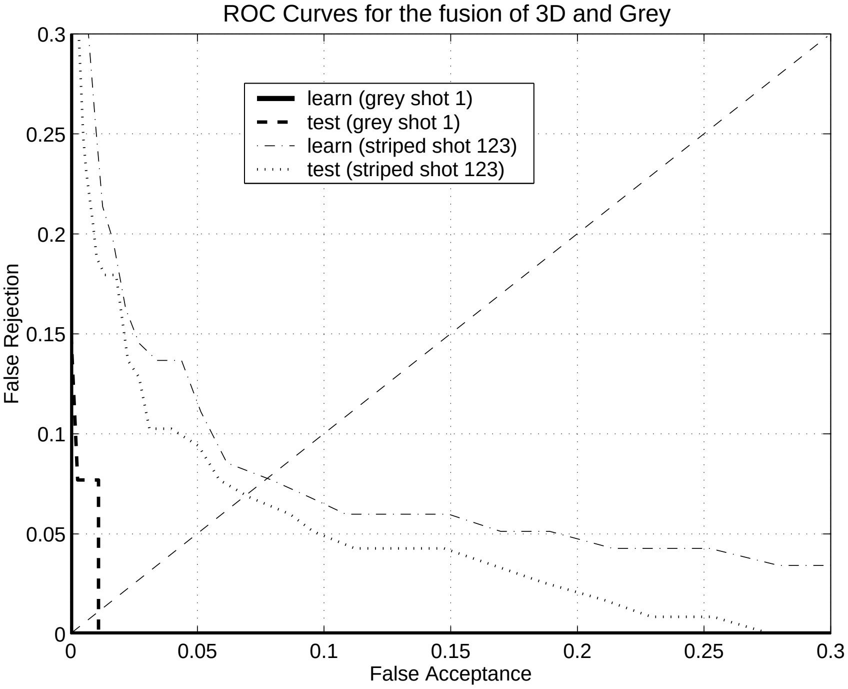


Figure 56: ROC Curves for the fusion of 3D and Grey, with grey or striped images, auto DB, no manual refinement

The difference in performance levels comes from the variability of the images used for tests. As already mentioned in the previous section, grey images only concern shot 1, with a diffuse illumination and a rather similar frontal pose for the two sessions. Striped images, shot 1, gives similar results thanks to the similar frontal pose and, consequently, the similar light reflection of the projected light. Striped images, all shots, introduce pose variations with lighting influences and 3D deficiencies. As visible in Table 13, the larger representation of the reference ('fusion') or of the test images ('temporal fusion') increases the recognition performance.

EER values have to be compared with caution, knowing that the number of False Rejection tests were small. These values were obtained using a fully automatic version of all processing steps and correspond to ***auto DB, auto*** of previous results. We are convinced that the figures are underestimated, according to the numerous sources of errors detailed in the section 3.6.

The current implementation takes 500 ms (Pentium 200 MHz) to perform central and lateral profile extraction. 100 ms is needed to compare two faces by the suggested 3D and grey analysis.

As presented in section 3.4.1, the simultaneous analysis of grey and geometrical information could be further investigated. Grey-level measures can help 3D acquisition and 3D profile extraction. Light reflection could be compensated thanks to the knowledge of the surface normals. Performance could be improved with little effort by comparing 3D and grey profiles simultaneously, using the same shift parameter.

3.6 Discussion

The important steps of processing of our 3D face verification approach are depicted in Figure 57. We can group the left components as belonging to the acquisition part and the right components to the 3D recognition part. Although each of these steps has been automated, results are provided with manual tuning as well, to guess the potential discriminative power of the approach and to see the influence of automation.

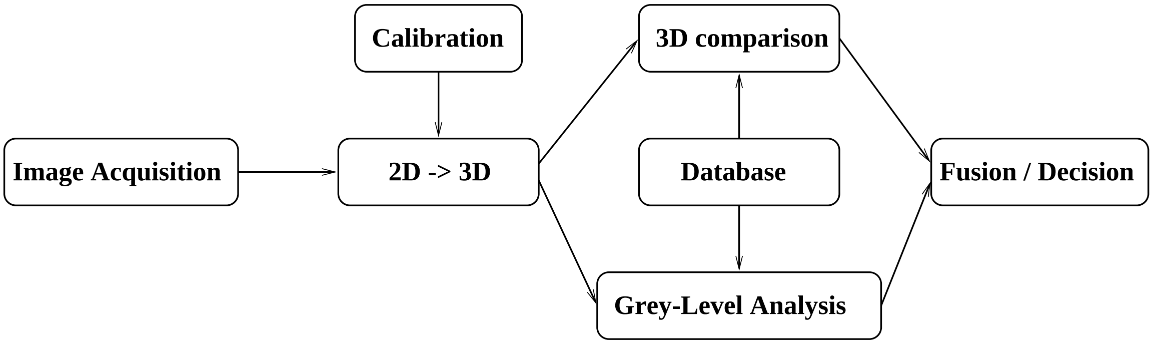


Figure 57: Synoptic with the different modules of the 3D verification system

The development of the 3D acquisition system and the 3D recognition software emerged from their mutual interaction.

The acquisition part was first developed and evaluated by the calibration procedure thanks to the residual error. Later, the 3D display of reconstructed objects or faces made us confident in the validity of the 3D surfaces. Many facial surfaces from the 3D_RMA database were then analysed to identify major problems related to the 3D face capture.

Several recognition approaches were experimented and led to the importance of a good surface capture of the nose. Recognition rates supported the general approach, acquisition and comparison included.

The inclusion of the grey analysis finally appeared as a way to improve the acquisition, the geometrical comparison and the recognition performance. Only the last point has been investigated so far.

Errors in the recognition process can be caused by the user's behaviour and by the processing of the different modules depicted in Figure 57.

a) Individual's cooperation

The person influences largely the results by the way he or she presents himself or herself in front of the camera. On the one hand, a right positioning in distance is essential, to avoid blurring of stripes, and a correct centring ensures that the whole face is captured.

On the other hand, people are expected to behave naturally, with a neutral expression and to take off disturbing objects such as glasses or scarf. User cooperation is the best acceptable requirement that solves these problems.

The main conclusion drawn from the tests is that individual's cooperation seemed natural. A small difference was noticed between session1 and session2 of the 3D_RMA database, attributed to a more relaxed behaviour in session2 (smiles) compared to session1.

b) Individual's particularities

Even in cases of correct individual's cooperation, parts of the face may be incorrectly captured due to typical facial components leading to difficult capture such as bushy beard, long hair or dark iris. Their impact could however be moderated by the analysis of the associated grey-level information.

Following the qualitative database analysis (see 2.8.2c)), typical acquisition problems were related to the beard, the nose and the eyes, by order of importance.

c) Hardware components

The camera (and the digitising board, if any) and the projector bring measurement errors.

Electronic noise affecting the camera and the digitising board may impair the detection of stripes in image. Grey-level saturation due to excessive light projection or spurious reflections prevents stripes to be correctly localised.

The lenses of the camera and projector imply a limited depth of field, possibly blurring stripes, impairing their correct detection. They also involve aberration resulting in imprecise 3D estimation.

Sources of hardware errors such as noise and lens aberrations are depicted by the impossibility to calibrate the system with no residual error. However, their effects are rather limited because the associated error is rather small, and they mainly affect the faces in a similar way.

The evaluation of prototype A revealed the effect of the limited resolution of the camera and the non-linearities in the image signals (Appendix C, Figure 71). These problems disappeared with the digital camera of prototype B and C.

d) Calibration

The precision of 3D coordinate measurements depends on the correct localisation of the points along the stripes, the good labelling of the stripes and the appropriate parameter estimation of the system. The limited precision of point localisation and lens or sensor distortion prevent the calibration to be perfect and induce systematic errors on 3D coordinates.

This kind of errors is however smooth and does not create discontinuities in the reconstructed surface. It is also likely that this systematic error will distort in the same way the captures obtained with the same system calibration, reducing its influence on comparison results as all the faces are presented consistently.

e) 3D extraction

The procedure used to get 3D coordinates from the acquired 2-D images relies on three phases that can all induce errors.

First, the stripes are localised and followed. Images with low contrast, high noise or saturation result in missing or erroneous 3D localisation. Persons with bushy beard or spectacles also suffer from bad 3-D sensing.

Secondly, stripes must be labelled and identified. Should this identification be wrong, associated 3D point coordinates could be estimated far from their real position, leading to large discontinuities in the surface.

Thirdly, 3D coordinates are computed for points along the stripes according to image position and stripe label. The correctness of these coordinates depends on calibration accuracy.

Stripe localisation and identification are critical for the quality of facial surfaces. Discontinuities in the surfaces may be induced, which largely affects the visual quality and the surface comparison algorithm. Most of the problems were localised in hair, beard and in the nose and eye regions.

f) 3D comparison

Once 3D representations are available, they are compared with reference 3D representations. The procedure followed here (see section 3.3.5) considers matching by distance minimisation depending on translation and rotation parameters. The minimum search can fail, due for instance to local minima or bad initial conditions, which typically explains false rejection cases. This type of error is strong in the sense that it usually leads to a wrong decision.

g) Grey extraction

The image source for grey measurement was either a second image, captured with the available ambient light and clearly underexposed, or the striped image, meant for 3D capture, and requiring grey-level average to sweep the influence of the projected stripes out.

Grey-level measurement considered the points related to the central and lateral profiles and was consequently dependent on the success of the profile extraction procedure.

h) Grey comparison

The grey comparison scheme suffers from the usual local minima and bad initial condition complications. In this respect, a concurrent comparison with the geometrical information when optimising the shift parameter would bring more robustness.

i) Decision

The decision rule used in the present work to answer the verification question ``Is this the right person?" is based on a simple threshold. The comparison of two 3D representations gives a distance of match. If this distance is bigger than the threshold, the two representations are supposed to be from different persons.

The choice of the threshold is important to solve the compromise between a low false rejection rate and a low false acceptance rate. Additional information, such as the natural variation of a given individual, or the presence of a beard can improve the system by choosing adaptive threshold.

j) Database

The database refers here to 3D descriptions that are stored as references of the individual. The quality of a person’s representation depends on the variety of allowable attitudes present in the database. For practical reasons such as memory limits and processing time, the number of references has been limited, considering a neutral expression and three different orientations.

For performance evaluation, the database is separated in test and reference descriptions. The number of references per individual is rather small, which normally leads to underestimated performances, because, from the lack of representation, the chance to present a description rather different from the references is higher.

Results showed that combining comparison scores of several references or several test descriptions improves the recognition performance (Table 13).

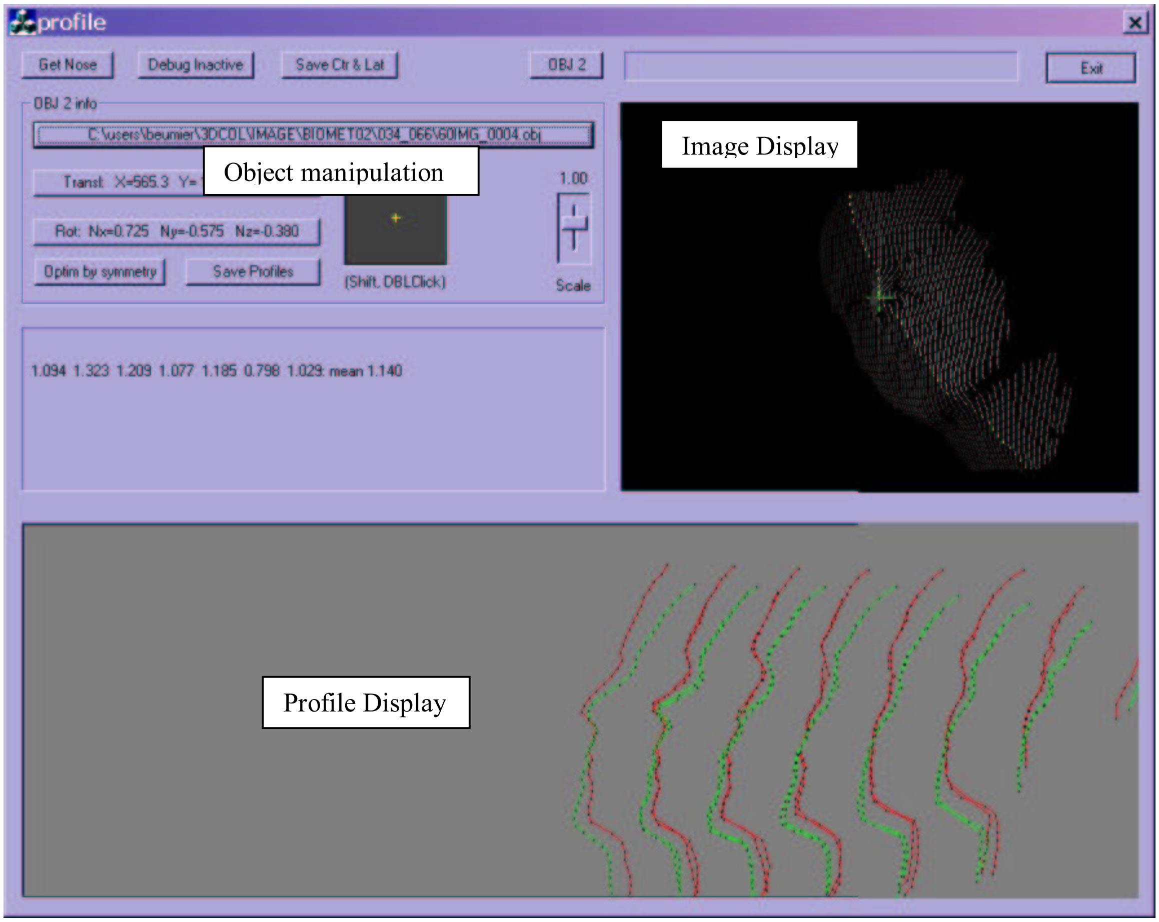
From all the presented errors, the most important and the less under control error relates to 3D extraction. It induces discontinuities in the 3D surface that may prevent correct 3D matching (due to a higher residual error or possible local minimum) or grey-level measurement. The sensitivity is less important for global surface matching which appears to be the method to be pursued now that computational power has increased.

3.7 BIOMET experiments (Prototype B)

This section reports on the preliminary experiments about face recognition with the BIOMET 3D database. About 500 3D faces have been recently digitised (refer to 2.8.3) so that only a few recognition experiments (3.7.3) with the central and lateral profiles and global surface matching have been carried out. We took the opportunity of developing software on Windows to built up new graphical interfaces for the profile extraction (3.7.1) and comparison (3.7.2) that did not exist when we tested the 3D_RMA database.

3.7.1 Profile extraction

The graphical interface for profile extraction contains three main areas (Figure 58).



C:\users\beumier\3DCOL\IMAGE\BIOMET02\034_066\60IMG_0004.obj

Transl: X=565.3 Y=

Rot: Nx=0.725 Ny=-0.575 Nz=-0.380

1.094 1.323 1.209 1.077 1.185 0.798 1.029: mean 1.140

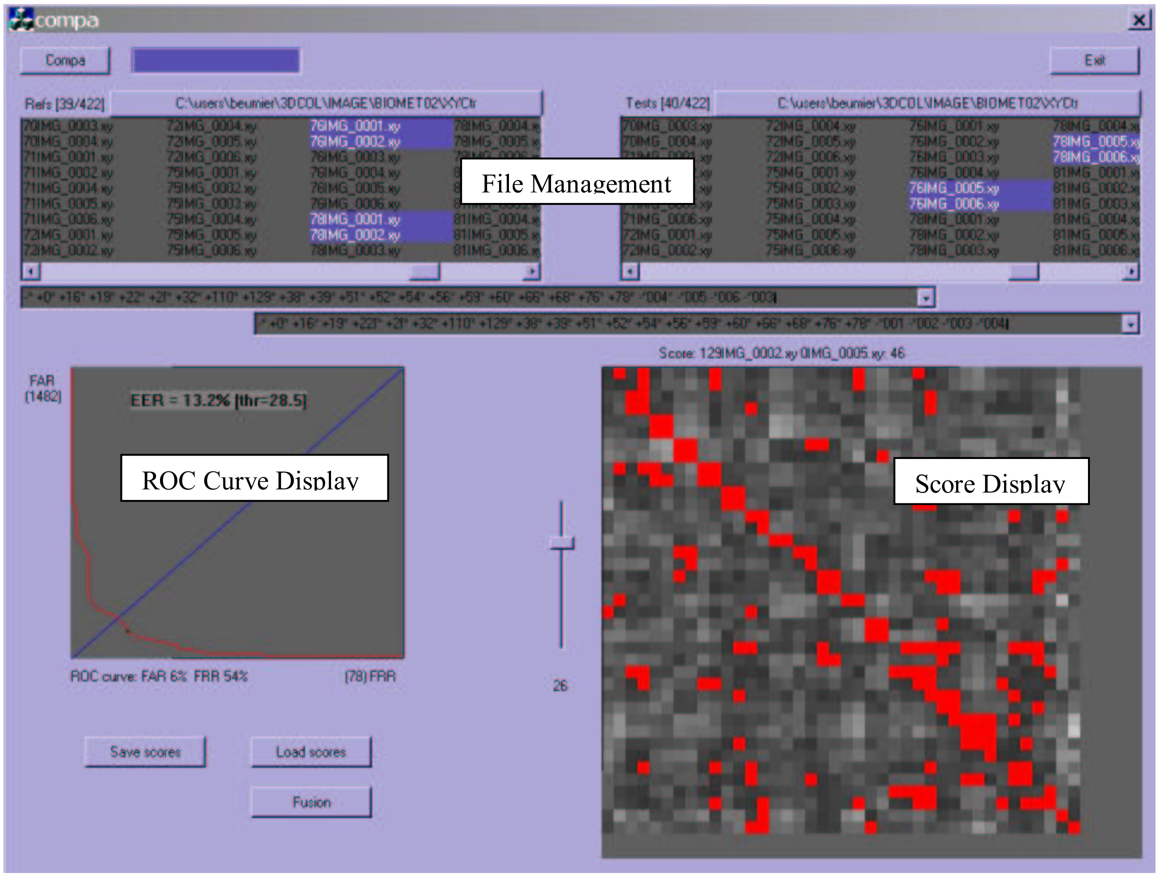
Figure 58: The graphical interface for profile extraction

The *Object Manipulation area* allows for the selection of a 3D file (.OBJ) and the modification of translation and rotation parameters. The current file is displayed in the *Image Display area*, highlighting the stripes. The *Profile Display area* displays up to 15 profiles extracted from the 3D object by parallel planes specified by the translation and rotation parameter values. Several 3D files can be loaded. This allows for the visual comparison of their profiles displayed in different colours.

The extraction of profiles can be manual or automatic. In the automatic mode, translation and rotation parameters are tuned to maximize the similarity of symmetrical profile pairs. A manual optimisation is always possible, looking for the similarity scores expressed by the distance in mm separating each profile pair. The profiles and the corresponding translation and rotation parameters can be stored to file. The central profile (passing through the nose) and one lateral profile (3 cm away from the central profile) can be stored for recognition experiments.

3.7.2 Profile comparison

The graphical interface for profile comparison contains three areas. The *File Management area* allows for the selection of files from a specified directory, with a dedicated filter command interpreter. The left part is reserved for specifying reference files and the right part refers to the test files.



Refs [39/422] C:\users\beumier\3DCOL\IMAGE\BIOMET02\XYCtr

Tests [40/422] C:\users\beumier\3DCOL\IMAGE\BIOMET02\XYCtr

-* +0* +16* +19* +22* +21* +32* +110* +129* +38* +39* +51* +52* +54* +56* +59* +60* +66* +68* +76* +78* -*004* -*005 -*006 -*003

-* +0* +16* +19* +221* +21* +32* +110* +129* +38* +39* +51* +52* +54* +56* +59* +60* +66* +68* +76* +78* -*001 -*002 -*003 -*004

Score: 129IMG_0002.xy 0IMG_0005.xy: 46

EER = 13.2% [thr=28.5]

ROC curve: FAR 6% FRR 54%

Figure 59: Graphical interface for profile comparison

The recognition experiment is based on the comparison of profiles selected in the test list relatively to files selected in the reference list. The experiment starts when the ‘Compa’ button is pressed and stops when the same button is pressed again or when all the files have been compared. A progress bar displays the degree of completion. The ROC curve is continuously computed and displayed in the *ROC Curve Display area* with the number of client (False Rejection) and impostor (False Acceptance) tests. The EER (Equal Error Rate) is automatically computed from the intersection of the ROC curve and the bisectrix (blue line). The full matrix of comparison scores is displayed in the *Score Display area* as an image with levels proportional to the scores. This presentation allows for the verification of particular values (with data file names) and highlights profiles with particular scores. A slider is used to modify the threshold, so that lower values are marked in red in the score image and the corresponding operational point is displayed on the ROC curve.

An experiment can be saved to file, storing the directory and filter of the reference and test profile lists, the list of profile files and the complete matrix of scores. Such a ‘.sco’ file can be loaded, with full context (directory, filter, …) either for verification or for conducting new and similar experiments. A fusion of score files has been implemented: the scores associated with each person are averaged.

3.7.3 Results

Recognition tests have been carried out for the two approaches of 3D face comparison presented in 3.3.

a) Central and lateral profiles

Central and lateral profiles were extracted by the graphical interface (Figure 58) thanks to the intrinsic symmetry of the face. The profiles were compared separately and their scores were fused. Results are presented in Table 14. More than 100 profiles can be compared per second.

	# client tests	**# impostor tests**	**Central** **EER**	**Lateral** **EER**	**Fusion** **EER**
Shot 123 - 456	670	52880	17.8 %	17.9 %	13.4 %
Shot 12 – 56	291	22923	16.6 %	17.6 %	13.0 %
Shot 1 – 6	72	5688	10.6 %	13.5 %	7.5 %

Table 14: 3D Face Recognition results for Prototype B, Central and Lateral profiles

A first experiment considered all the profiles of all the persons. Shots 4,5,6 are compared to shots 1,2,3. There are 670 client tests and 52880 impostor tests. The Equal Error Rates of nearly 18 % are similar to the scores obtained with session1 on the full database of Prototype A (Figure 50).

In a second experiment, shots 5,6 were compared to shots 1,2, bringing the same level of error rates.

In the last experiment, shot 6 was compared to shot 1, both corresponding to a frontal posture. This posture has the advantage of a correct and balanced acquisition of the facial surface. Shots 2 and 3 have left or right rotation that reduces the symmetry of facial coverage and often impairs nose capture. Shots 4 and 5 modify the throat capture, which largely influences the extracted profiles. The Equal Error Rate after fusion (7.5 %) of central and lateral scores is similar to Prototype A experiments (Table 13, Shot 1), but all the extracted 3D representations were considered here.

b) Global surface matching

Due to the lack of time, we implemented global surface matching in a two-step procedure. The first step extracts up to 15 parallel planar profiles (profile interface of Figure 58). This is based on the facial symmetry and solves three of the six translation and rotation parameters. The second step tunes the remaining three parameters in order to minimise the distance between the profile curves (see 3.3.4c). The matching score is the mean residual distance of each profile weighted by its number of points. A low value corresponds to a good match. In the current implementation, two seconds are needed to compare two facial surfaces.

We analysed the discrepancies between the different shots. The throat, mainly affected in shots 4 and 5 with up or down rotation, plays an important role by its large area and variability between people. It is however variable with posture. The left or right rotation of shots 2 and 3 influences the symmetry of the captured area and the visibility of the nose, very important for localisation and matching.

Table 15 summarises the experiments.

	# client tests	**# impostor tests**	**EER**	**Fusion (EER)**
Shots 123 – 456	651	51324	14.1 %	9.6 %
Shots 12 – 56	268	16688	8.9 %	7.8 %
Shots 12 – 6	140	11092	3.7 %	1.7 %
Shots 1 – 6	70	5546	3.6 %	---

Table 15: 3D Face Recognition results for Prototype B, with Global Surface Matching

In these experiments, the persons without one of the required shots were discarded.
Comparing Shots 456 with shots 123 leads to an important error rate due to the large differences in the representations of shots 2, 3, 4 and 5, as explained above. As expected, the ‘Shots 12-56’ and ‘Shot 1-6’ comparison error rates are lower thanks to the reduced variability between the shots. Comparing shots 1 and 6 is particularly promising as those shots emanate from the same point of view.

For the tests containing several reference shots (12 or 123), the average of the scores obtained for each person was used as matching score in ‘fusion’ experiments. The gain in error rate reduction is clear.

The main conclusion about this experiment is the necessity to dispose of a good representation of the faces. Several shots for each important posture should be available, or the posture could be required frontal, without necessitating much user cooperation. The experience we gained with profile recognition ([Beumier97d]) promotes for five or six references to account for different sources of errors.

Compared to the results obtained with prototype A, the present error levels are promising, due to the fact that no capture with extracted 3D information has been rejected and provided that the quality of 3D capture during the BIOMET campaign was bad for 20 % of the images. Moreover, the matching approach is currently performed in two steps, which may prevent the global minimum to be reached. Finally, the inclusion of grey-level clues has not been implemented yet and is expected to largely contribute to the reduction of error rates.

3.8 Conclusions

Face recognition from 3D facial surface and grey-level clues has been presented.

The 3D geometrical comparison is based on the matching of facial surfaces through the optimisation of the three rotation and three translation parameters in order to reduce the distance between the two surfaces. In a first experiment, the global matching is achieved by minimising the distances between corresponding planar profiles obtained from parallel planar cuts of the facial surfaces. Although promising, the implementation was too slow by the time of development. We prefer to conduct a second experiment based on the extraction of the central and lateral profiles, thanks to the intrinsic symmetry of the head. The second approach presents the following properties:

- Normalisation of the 3D data relatively to three parameters (two rotation and one translation) thanks to the intrinsic symmetry of the head;
- compact representation with two planar curves;
- rapid comparison with only one parameter (shift).

The presented grey-level analysis was successfully integrated in the existing 3D comparison method from central and lateral profiles:

- Grey measurements are obtained directly from a grey-level image in alignment with the striped image or directly from the striped image;
- Geometrical compensation of the grey values is based on Euclidian distances, derived from the extracted profiles;
- The grey profile comparison procedure benefits from simplicity and quickness.

The recognition rates, improved by the combination of 3D and grey data, supports the approach. The time latency of less than 1 sec to compare a 3D representation with the claimed references from the database is compatible with a practical application. Background removal, translation and scale independence are important assets of the method that should further guarantee application success.

We also showed that using the striped images for grey measurement implied a small performance penalty in recognition but allows easy 3D and intensity capture from the same image.

Tests on the BIOMET database (prototype B) were recently carried out. The extraction of the profiles seems better from the higher quality of the 3D surfaces. The major problems with the central and lateral approach remain the quality in the nose region and the presence of 3D discontinuities. The global surface matching is therefore more robust, but results clearly suffered from the lack of captures from similar postures. The comparison of the two frontal shots of the whole database gave promising results. Quantitative results with geometrical comparison are better than those obtained with prototype A. No grey-level comparison has been carried out.

Conclusions and perspectives

Objectives

The purpose of this thesis was to design a ***realistic face recogniser*** based on facial surface analysis. Our experience in 2D face recognition through frontal and profile views led us to the conclusion that the 3D geometrical information was particularly well suited to build up a realistic and robust face recognition solution. The range information simplifies the localisation of the face, provides for robust and normalised features and addresses the two major limitations of 2D face analysis for recognition: the point of view and illumination dependency. The inclusion of grey-level clues and their combination with geometrical features are an additional asset that allows to exploit most of the available information present in the facial surface.

The two key issues to be addressed when designing and realising a 3D face recognition system are the acquisition of 3D facial surfaces and the comparison of the acquired surfaces. One important question was related to the possibility of distinguishing people from one another through facial surface comparison, and to achieve this, a statistical analysis through extensive testing had to be carried out. During all the developments, particular attention was devoted to the possibility to automatise the designed algorithms and to rely as little as possible on expensive solutions.

Acquisition

Facial surfaces have been acquired with a ***structured light*** approach, including a camera and a slide projector, and taking advantage of the rapidity, low cost and hardware simplicity of that solution. Several slide patterns with different structured elements have been designed in order to improve three criteria: the precision in element detection (localisation), the correctness of element identification (labelling) and the possibility to recover object texture, as independently from the projected pattern as possible. Parallel stripes appeared to be the best projected pattern due to continuity in one direction and the facility of precise detection.

Three prototypes were realised. Prototype A is highly contrasted and suits the low resolution camera we used in the beginning of the project, but is poor at recovering texture. Prototype B addresses the problem of labelling and texture recovery thanks to colour striping. The limited contrast, especially for certain skin colours, reduces the accuracy and correctness of stripe localisation. Prototype C makes use of black and white striping to maximise the stripe visibility while limiting the influence on texture capture. Stripes are labelled thanks to the vertical position of dots.

Especially for prototype C, most of the facial information is captured by one ***frontal view***, extracting texture information from the same image capture. Other views suffer from the lack of geometrical information in areas such as the nose or the throat. Difficulties linked to the presence of beard or hair cannot be alleviated. Defects due to glasses or opened mouth can be avoided through user cooperation.

A ***qualitative evaluation*** has been performed on the databases acquired with prototypes A and B. Except for highly non frontal captures or faces with bushy beard or long hair, most of the facial surface was acquired in one shot. The texture captured by prototype B is adequate for face recognition. Prototype C, still in the feasibility stage, is expected to bring higher resolution from higher stripe density, larger coverage thanks to improved stripe contrast, and better texture recovery as a result of the periodicity of the projected pattern.

Face recognition

Our 3D face recogniser results from the combination of geometrical and grey-level curve comparison.

The ***geometrical comparison*** consists in optimising 6 parameters of rotation and translation to match corresponding 2D profiles extracted from the facial surfaces by parallel planes. In a first approach called global surface matching, the geometrical matching score between two 3D faces is the residual distance between pairs of corresponding profiles after optimisation. In a second approach called Central and Lateral profiles, the matching score is based on the similarity of local slope values along the central and lateral profiles. These are extracted thanks to the symmetry of the head, solving three of the six parameters. They are then converted into local slopes measures that are easily compared with a 1-parameter search (shift).

The second approach facilitated the inclusion of a ***grey-level analysis***. Grey-level values are measured along the profiles, with an average perpendicular to the profile. The characteristics to be compared are differential grey-level values along the profiles, which are largely independent of ambient lighting conditions. The geometrical and grey-level comparison scores of central and lateral profiles were combined with a linear classifier.

The ***results*** showed the limited performance of each profile recogniser (typical EER of 10 to 15%) but at the same time, the benefits of combining them (EER of a few percents). The global surface matching approach is more robust and retained recently our attention thanks to the increase in computational power. Facial comparison is mainly affected by the discontinuities of the profiles due to absent or incorrect 3D acquisition areas. The comparison algorithm may also lead to wrong conclusions when it converges to a local minimum during matching. The recognition figures are finally dependent on the content of the database, where factors such as the variability of the population and the quality of the representation of each individual play an important role.

Perspectives

Among our 3D ***acquisition developments***, Prototype C appears as the best solution to capture a high density of 3D points with texture. An implementation with a flash lamp and slide made of glass should enhance the acquisition quality and enlarge the acquisition depth of field. Infra-red projection is possible, for improved user comfort. A better connection of the camera to the computer would reduce capture delays and allow real-time applications.

Although the whole recognition chain was designed to be automatic, limited manual intervention was necessary for 3D acquisition and recognition, mainly due to geometrical discontinuities in captured 3D surfaces due to incorrect stripe localisation and labelling. The ***frontal pose*** reduces by far those inconveniences and is a natural position to be taken on by an individual in a cooperative scenario.

A practical implementation could take advantage of the analysis of a ***sequence of captures*** to build up a better description of the face, either thanks to the more complete information or by appropriately selecting a good quality capture. A rough 3D analysis and an intuitive human machine interface could then guide the individual to a correct posture, if requested, or could select the appropriate capture from the sequence.

As far as the recognition phase is concerned, the capture or selection of frontal poses also leads to better performances when comparing facial surfaces. The database of references could then consist of more frontal shots. As another way to improve the ***description of people***, reward and punishment figures can help giving more weight to typical reference shots present in the database. In the verification paradigm, the claimed identity can be used to accumulate reward points to the reference shots that participated in the correct verification of the person and punishment points can be attributed to database entries which led to a wrong decision. This log information is also a means of updating database content to adapt the system to changes in people.

The whole chain of processing for 3D face acquisition and recognition benefits from ***grey-level analysis***. Our experiments only showed the advantage of combining 3D and grey comparison scores. But 3D acquisition with structured light could also reduce the confidence of 3D capture in regions where the underlying surface appears dark or where image contrast is poor. And the recognition approach could integrate in the same cost function 3D and grey-level criteria to lead to a coherent minimum and avoid some of the local minima.

Concluding from the tests on geometrical and grey profile comparison, we would now propose the ***global surface approach***, including at the same time geometrical and grey information. This heavy optimisation with six parameters (rotation and translation) could be speeded up thanks to initial parameter values based on natural postures and on facial symmetry. Many local minima could be avoided this way.

Bibliography

[Achermann97] B. Achermann, X. Jiang, H. Bunke,
"Face Recognition Using Range Images",
Proc. Int. Conf. on Virtual Systems and MultiMedia '97 (VSMM '97), Geneva, Switzerland, pp 129-136, Sep 1997.

[Achermann00] B. Achermann, H. Bunke,
"Classifying Range Images of Human Faces with Hausdorff Distance",
Int Conf on Pattern Recognition 2000, Barcelona, Spain, pp 813-817, Sep 2000.

[Acheroy96a] M.P. Acheroy, C. Beumier, J. Bigün, G. Chollet, B. Duc, S. Fischer, D. Genoud, P. Lockwood, G. Maitre, S. Pigeon, I. Pitas, K. Sobatta, L. Vandendorpe,
"Multi-modal person verification tools using speech and images",
Proceedings of the European Conference on Multimedia Applications, Services and Techniques (ECMAST '96),Louvain-La-Neuve, May 28-30 1996, pp 747-761.

[Ahn99] S.C. Ahn, H.-G. Kim,
"Automatic FDP (Facial Definition Parameters) Generation for a Video Conferencing System",
International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging, Santorini, Greece, Sept 1999, pp 16-19.

[Aibara92] T. Aibara, K. Ohue, Y. Oshita,
"Human face profile recognition by a P-Fourier descriptor",
Optical Engineering, Vol. 32 No. 4, 861-863, Apr 1993.

[Aizawa95] K. Aizawa, T. Huang,
"Model-Based Image Coding: Advanced Video Coding Techniques for Very Low Bit-Rate Applications",
Proceedings of the IEEE, Vol. 83, No. 2, Feb 1995.

[Akimoto93] T. Akimoto, Y. Suenaga, R. Wallace,
"Automatic Creation of 3D Facial Models",
IEEE Computer Graphics & Applications, pp 16-22, Sep 1993.

[Asada86] M. Asada, H. Ichikawa and S. Tsuji,
"Determining of Surface Properties by Projecting a Stripe Pattern",
8th Int. Conf. on Pattern Recognition, Vol 2, Paris, France, pp 1162-1164, Oct 1986.

[Asada88] M. Asada, H. Ichikawa and S. Tsuji,
"Determining of Surface Orientation by Projecting a Stripe Pattern",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 10, No 5, pp 749-754, Sep 1988.

[Augusteijn93] M. Augusteijn and T. Skufca,
"Identification of Human Faces through Texture-Based Feature Recognition and Neural Network Technology", IEEE Conf. Neural Networks, pp 392-398, 1993.

[Baron81] R.J. Baron,
"Mechanisms of human facial recognition",
International Journal of Man-Machine Studies,Vol 15, pp 137-178.

[Bartlett98] M.S. Bartlett, H.M. Lades, and T.J.Sejnowski,
"Independent component representations for face recognition",
Proc. of the SPIE Symposium on Electronic Imaging: Science and Technology; Conference on Human Vision and Electronic Imaging III, San Jose, California, January, 1998.

[Battle98] J. Battle, E. Mouaddib,
“Recent progress in coded structured light as a technique to solve the correspondence problem: a survey”,
Pattern Recognition, Vol 31, No 7, pp 963-982, 1998.

[Bernardini02] F. Bernardini, H.E. Rushmeier,
“3D Model Acquisition Pipeline”,
Computer Graphics Forum, Vol. 21, No 2, pp 149-172, 2002.

[Beumier95a] C. Beumier, M.P. Acheroy,
“Automatic Face Identification”,
In *Applications of Digital Image Processing XVIII*, SPIE, vol. 2564, July 1995, pp 311-323.

[Beumier96b] C. Beumier, M.P. Acheroy,
« Reconnaissance automatique de visages »,
In BIVA VISION'S TODAYS SOLUTION, Séminaire 14/3-15/3, Château Gravenhof, Dworp, Belgium, 1996.

[Beumier97a] C. Beumier, M.P. Acheroy,
“Person Authentication with structured light”,
In Deliverable 3.1.1 “multi-modal basic algorithm components”,
Project M2VTS, Programme ACTS, Oct 96.

[Beumier97b] C. Beumier, M.P. Acheroy,
“Automatic Face Recognition by 3-D Analysis”,
In ORBEL 11 (Belgian Operations Research Society), Namur, Belgium, January 16-17, 1997.

[Beumier97c] C. Beumier, M.P. Acheroy,
“Person Authentication with structured light”,
In Deliverable 3.2.1 “algorithm complexity evaluation and selection”,
Project M2VTS, Programme ACTS, Jan 97.

[Beumier97d] C. Beumier, M.P. Acheroy,
“Automatic Profile Identification”,
In *FIRST INTERNATIONAL CONFERENCE ON AUDIO AND VIDEO BASED BIOMETRIC PERSON AUTHENTICATION (AVBPA)*,
Crans-Montana, Switzerland, March 12-14 1997, pp 145-152.

[Beumier98a] C. Beumier, M.P. Acheroy,
“Automatic 3D Face Authentication”,
In *Workshop on Advances in Facial Image Analysis and Recognition Technology*, Post ECCV'98. Freiburg, Germany, June 6, 1998.

[Beumier98b] C. Beumier, M.P. Acheroy,
“Automatic Face Authentication from 3D Surface”,
In *British Machine Vision Conference BMVC 98*,
University of Southampton UK, 14-17 Sep, 1998, pp 449-458.

[Beumier98e] C. Beumier, M.P. Acheroy,
“Final Report on the structured light modality”,
In Deliverable 3.3.1 “algorithm refinement”, Project M2VTS, Programme ACTS, Jul 98.

[Beumier99b] C. Beumier, M. Acheroy,
“3D Facial Surface Acquisition by Structured Light”,
In *International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging],*
Santorini, Greece, 15-17 Sep, 1999, pp 103-106.

[Beumier99c] C. Beumier, M. Acheroy,
"SIC_DB: Multi-Modal Database for Person Authentication",
In *Proceedings of the 10th International Conference on Image Analysis and Processing*, Venice, Italy, 27-29 Sep, 1999, pp 704-708.

[Beumier00a] C. Beumier, M. Acheroy,
"Automatic 3D Face Authentication", In *Image and Vision Computing*, Vol. 18, No. 4, pp 315-321.

[Beumier00c] C. Beumier, M. Acheroy,
"Automatic Face Verification from 3D and Grey Level Clues",
RECPAD2000, 11th Portuguese conference on pattern recognition,
Porto, Portugal, May 11-12, 2000.

[Beumier00d] C. Beumier, M. Acheroy,
"Automatic Face Recognition",
Proceedings symposium IMAGING, Eindhoven, The Netherlands, May 18, 2000, pp 77-89.

[Beumier00f] C. Beumier, M. Acheroy,
"Face Verification from 3D and grey-level clues",
In *Pattern Recognition Letters*, Vol. 22, 2001, pp 1321-1329.

[Beumier01a] C. Beumier,
« Reconnaissance de personnes par surface faciale »,
Journées de travail GDR-PRC ISIS (CNRS) « Numérisation et Modélisation 3D », ENST, Paris, Jan 24-25, 2001.

[Beumier01b] C. Beumier,
"Automatic Face Recognition with 3D Support",
Biometrics as a Business Strategy, Radisson SAS Royal Hotel, Copenhagen, Denmark, Apr 25-26, 2001.

[Beumier01c] C. Beumier,
« Face Recognition", Slides,
Biometrie studiedag, Ingenieurshuis - KVIV, Antwerpen, Oct 25, 2001.

[Beumier03a] C. Beumier,
"Facial Surface Acquisition with Colour Striping",
ICANN/ICONIP 2003, Istanbul, Turkey, June 25-29 2003, pp 548-551.

[Beymer93] D. Beymer,
"Face Recognition Under Varying Pose",
AI Memo No 1461, CBCL Paper No 89, MIT 1993.

[BIOMET],
Incentive research project of GET (RD123, RE408), France.

[Blake93] A. Blake, D McCowen, H.R. Lo, P.J. Lindsey,
"Trinocular Active Range-Sensing",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 15, No 5, pp 477-483, May 1993.

[Blanc-Garin86] J. Blanc-Garin,
"FACES AND NON-FACES IN PROSOPAGNOSIC PATIENTS",
In H. Ellis, M. Jeeves, F. Newcombe and A. Young editors, "Aspects of Face Processing",
Doordrecht, Netherlands: Martinus Nijhoff Publishers, p. 273-278, 1986.

[Boyer87] K. Boyer, A. Kak,
"Color-Encoded Structured Light for Rapid Active Ranging",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 9, No 1, pp 14-28, Jan 1987.

[Bozdagi94] G. Bozdagi, A. Tekalp, and L. Onural,
"An Improvement to MBASIC Algorithm for 3-D Motion and Depth Estimation",
IEEE Trans. on Image Processing, Vol. 3, No. 5, pp 711-716, Sept 1994.

[Brechb95] C. Brechbühler, G. Gerig, O. Kübler,
"Parameterization of Closed Surfaces for 3-D Shape Description",
Computer Vision and Image Understanding, Vol. 61, No 2, pp 154-170, Mar 1995.

[Brunelli93] R. Brunelli, T. Poggio,
"Face Recognition: Features versus Templates",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 15, No 10, Oct 1993, pp 1042-1052.

[Brunelli95] R. Brunelli, D. Falavigna, T. Poggio, L. Stringa,
"Automatic person recognition by acoustic and geometric features",
Machine Vision and Applications (1995), 8, pp 317-325.

[Carrihill85] B. Carrihill, R. Hummel,
"Experiments with the Intensity Ratio Depth Sensor",
Computer Vision, Graphics and Image Processing, 32, pp 337-358, 1985.

[Carter90] J. Carter, T. Mathews, C. Shadle,
"A three-dimensional measurement system for speech research based on structured light",
SPIE Vol 1349 Applications of Digital Image Processing XIII, 1990.

[Cartoux89] J.Y. Cartoux, J.T. Lapresté, M. Richetin
"Face Authentification or Recognition by Profile Extraction from Range Images",
IEEE Computer Society Workshop on Interpretation of 3D scenes (1989), pp 194-199.

[Caspi98] D. Caspi, N. Kiryati, J. Shamir,
"Range Imaging With Adaptive Color Structured Light",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 20, No 5, pp 470-480, May 1998.

[Chang97] S. Chang, M. Rioux, J. Domey,
"Face recognition with range images and intensity images",
Optical Engineering, Vol. 36, No 4, pp 1106-1112, Apr 97.

[Chang03] K. Chang, K. Bowyer, P. Flynn,
"Multi-Modal 2D and 3D Biometrics for Face Recognition",
Int. Workshop on Analysis and Modeling of Faces and Gestures 2003,
Nice, France, pp 187-194, Oct 2003.

[Chellappa95] R. Chellappa, C.L. Wilson and S. Sirohey,
"Human and Machine Recognition of Faces: A Survey",
Proceedings of the IEEE, Vol. 83, No. 5, pp. 705-740, May 1995.

[Chen93] Z. Chen, S.-Y. Ho, D.-C. Tseng,
"Polyhedral Face Reconstruction and Modeling from a Single Image with Structured Light",
IEEE Trans. on Systems, Man, and Cybernetics", Vol 23, No 3, 1993.

[Chen97] C.S. Chen, Y.P. Hung, C.C. Chiang, J.L Wu,
"Range Data Acquisition Using Color Structured Lighting and Stereo Vision",
Image and Vision Computing, vol. 15, pp. 445-456, 1997.

[Chin86] R.T. Chin, C.R. Dyer,
“Model-based recognition in robot vision”,
ACM Computing Surveys, 18, pp 67-108, March 86.

[Choi91] C.S. Choi, H. Harashima, T. Takebe,
“Analysis and synthesis of facial expressions in knowledge-based coding of facial image sequences”,
In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing,
Toronto, Vol 4, May 1991, pp 2737-2740.

[Chua00] C-S. Chua, F. Han, Y-K. Ho,
“3D Human Face Recognition Using Point Signature”,
Int. Conf. on Automatic Face and Gesture Recognition 2000,
Grenoble, France, pp 233-238, March 2000.

[Cox95] J. Cox, J. Ghosn, P. Yianilos,
“Feature-Based Face Recognition Using Mixture-Distance”,
NEC Research Institute, Technical Report 95.

[Craw87] I. Craw, H. Ellis and J.R. Lishman,
“Automatic extraction of face-features”,
In Pattern Recognition Letters, Vol. 5 (1987), pp 183-187.

[Dahler87] J. Dahler,
“Problems in Digital Image Acquisition with CCD Camera”,
Proc. of ISPRS Intercommission Conference, Interlaken, Switzerland, 1987, pp 48-59.

[Dickinson93] S. J. Dickinson,
“Part-Based Modeling and Qualitative Recognition”,
In Three Dimensional Object Recognition Systems, Advances in Image Communication, Vol. 1

[Donato99] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski,
“Classifying Facial Actions”,
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 21, No. 10, Oct 1999, pp 974-989.

[Ekman78] P. Ekman and W. Friesen,
“Facial Action Coding System: A Technique for the Measurement of Facial Movement”,
Palo Alto, Calif., Consulting Psychologists Press, 1978.

[Etemad97] K. Etemad and R. Chellappa,
“Discriminant Analysis for Recognition of Human Face Images”,
1st Int. Conf on Audio and video-based Biometric Person Authentication (AVBPA),
Crans-Montana, Switzerland, March 12-14 1997, pp 127-142.

[Faugeras86] O.D. Faugeras, G. Toscani,
“Calibration Problem for Stereo”,
Proc. Int. Conf. Computer Vision and Pattern Recognition, June 1986, pp 15-20.

[Fisher1936] R.A. Fisher,
“The use of multiple measurements in taxonomic problems”,
Annals of Eugenics, Vol. 7, pp. 179-188, 1936.

[Flynn89] J. Flynn, A. Jain,
“On Reliable Curvature Estimation”,
IEEE Conf. on Computer Vision and Pattern Recognition, CVPR '89, June 4-8, pp 110-116, 1989.

[Forster01] F. Forster, P. Rummel, M. Lang, B. Radig,
“The HISCORE camera - A real time three dimensional and color camera”,
International Conference on Image Processing ICIP2001, Oct 7-10 2001, Thessaloniki, Greece.

[Fraser86] I.H. Fraser and D.M. Parker,
“REACTION TIME MEASURES OF FEATURE SALIENCY IN A PERCEPTUAL INTEGRATION TASK”,
In H. Ellis, M. Jeeves, F. Newcombe and A. Young editors, “Aspects of Face Processing”,
Doordrecht, Netherlands: Martinus Nijhoff Publishers, pp. 45-52, 1986.

[Galton1910] Sir F. Galton,
“Numeralized profiles for classification and recognition”,
In Nature 83, pp. 127-130, 31 march 1910.

[Gardel03] A. Gardel, J.L. Lazaro, J.M. Lavest,
“Influence of Mechanical Errors in a Zoom Camera”,
Image Anal Stereol 2003, vol. 22 pp 21-25.

[Gärtner96] H. Gärtner, P. Lehle, H. Tiziani,
“New, Highly Efficient, Binary Codes for Structured Light Methods”,
SPIE, Vol. 2599, 1996.

[Geng96] Z.J.Geng,
“Rainbow three-dimensional camera: new concept of high-speed three-dimensional vision systems”,
Optical Engineering, Vol. 35, pp376-383, Feb 1996.

[Gordon91] G. Gordon,
“Face recognition based on depth maps and surface curvature”,
SPIE Geometric methods in Computer Vision, Vol 1570, San Diego, 1991.

[Gordon92] G. Gordon,
“Application of Morphology to Feature Extraction for Face Recognition”,
Non-linear Image Processing III, Proc. SPIE, Vol. 1658, Feb 1992, San Jose, USA.

[Gordon92b] G. Gordon,
“Face Recognition Based on Depth and Curvature Features”,
Computer Vision & Pattern Recognition, Champaign, Illinois, 1992.

[Goudail96] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma and N. Otsu,
“Face Recognition System Using Local Autocorrelations and Multiscale Integration”,
IEEE Trans. on Pattern Analysis and Machine Intelligence, VOL. 18, NO. 10, Oct 1996, pp 1024-1028.

[Griffin92] P. M. Griffin, L. S. Narasimhan, S. R. Yee,
“Generation Of Uniquely Encoded Light Patterns For Range Data Acquisition”,
Pattern Recognition, Vol. 25, No. 6, pp 609-616, 1992.

[Guisser92] L. Guisser, R. Payrissat, S. Castan,
“A New 3-D Surface Measurement System Using a Structured Light”,
IEEE Conf. on Computer Vision and Pattern Recognition, CVPR '92, pp 784-786, Jun 1992.

[Haig86] N.D. Haig,
“INVESTIGATING FACE RECOGNITION WITH AN IMAGE PROCESSING COMPUTER”,
In H. Ellis, M. Jeeves, F. Newcombe and A. Young editors, “Aspects of Face Processing”,
Doordrecht, Netherlands: Martinus Nijhoff Publishers, pp. 410-425, 1986.

[Hall-Holt01] O. Hall-Holt, S. Rusinkiewicz,
"Stripe Boundary Codes for Real-Time Structured Light Range Scanning of Moving Objects",
ICCV 8th International Conference on Computer Vision, Vancouver, Canada, pp. 359-366, 2001.

[Harmon78] L. Harmon, S. Kuo, P. Ramig, and U. Raudkivi,
"Identification of Human Face Profiles by Computer",
Pattern Recognition, Vol. 10, pp 301-312, 1978.

[Harmon81] L. Harmon, M. Khan, R. Lasch, and P. Ramig,
"Machine Identification of Human Faces",
Pattern Recognition, Vol. 13, No. 2, pp 97-110, 1981.

[Hata92] S. Hata,
"Shape Detection of Small Specular Surface using Color Stripe Lighting",
Proc. Int. Conf. on Pattern Recognition, pp. 554-557, 1992.

[Hoffman87] R. Hoffman, A.K. Jain,
"Segmentation and Classification of Ranges Images",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 9, No 5, pp 608-620, Sep 1987.

[Horn99] E. Horn, N Kiryati,
"Toward optimal structured light patterns",
Image and Vision Computing 17, pp 87-97, 1999.

[Hsieh98] Y.-C. Hsieh,
"A note on the structured light of three-dimensional imaging systems",
Pattern Recognition Letters, Vol 19, pp 315-318, 1998.

[Hu89] G. Hu, G. Stockman,
"3-D Surface Solution Using Structured Light and Constraint Propagation",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 11, No 4, pp 390-402, Apr 1989.

[Hu96] J.-H. Hu, R.-S. Wang, Y. Wang,
"Compression of personal identification pictures using vector quantization with facial feature correction",
Optical Engineering, Vol. 35 No. 1, Jan 1996, pp 198-203.

[HuangC93] C.-L. Huang, C.-W. Chen,
"Human Facial Feature Extraction for Face Interpretation and Recognition",
Pattern Recognition, Vol. 25, No. 12, pp 1435-1444, 1992.

[Huang93] H.-C. Huang, M. Ouhyoung, J.-L. Wu,
"Automatic feature point extraction on a human face in model-based image coding",
Optical Engineering, Vol. 32, No. 7, pp 1571-1580, Jul 1993.

[Hurt84] S.L. Hurt, A. Rosenfeld,
"Noise Reduction in Three-Dimensional Digital Images",
Pattern Recognition Vol. 17, No. 4, pp. 407-421, 1984.

[Ittner85] D. Ittner, A. Jain,
"3-D Surface Discrimination From Local Curvature Measures",
IEEE Conf. on Computer Vision and Pattern Recognition, CVPR '85.

[Jain97] A.K. Jain, L. Hong, S. Pankanti, R. Bolle,
"An Identity-Authentication System Using Fingerprints",
Proceedings of the IEEE, Vol. 85, NO. 9, Sept 1997, pp 1365-1380.

[Jain99] A. Jain, R. Bolle, S. Pankanti editors,
"BIOMETRICS Personal Identification in Networked Society",
Kluwer Academic Publishers, 1999.

[Jalkio85] J. Jalkio, R. Kim, S. Case,
"Three dimensional inspection using multistripe structured light",
Optical Engineering, Vol 24, No 6, pp 966-974, 1985.

[Jarvis83a] R. Jarvis,
"A Laser Time-of-Flight Range Scanner for Robotic Vision",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 5, No 5, pp 505-512, Sep 1983.

[Jarvis83b] Jarvis R.A.,
"A perspective on range finding techniques for computer vision",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 5, No 3, pp 122-139, Mar 1983.

[Jarvis93] R. Jarvis,
"Range Sensing for Computer Vision",
In Three-Dimensional Object Recognition Systems, Advances in Image Communication, Vol 1, A.K. Jain and P.J. Flynn (Editors), Elsevier Science Publisher 1993, pp 17-56.

[Jia95] X. Jia, and M. Nixon,
"Extending the Feature Vector for Automatic Face Recognition",
IEEE Trans. on Pattern Analysis and Machine Intelligence,
Vol. 17, No. 12, Dec 1995, pp 1167-1176.

[Kamel93] M. Kamel, H. Shen, A. Wong, R. Campeanu,
"Sytem for the recognition of human faces",
IBM SYSTEMS JOURNAL, Vol. 32, NO 2, 1993, pp 307-320.

[Kamel94] M. Kamel, H. Shen, A. Wong, T. Hong, R. Campeanu,
"Face recognition using perspective invariant features",
Pattern Recognition Letters, Vol. 15 (1994), pp 877-883.

[Kaufman76] G. Kaufman and K. Breeding,
"The Automatic Recognition of Human Faces from Profile Silhouettes",
IEEE Trans. on Systems, Man, and Cybernetics, Vol. 6, No. 2, pp 113-120, Feb 1976.

[Kaya72] Y. Kaya and K. Kobayashi,
"A Basic Study on Human Face Recognition",
Frontiers of Pattern Recognition, S. Watanabe, Ed. New York 1972, pp 265-289.

[Kerin90] M. Kerin, T. Stonham,
"Face Recognition Using a Digital Neural Network with Self-Organising Capabilities",
Int. Conf. on Pattern Recognition, June 1990, pp 738-741.

[Kirby90] M. Kirby and L. Sirovich,
"Application of the Karhunen-Loève Procedure for the Characterization of Human Faces",
IEEE Trans. on Pattern Analysis and Machine Intelligence, VOL. 12, NO. 1, Jan 1990, pp 103-108.

[Kohonen89] T. Kohonen,
"Self-Organization and Associative Memory",
Springer series, 3th edition, 1989.

[Kotro97] C. Kotropoulos, I. Pitas, S. Fisher, B. Duc,
"Face Authentication Using Morphological Dynamic Link Architecture",

FIRST INTERNATIONAL CONFERENCE ON AUDIO AND VIDEO BASED
BIOMETRIC PERSON AUTHENTICATION (AVBPA),
Crans-Montana, Switzerland, Mar 1997, pp 169-176.

[Kurada95] S. Kurada, G. Rankin, K. Sridhar,
“A trinocular vision system for close-range position sensing”,
Optics and Laser Technology, Vol 27, No 2, pp 75-79, 1995.

[Kruger97] N. Krüger, M. Pötzsch, T. Maurer, M. Rinne,
“Estimation of Face Position and Pose with Labeled Graphs”,
British Machine Vision Conference 97.

[Lam98] K.-M. Lam, H. Yan,
“An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 7, Jul 1998, pp 673-686.

[Lapreste88] J.T. Lapresté, J.Y. Cartoux, M. Richetin,
“Face Recognition From Range Data By Structural Analysis”,
NATO ASI series, Vol. F45, Syntactic and Structural Pattern Recognition, Edited by G. Ferraté et al.,
Springer-Verlag Berlin Heidelberg, 1988.

[Laughery89] R. Laughery and M. Wogalter,
“Forensic Applications of facial memory research”,
In A. Young and H. Ellis, editors, “Handbook of research on Face Processing”,
Amsterdam, Netherlands: North-Holland, p. 544, 1989.

[Lavest98] Lavest JM, Viala M, Dhôme M,
“Do we really need an accurate calibration pattern to achieve a reliable camera calibration”,
Proc. of the 5th European Conf. on Computer Vision, 1998, June 2-6, Freiburg, Germany.

[Lee90] J.C. Lee and E. Milios,
“Matching Range Images of Human Faces”,
Proc. IEEE Soc. 3rd Int. Conf. on Computer Vision, 1990, pp 722-726.

[LeMoigne88] J.J. Le Moigne, A.M. Waxman,
“Structured Light Patterns for Robot Mobility”,
IEEE JOURNAL OF ROBOTICS AND AUTOMATION, Vol. 4, No. 5, Oct 1988.

[Lucas97] S.M. Lucas,
“Face Recognition with the continuous n-tuple classifier”
British Machine Vision Conference BMVC'97, Univ. of Essex, 8-11 Sept 1997.

[M2VTS],
“Multi-Modal Verification for Telesurveillance and Security Application”,
European project, ACTS programme (AC102), 1995-1998.

[Manjunath92] B. Manjunath, R. Chellappa, C. von der Malsburg,
“A Feature Based Approach to Face Recognition”,
IEEE Computer Society Conf. on Computer Vision and Pattern Recognition,
Proc. CVPR '92, Champaign, IL, June 15-18, pp 373-378, 1992.

[Mannaert90] H. Mannaert, A. Oosterlinck,
“Self-organizing system for analysis and identification of human faces”,
SPIE Vol. 1349 Applications of Digital Image Processing XIII (1990), pp 227-232.

[Maruyama93] M. Maruyama, S. Abe,
"Range Sensing by Projecting Multiple Slits with Random Cuts",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 15, No 6, June 1993, pp 647-651.

[Medioni03] G. Medioni, R. Waupotitsch,
"Face Modeling and Recognition in 3-D",
Int. Workshop on Analysis and Modeling of Faces and Gestures 2003,
Nice, France, pp 232-233, Oct 2003.

[Monga95] O. Monga, S. Benayoun,
"Using Partial Derivatives of 3D Images to Extract Typical Surface Features",
Computer Vision and Image Understanding, Vol 61, No 2, pp 171-189, Mar 1995.

[Morano98] R. Morano, C. Ozturk, R. Conn, S. Dubin, J. Nissanov,
"Structured Light Using Pseudorandom Codes",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 20, No 3, pp 322-327, Mar 98.

[Mundy91] J. Mundy,
"Model-based vision, an operational reality ?",
Applications of Digital Image Processing XIV, SPIE Vol. 1567, pp 124-141.

[Müller89] E. Müller,
"Fast three-dimensional form measurement system",
Optical Engineering, Vol. 34, pp 2754-2756, Sept 1995.

[Nakamura91] O. Nakamura, S. Mathur, and T. Minami,
"Identification of Human Faces Based on Isodensity Maps",
Pattern Recognition, Vol. 24, No. 3, pp 263-272, 1991.

[Nixon85] M. Nixon,
« Eye Spacing Measurement for Facial Recognition",
SPIE Vol. 575 Applications of Digital Image Processing VIII (1985), pp 279-285.

[Nouri95] Taoufik Nouri,
"Three-dimensional scanner based on fringe projection",
Optical Engineering, Vol 34, No 7, pp 1961-1963.

[NumRecipes],
"Numerical Recipes",
http://www.nr.com.

[Ozeki86] O. Ozeki, T. Nakano, S Yamamoto,
"Real-Time Range Measurement Device for Three-Dimensional Object Recognition",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 8, No 4, pp 550-554, Jul 86.

[Pardas99] M. Pardas,
"Automatic Face Analysis for Model Calibration",
International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging,
Santorini, Greece, Sep 1999, pp 12-15.

[Parkin86] A. Parkin and P. Williamson,
"PATTERNS OF CEREBRAL DOMINANCE IN WHOLISTIC OR FEATURAL STAGES OF FACIAL PROCESSING",
In H. Ellis, M. Jeeves, F. Newcombe and A. Young editors, "Aspects of Face Processing",
Doordrecht, Netherlands: Martinus Nijhoff Publishers, pp. 223-227, 1986.

[Pentland91] A. Pentland, S. Sclaroff,
“Closed-Form Solutions for Physically Based Shape Modeling and Recognition”,
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 13, No 7, pp 715-729, Jul 91.

[Perneel90] C. Perneel and M. Acheroy,
“Face Recognition with Artificial Neural Networks”,
IMACS Annals on Computing and Applied Mathematics,
Proceedings MIMS2 '90, Sept. 3-7, Brussels.

[Perrett86] D. Perrett, A. Mistlin, D. Potter, P. Smith, A. Head, A. Chitty, R. Broennimann,
D. Milner and M. Jeeves,
“FUNCTIONAL ORGANIZATION OF VISUAL NEURONES PROCESSING FACE IDENTITY”,
In H. Ellis, M. Jeeves, F. Newcombe and A. Young editors, “Aspects of Face Processing”,
Doordrecht, Netherlands: Martinus Nijhoff Publishers, pp. 187-198, 1986.

[Pigeon97] S. Pigeon, L. Vandendorpe,
“Profile Authentication Using a Chamfer Matching Algorithm”,
Int. Conf. on Audio- and Video-based Person Authentication (AVBPA'97),
Crans-Montana, Switzerland, March 1997,
Lecture Notes in Computer Science, no 1206, Springer, pp 185-192.

[Pigeon98] S. Pigeon and L. Vandendorpe,
“Image-based Multimodal Face Authentication”,
In Signal Processing, Vol. 69, no. 1, pp. 59-79, October 1998.

[Posdamer82] J.L. Posdamer, M.D. Altschuler,
“Surface Measurement by Space-encoded Projected Beam Systems”,
Computer Graphics and Image Processing, Vol. 18, 1982.

[Proesmans96] M. Proesmans, L. Van Gool,
“Recognition Of Suspects Through 3D Photographs”,
Proc. SPIE 96, Photonics East 1996.

[Proesmans97] M. Proesmans, L. Van Gool,
“One-shot 3D-shape and Texture Acquisition of Facial Data”,
In Lecture notes in Computer Science, Vol 1206, pp 411-418.
Audio- and Video-based Biometric Person Authentication, Crans-Montana, Switzerland, 12-14 Mar 1997.

[Proesmans97b] M. Proesmans, L. Van Gool,
“Reading between the lines - a method for extracting dynamic 3D with texture”,
In ACM Symposium on virtual reality software and technology, VRST 97, pp 95-102, Sept 15-17, 1997
Lausanne, Switzerland.

[Redert02] A. Redert, E. Hendriks, J. Biemond,
“Accurate and robust marker localization algorithm for camera calibration”,
1st Int. Symp. On 3D Data Processing Visualization and Transmission (3DPVT2002), Padova, Italy, Jun 19-21, 2002.

[Reinders92] M.J.T. Reinders, B. Sankur and J.C.A. van der Lubbe,
“Transformation of a general 3D facial model to an actual scene face”,
Int. Conf. on Pattern Recognition, IEEE pp 75-78, 1992.

[Richard99a] G. Richard, Y. Menguy, I. Guis, N. Suaudeau, J. Boudy, P. Lockwood, C. Fernandez, F. Fernandez, C. Kotropoulos, A. Tefas, Ioannis Pitas, R. Heimgartner, P. Ryser, C. Beumier, P. Verlinde, S. Pigeon, G. Matas, Josef Kittler, J. Bigün, Y. Abdeljaoued, E. Meurville, L. Besacier, M. Ansorge, G. Maitre, J. Luettin, S. Ben-Yacoub, B. Ruiz, K. Aldama, J. Cortes,
"Multi Modal Verification for Teleservices and Security Applications (M2VTS)",
IEEE International Conference on Multimedia Computing and Systems, ICMCS 1999, Florence, Italy, 7-11 June, 1999, Proceedings, Volume I. IEEE Computer Society, pp 1061-1064.

[Rocchini01]C. Rocchini, P. Cignoni, C. Montani, P. Pingi, R. Scopigno,
"A low cost 3D scanner based on structured light",
EUROGRAPHICS 2001, Vol. 20, No 3, pp. 299-308.

[Romdhani02] S. Romdhani, V. Blanz, T. Vetter,
"Face Identification by Fitting a 3D Morphable Model using Linear Shape and Texture Error Functions",
European Conference on Computer Vision 2002, Copenhagen, Denmark, pp 3-19, May 2002.

[Rowley98] H. Rowley, S. Baluja, and T. Kanade,
"Neural Network-Based Face Detection",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1, Jan 1998, pp 23-37.

[Saber98] E. Saber, A. Tekalp,
"Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions"
Pattern Recognition Letters, Vol. 19 (1998), pp 669-680.

[Salicetti03] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. Leroux-les-jardins, Jan Lunter, Yang Ni, D. Petrovska-Delacrétaz,
"BIOMET: A Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities",
AVBPA 2003, Guildford, UK, June 9-11, 2003.

[Salvi98] J. Salvi, J. Battle, E. Mouaddib,
"A robust-coded pattern projection for dynamic 3D scene measurement",
Pattern Recognition Letters, 19, pp 1055-1065, 1998.

[Salzen86] E.A. Salzen, E.A. Kostek and D.J. Beavan,
"THE PERCEPTION OF ACTION VERSUS FEELING IN FACIAL EXPRESSION",
In H. Ellis, M. Jeeves, F. Newcombe and A. Young editors, "Aspects of Face Processing",
Doordrecht, Netherlands: Martinus Nijhoff Publishers, pp 326-339, 1986.

[Samal92] A. Samal and P. Iyengar,
"Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey",
Pattern Recognition, Vol. 25, No. 1, pp 65-77, 1992.

[Sato82] Y. Sato, H. Kitagawa, H. Fujita,
"Shape Measurement of Curved Objects Using Multiple Slit-Ray Projections",
IEEE Trans. on Pattern Analysis and Machine Intelligence,
Vol 4, No 6, pp 646, Nov 82.

[Sato86] K. Sato, H. Yamamoto, S. Inokuchi,
"Tuned Range Finder for High Precision 3D Data",
8th Int. Conf. on Pattern Recognition, pp 1168-1171, Paris, France, 0ct 27-31, 1986.

[Saulnier94] A. Saulnier, M.-L. Viaud, D. Geldreich,
"Analyse et Synthèse en temps réel du Visage pour la Télévirtualité",
Imagina 94, pp 173-182.

[Sergent86] J. Sergent,
"MICROGENESIS OF FACE PERCEPTION",
In H. Ellis, M. Jeeves, F. Newcombe and A. Young editors, "Aspects of Face Processing",
Doordrecht, Netherlands: Martinus Nijhoff Publishers, pp. 17-33, 1986.

[Shrikhande89] N. Shrikhande, G. Stockman,
"Surface Orientation from a Projected Grid",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 11, No 6, pp 650-655, Jun 89.

[Stein93] F. Stein, G. Medioni,
"Structural Indexing: Efficient Three Dimensional Object Recognition",
In "Three Dimensional Object Recognition Systems", Advances in Image Communication, Vol. 1.

[Stonham86] T. Stonham,
"Practical Face Recognition and Verification with Wisard",
In Aspects of Face Processing, Doordrecht, NL: Martinus Nijhoff Publishers, pp 426-441, 1986.

[Strand85] T. Strand,
"Optical three-dimensional sensing for machine vision",
Optical Engineering, Vol 24, No 1, pp 33-40, 1985.

[Sung98] K.-K. Sung and T. Poggio,
"Example-Based Learning for View-Based Human Face Detection",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1, Jan 1998, pp 39-50.

[Suzuki95] S. Suzuki, Y. Tatsuno, N. Yokoya, H. Iwasa and H. Takemura,
"Analysis and Synthesis of Human Facial Expression Using Range Images",
Second Asian Conference on Computer Vision, ACCV'95, Dec 1995, Singapore.

[Tajima90] J. Tajima, M. Iwakawa,
"3-D Data Acquisition By Rainbow Range Finder",
Proc. ICPR, pp. 309-313, 1990.

[Turk90] M. Turk, and A. Pentland,
"Recognition in Face Space",
SPIE Vol. 1381 Intelligent Robots and Computer Vision IX: Algorithms and Techniques (1990), pp 43-54, 1990.

[Turk91] M. Turk and A. Pentland,
"Eigenfaces for Recognition",
Journal of Cognitive Neuroscience, Vol. 3, No 1, pp 71-86, 1990.

[Vuylsteke90] P. Vuylsteke, A. Oosterlinck,
"Range Image Acquisition with a Single Binary-Encoded Light Pattern",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 12, No 2, pp 148-164, Feb 1990.

[Wang87] Y. Wang, A. Mitiche, J. Aggarwal,
"Computation of Surface Orientation and Structure of Objects Using Grid Coding",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 9, No 1, pp 129-136, Jan 1987.

[Wang89] Y.F. Wang, P. Liang,
"A new method for Computing Intrinsic Surface Properties",
Proc. Computer Vision and Pattern Recognition CVPR89, pp 235-240, June 4-8, 1989.

[Wang91] Y. Wang,
"Characterizing Three-Dimensional Surface Structures from Visual Images",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 13, No 1, Jan 1991.

[Wang02] Y. Wang, C-S Chua, Y-K Ho,
"Facial feature detection and face recognition from 2D and 3D images",
Pattern Recognition Letters, vol. 23, pp 1191-1202, 2002.
[Weng92] J. Weng, P. Cohen, M. Herniou,
"Camera Calibration with Distortion Models and Accuracy Evaluation",
IEEE Trans. on Pattern Analysis And Machine Intelligence, Vol.14, No. 10, Oct 92.

[Wilder93] J. Wilder, S. Juth, A. Tsai, X. Zhang,
"Face Recognition Using The Neural Tree Network",
SPIE 2093, Proceedings Europto Substance Identification Analytics, 4-8 Oct 1993.

[Wildes97] R.P. Wildes,
"Iris Recognition: An Emerging Biometric Technology",
Proceedings of the IEEE, vol. 85, No. 9, pp 1348-1363, Sept 1997.

[Wiskott96] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg,
"Face Recognition by Elastic Bunch Graph Matching",
Institut für Neuroinformatik, Universit\"at Bochum, FRG, Internal Report 96-08.

[XM2VTSDB],
"The Extended M2VTS Database",
http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb.

[Yacoob94] Y. Yacoob, L. Davis,
"Labeling of Human Face Components from Range Data",
CVGIP: Image Understanding, Vol. 60, No 2, pp 168-178, Sep 1994.

[Yamamoto86] H. Yamamoto, K. Sato, S. Inokuchi,
"Range Imaging System Based on Binary Image Accumulation",
8th Int. Conf. on Pattern Recognition, pp 233-235, Paris, France, 0ct 27-31, 1986.

[YangM02] M-H Yang, D.J Kriegman, N. Ahuja,
"Detecting Faces in Images: A survey",
IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, Jan 2002, pp 34-58.

[Yee94] R. Yee, P. Griffin,
"Three-dimensional imaging system",
Optical Engineering, Vol 33 No 6, Jun 1994.

[Young89] A. Young and H. Ellis, editors,
"Handbook of research on Face Processing",
Amsterdam, Netherlands: North-Holland 1989.

[Yow97] K. Yow, R. Cipolla,
"Scale and Orientation Invariance in Human Face Detection",
British Machine Vision Conference 1997.

[Yuille89] A. Yuille, D. Cohen, and P. Hallinan,
"Feature extraction from faces using deformable templates",
Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, June 4-8, 1989, pp 104-109.

[Zhang02] L. Zhang, B. Curless, S.M. Seitz,
“Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming”, 1st Int. Symposium on 3D Data Processing, Visualization and Transmission (3DPVT’02), June 19-21, 2002, Padova, Italy.

3D analysis

[Asada86] [Asada88] [Bernardini02] [Brechb95] [Chin86]
[Flynn89] [Hoffman87] [Hurt84] [Ittner85] [Monga95]
[Pentland91] [Shrikhande89] [Wang87] [Wang89] [Wang91]

3D capture

[Battle98] [Blake93] [Boyer87] [Carter90] [Carrihill85]
[Caspi98] [Chen93] [Chen97] [Forster01] [Gärtner96]
[Geng96] [Griffin92] [Guisser92] [Hall-Holt01] [Hata92]
[Horn99] [Hsieh98] [Hu89] [Jalkio85] [Jarvis83a]
[Jarvis83b] [Jarvis93] [Kurada95] [LeMoigne88] [Maruyama93]
[Morano98] [Müller89] [Nouri95] [Ozeki86] [Posdamer82]
[Proesmans96] [Proesmans97] [Proesmans97b] [Rocchini01] [Salvi98]
[Sato82] [Sato86] [Strand85] [Tajima90] [Vuylsteke90]
[Yamamoto86] [Yee94] [Zhang02]

3D Face Recognition

[Achermann97] [Achermann00] [Cartoux89] [Chang97] [Chang03]
[Chua00] [Gordon91] [Gordon92] [Gordon92b] [Lapreste88]
[Lee90] [Medioni03] [Romdhani02] [Wang02] [Yacoob94]

Biometry

[Acheroy96a] [Richard99a] [Salicetti03] [Wildes97]

Calibration

[Faugeras86] [Gardel03] [Lavest98] [Redert02] [Weng92]

Face detection

[Rowley98] [Saber98] [Sung98] [YangM02] [Yow97]

Face modelling

[Aizawa95] [Akimoto93] [Bozdagi94] [Ekman72] [HuangC93]
[Huang93] [Kruger97] [Reinders92] [Saulnier94] [Yuille89]

Face recognition: featural

[Cox95] [Craw87] [Jia95] [Kamel93] [Kamel94]

[Kaya72] [Manjunath92] [Mannaert90] [Nixon85] [Wiskott96]

Face recognition : holistic

[Bartlett98] [Beymer93] [Etemad97] [Goudail96] [Kerin90]
[Kirby90] [Lucas97] [Nakamura91] [Perneel90] [Stonham86]
[Turk90] [Turk91] [Wilder93]

Face recognition : Miscellaneous

[Augusteijn93] [Brunelli93] [Brunelli95] [Chellappa95] [Hu96]
[Lam98] [Pigeon98] [Samal92]

Face recognition : profile

[Aibara92] [Beumier95a] [Beumier97d] [Galton10] [Harmon78]
[Harmon81] [Kaufman76] [Pigeon97]

Publications related to this thesis

3D capture

[Beumier99b] [Beumier03a]

3D Face recognition

[Beumier97b] [Beumier98a] [Beumier98b] [Beumier00a] [Beumier00c]
[Beumier00f] [Beumier01a] [Beumier01b]

Biometry

[Acheroy96a] [Richard99a] [Beumier99c] [Salicetti03]

Face recognition

[Beumier95a] [Beumier96b] [Beumier97d] [Beumier00d] [Beumier01c]

Technical reports

[Beumier97a] [Beumier97c] [Beumier98e]

Others

[Beumier98c] C. Beumier,
« Face Recognition at the SIC”,
CVSSP External Seminars 98/99 in Surrey (UK), Oct 13th 1998.
No paper.

[Beumier98d] C. Beumier,
“Automatic Profile Identification”,
One Day BMVA Technical Meeting: 'Personal Identity Verification',
Oct 14th 1998, London UK.
No paper.

Publications on other topics :

[Villers95b] D. Villers, C. Fougnies, L. Paternostre, C. Beumier, M. Dosière,
“The use of an imaging plate as a detecting system in X-ray diffraction of polymers”,
In *Nuclear Instruments and Methods in Physics Research*, B97 1995, pp. 265-268.

[Beumier00b]
C. Beumier, M. Acheroy,
``Motion estimation of a hand-held mine detector",
Signal Processing Symposium, Hilvarenbeek, The Netherlands, 23-24th March 2000.

[Beumier00e] C. Beumier, P. Druyts, Y. Yvinec, M. Acheroy,
"Real-Time Optical Position Monitoring using a Reference Bar",
Signal Processing and Communications (SPC2000), IASTED International Conference, Marbella, Spain, Sept 19-22, 2000, pp 468-473.

Appendix A Biometry

Interesting pages on the internet:

http://online.biometry.com
http://www.cbel.com/biometrics_security/
http://www.biometrics.org/
http://biometrie.online.fr/

Among biometry, European Projects with Face:

Multi-modal Verification for Tele-services and Security Applications (M2VTS)

The objective is to address the issue of secured access to local and centralised services in a multi-media environment thanks to a multi-modal (speech, image, …) approach bringing robustness and increased performance over individual technologies.

A WALK-BY BIOMETRIC IDENTIFICATION SYSTEM BASED ON FACE RECOGNITION (WABY)

The objective is to validate 'walk-by' biometric identification system, realised by a combination of real-time face recognition and hands-free RF card technology.
Using existing face recognition technology (ZN-Face by ZN GmbH) for still images and face recognition technology (FaceSnap by C-VIS GmbH) for real-time video.
Long range RF card reader by Nedap NV.
Supporting end-users are Rabobank and Schiphol Airport.

Biometric Access Control for Networked and e-Commerce Applications (BANCA)

The objective of this project is to develop and implement a complete secured system with enhanced identification, authentication and access control schemes for applications over Internet such as tele-working and Web-banking services.

User friendly face access control system for physical access and healthcare applications (UFACE)

To develop and demonstrate user friendly secure access control in financial services and healthcare. A facial biometric will be combined with a smart-card to create a personalised token.

High Speed 3D- and Colour Interface to the Real World (HISCORE)

HISCORE will integrate a new approach to high speed 3-Dimensional image acquisition for the very important mid- distance range and build an affordable real-time 3D and colour camera. This system will be validated in two application areas: face recognition with relevance to access control and security applications, and hand gesture recognition with relevance to new human-computer interfaces.

Some face recognition product providers:

AGMA Morpho Co. (F)
ASTRON Advanced Technology Solutions (PL)
Aurora Computer Services Ltd (UK)
Axis Software pvt ltd (India)
Biometric Access Company (BAC, Texas, USA)
BS-Control (Germany)
C-VIS Computer Vision and Automation GmbH (Germany)
DCS (Berlin, Germany)
Eyematic (USA)
FACE Technologies (South Africa)
Fraunhofer IIS (Germany)
ImageWare Software, Inc. (USA)
IMAGIS Technologies Inc.
Malin Systems (UK)
Miros
Neusciences Biometrics (UK)
Plettac electronics (Germany)
Sintec Image recognition system (USA)
Unidas (Germany)
Visec FIRE (Germany)
Visionics Corporation (USA)
Visionsphere Technologies Inc. (Canada)
Visual Automation (UK)

Appendix B Label sequence

B.1 Introduction

The objective of the label sequence is to allow identification of the items possessing the labels. As the number of items is rather large, and the coding possibility of each item is limited, unique identification is realised with code distribution among neighbour items.

We present in this appendix the algorithms developed to build up label sequences for the slide designs used by the different prototypes discussed in Chapter 2. Other sequences can be designed with the same algorithm or easily created from partial sequences or repetitions of obtained sequences.

The problem of sequence creation with unique neighbouring n-tuple of m-symbol can also make use of de Bruijn sequences [Griffin92, Hsieh98]. Basically, any pseudorandom code generation is welcome, and the approach presented in [Morano98] has the additional advantage to introduce a constraint on Hamming distance between successive elements so that some errors can be detected and corrected.

B.2 Binary sequence with unique words

We present in this section the algorithm to create a binary sequence so that words (of n successive bits) are unique in the sequence.

Algo:

CreateSequence

```
a) Use first unused word as starting sequence (length = n bits). Mark in list that word as
used.
b) Add bit 0 to sequence. If word with last n bits is not used, mark it as used and go to b)
c) Add bit 1 to sequence. If word with last n bits is not used, mark it as used and go to b),
else return the sequence to caller.
```

Create the longest sequence

```
Mark all possible words (2 exp n) except 0 as unused. currentSequence is '00..0'.
While unused word exists
        Call CreateSequence which returns newSequence and get first n-1 bits
        Find n-1 bit pattern in current list
        Insert newSequence in currentSequence where n-1 bit pattern was found
```

The property that is used to create longer sequences is that each sequence created with CreateSequence begins and ends with the same n-1 bits (check in the example). Because these n-1 bits stopped the sequence, the two possibilities (0 and 1) for the nth bit were used, so that the same n-1 bits appear in the current sequence. It is thus possible to insert the new created sequence in the current sequence because the n-1 bits seen from the left and right of the insertion are the same as before insertion.

Let us detail the algorithm for n = 4.

Words range from 0 to 15, all marked unused at initialisation.
The table is empty, except 0. currentSequence is initialised with 0000.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
X															

First unused is 1 (0001), CreateSequence fills the table
(succession of added bits: (0001), 0, 0, 0, stop because 0000 and 0001 already used.)

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	X	X		X				X							

and returns sequence 0001000
n-1=3 first bits are 000 which are found at position 0 of currentSequence
Insertion of returned sequence gives new currentSequence 000-1000-0

First unused is 3 (0011), CreateSequence fills the table ((0011), 0, 0, 1, stop (2 and 3 already used))

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	X	X	X	X		X		X	X			X			

Returned sequence is 0011001.
001 is in current list at position '2'.
Insertion leads to currentSequence 0001-1001-0000

First unused is 5 (0101), CreateSequence fills the table with ((0101), 0, stop)

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	X	X	X	X	X	X		X	X	X		X			

and returns 01010 which is inserted where 010 appears in currentSequence to give new currentSequence: 000110010-10-000

First unused is 7 (0111), table will be filled

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	

The returned sequence: 0111 0 1 1 is inserted where 011 occurs.
New currentSequence: 00011-1011-001010000

First unused is 1111 which is the last returned sequence, inserted where 111 occurs
The current and full sequence is 000111-1-011001010000. This corresponds to the word sequence 1, 3, 7, 15, 14, 13, 11, 6, 12, 9, 2, 5, 10, 4, 8, 0.

We take the opportunity of this example to highlight some ***properties***.

1. The sequence contains all the possible (2^n) words, so that the binary sequence is 2^n + n-1 bit long.

2. The binary sequence (0001111011001010000) is cyclical: the sequence can be started everywhere, which is useful to skip some parts (for instance 1111 or 0000).

3. The binary sequence is not unique, even when starting with the same word (cyclic). There are many other solutions, like the inverted order or the complementary values (0 and 1 inverted). Insertion of subsequences can sometimes be made at different places.

We give hereafter the binary sequences for different n values.

n = 5
0 1 3 7 15 31 30 29 27 23 14 28 25 19 6 12 24 17 2 5 11 22 13 26 21 10 20 9 18 4 8 16

000001111101110011000101101010010000

n = 6
0 1 3 7 15 31 63 62 61 59 55 47 30 60 57 51 39 14 28 56 49 35 6 13 27 54 45 26 52 41 19 38 12 24 48 33 2 5 11 23 46 29 58 53 43 22 44 25 50 37 10 21 42 20 40 17 34 4 9 18 36 8 16 32

000000111111011110011100011011010011000010111010110010101000100100000

n = 7
0 1 3 7 15 31 63 127 126 125 123 119 111 95 62 124 121 115 103 79 30 60 120 113 99 71 14 29 58 116 105 83 39 78 28 56 112 97 67 6 13 27 55 110 93 59 118 109 91 54 108 89 51 102 77 26 52 104 81 35 70 12 24 48 96 65 2 5 11 23 47 94 61 122 117 107 87 46 92 57 114 101 75 22 44 88 49 98 69 10 21 43 86 45 90 53 106 85 42 84 41 82 37 74 20 40 80 33 66 4 9 19 38 76 25 50 100 73 18 36 72 17 34 68 8 16 32 64

000000011111110111110011110001110100111000011011101101100110100011000
00101111010111001011000101011010101001010000100110010010001000000

n = 8
0 1 3 7 15 31 63 127 255 254 253 251 247 239 223 191 126 252 249 243 231 207 159 62 124 248 241 227 199 143 30 61 122 244 233 211 167 79 158 60 120 240 225 195 135 14 29 59 119 238 221 187 118 236 217 179 103 206 157 58 116 232 209 163 71 142 28 56 112 224 193 131 6 13 27 55 111 222 189 123 246 237 219 183 110 220 185 115 230 205 155 54 108 216 177 99 198 141 26 53 106 212 169 83 166 77 154 52 104 208 161 67 134 12 25 51 102 204 153 50 100 200 145 35 70 140 24 48 96 192 129 2 5 11 23 47 95 190 125 250 245 235 215 175 94 188 121 242 229 203 151 46 92 184 113 226 197 139 22 45 91 182 109 218 181 107 214 173 90 180 105 210 165 75 150 44 88 176 97 194 133 10 21

43 87 174 93 186 117 234 213 171 86 172 89 178 101 202 149 42 85 170 84 168 81 162
69 138 20 40 80 160 65 130 4 9 19 39 78 156 57 114 228 201 147 38 76 152 49 98 196
137 18 37 74 148 41 82 164 73 146 36 72 144 33 66 132 8 17 34 68 136 16 32 64 128

000000001111111101111110011111000111101001111000011101110110011101000
111000001101111011011100110110001101010011010000110011001000110000001
011111010111100101110001011011010110100101100001010111010101100101010
10001010000010011100100110001001010010010000100010000000

6-uple sequence with unique pairs

The creation of n-tuple sequences follows the same algorithm but is more complex. First, the number of possibilities (n) to check is larger and secondly, special cares help distributing the n labels more uniformly.

L \ R	**0**	**1**	**2**	**3**	**4**	**5**
0	30	0	6	12	19	23
1	18		1	8	14	27
2	22	7		2	25	16
3	29	13	21		3	9
4	11	26	15	20		4
5	5	17	24	28	10	

Table 16: Order of attribution

Prepare Table 16 by setting all entries as unused.
Prepare a table 'numT' with 6 entries to keep track for each label of the number of occurrences. These entries are initialised with 0.

We start with label 0. We increment numT[0].

We scan numT and get the first entry with minimal occurrence. It is numT[1] = 0 because numT[0] = 1. We mark [row0, col1] as used. We increment numT[1]. Sequence is 01.

We scan numT and get 2 (numT[2] = 0). 12 is marked and numT[2] incremented. Sequence is 012.

The process is repeated until the current sequence is
"0123450213540314251043205241530"
and Table 16 is nearly full as depicted.
It was obtained in one cycle. We have to add pairs 11, 22, 33, 44, 55 by inserting 11 where there is a 1, 22 where there is a 2, etc.

We finally obtain, for instance, "0112345021335403142251044320552415300".
The solution is not unique.

Appendix C Prototype A

C.1 *Foreword*

This annex contains the report that was written by the time of the developments of Prototype A, in 1998. The text has been slightly modified to conform to the notations adopted in the presentation of Chapter 2.

C.2 *System design*

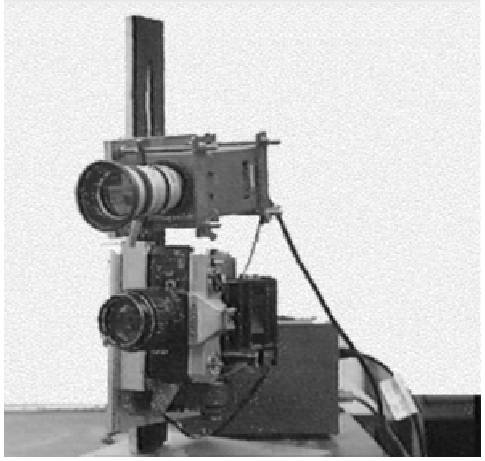


Figure 60: Hardware prototype A

The system captures 3D surfaces thanks to the structured light principle. A slide projector projects a light pattern on the scene captured by a camera. The position of pattern elements in the image allows for the determination in the space of the 3D position of the underlying surface by triangulation.

C.2.1 Camera

A normal CCD camera, inducing little geometric distortion and providing for 500 lines, fulfils our resolution and quality requirements. The light pattern has been designed to work with black and white cameras. We used the Panasonic BL 600 model. The image acquisition board (Matrox Meteor) with the same resolution (768 x 576 pixels) has been retained as an acceptable and low cost solution.

We opted for a 25 mm camera lens. With a smaller focal length, the distortion usually becomes important. A longer focal length implies less magnification and thus smaller areas of the scene visible by the camera. With a focal length of 25 mm and working at a distance of 150 cm, the magnification is 60. As the sensor is 1/4 inch (0.635 cm), the visible area is 15 inches (38 cm) long, what is comfortable as heads are about 25 cm high.

C.2.2 Projector

A normal projector for 24x36 slides has been used. Its low price and the facility to create slides for it have simplified the integration. Several slides with different stripe densities have been tried. The selected slide has 134 (parallel) stripes, half of which are thin and the other half thick. Stripe thickness helps coding the identity of each stripe among seven neighbour stripes.

A 150 W white light lamp is used. This lamp gives a very good contrast but is too powerful. It dazzles the individual and requires a fan cooler.

The slide for projection is 24 x 36 mm. This allows for rather a high focal length because the needed magnification is not so high. Working at a distance of 150 cm with a 100 mm lens implies a magnification of 15, projecting light on a rectangle of 36 x 54 cm.

C.2.3 Slide design

We designed our own stripe pattern as a 1-dimensional set of parallel stripes consisting of thin and thick lines. This choice was motivated by the continuity of the stripes in one direction leading to an easy detection and high resolution in this direction.

First, the thick/thin stripe arrangement had to be produced. We wrote a program that delivers a bit pattern so that any sequence of n (here, 7) bits only appears once. The binary information is encoded in the width of the stripes (Figure 61). The ratios 1/3 and 2/3 for thin and thick stripes have been chosen for the balanced (50 %) light transmission and the high thin / thick immunity (the ratio of black and white widths is 2 for thick and 1/2 for thin). A constant distance between left edges of stripes eases the stripe to angle conversion.

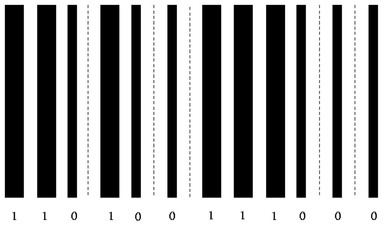


Figure 61: Part of the slide pattern

Such a slide design is compatible with a monochromatic projection, and thus with infra-red lighting, which allows for the discrete projection of invisible light. Grey-level hardware is also less expensive. The density of stripes for a projection at about 1m20 is more than 2 stripes per cm, which corresponds to a slide with about 100 stripes. We thus selected stripe identity encoding on 7 bits (134 stripes).

C.2.4 Slide production

We first considered printing the pattern on a slide thanks to a laser printer what resulted in poor quality (accuracy and noise). We then took a picture with a slide film of a computer screen displaying the slide pattern. The resulting slide was better, although its translucence is not optimal, what limits the contrast. We then discovered the necessity to set a plastic or glass support to prevent the slide from warping.

Laser scanning was considered for the advantages of accuracy, high contrast and larger depth of focus, but the solutions seemed too slow. CGH (Computer Generated Holograms) could help projecting the complete pattern with laser light without scanning but the design for our pattern is difficult and the projection would be subject to depth of focus.

We finally used a microelectronics solution to produce a high quality stripe pattern on glass. This brought the advantages of a high accuracy and opacity of the stripes, high translucence and little deformation of glass with heat or time.

C.2.5 Discussion

For a system with the camera and projector characteristics mentioned before, Table 17 gives values concerning the field of view ('view' relates to the rectangle seen by the camera and 'proj' to the rectangle of projected light, both at the distance of work, 'distance') and image resolution ('Pixels in face' and 'Pixels per stripe') obtained for different distances of work and two stripe densities (134 and 200 stripes on the slide). The numbers of 'Pixels in the face' and 'Stripes in the face' were computed for a facial size of 20 cm.

Distance [cm]	View horiz [cm]	View vertic [cm]	Proj horiz [cm]	Proj vertic [cm]	Stripes in face (134)	Stripes in face (200)	Pixels in face	Pixels/stripe (134)	Pixels/stripe (200)
125	31.7	23.8	45	30	60	89	485	8.1	5.5
150	38.1	28.6	54	36	50	74	403	8.1	5.5
175	44.4	33.3	63	42	43	63	346	8.1	5.5
200	50.8	38.1	72	48	37	56	302	8.1	5.5

Table 17: Sets of field of view and stripe resolution on image for different distances

The stripe resolution in pixels in the image is computed here in the case of horizontal projection. Due to the adopted diagonal striping (see next subsection), this number is less (about 25 % less, corresponding to an inclination of 40°). We see from the table that the stripe width on the image is independent from the distance of work. For a larger distance, the stripes will be wider on the object, but the object is seen smaller from the camera. So, for a given setting of the system (lenses and slide), mainly the surface orientation has an influence on the image stripe width.

With the given focal lengths, projected light power and working distance at 1m50, the depth of focus is about 30 cm and the field of acquisition covers 40 cm x 30 cm. This is compatible with the cooperative scenario for a sitting attitude.

C.2.6 Sitting attitude

Sitting on a chair has the first advantage to properly place the person in the field of view and field of focus of the acquisition system. A second advantage is to reduce the deviation of the population height (probably dividing by two the variation). However, the way people sit accounts for a big proportion in the height variation and positioning so that user cooperation is also required in that respect.

As an evaluation of the gain for a sitting scenario, let us consider Table 18. In this table, several cases of human height ranges are considered (minH - maxH). Assuming the head size is one seventh of people's height, the level of the chin for the taller and smaller persons are derived. Expecting that the variation in a sitting attitude ('delta (sit)') is half the variation in height ('delta'), the 'Needed range' to grab the whole face of people of height in the range minH - maxH is the variation 'delta (sit)' + head size (supposed to be height/7).

MinH [cm]	MaxH [cm]	Min chin (minH-1/7)	Max chin (maxH-1/7)	Delta [cm]	Delta sit [cm]	Needed range
150	200	129	171	42	21	50
155	195	133	167	34	17	45
160	190	137	163	26	13	40

Table 18: The needed range to grab the whole face for sitting people of height in the range minH-maxH

C.2.7 System setup

Because of the compromise between resolution and volume of acquisition, we opted for a sitting attitude. This is compatible with the cooperative scenario. It ensures that the distance with the 3D acquisition system is within focus and it reduces vertical variation between people so that the head is within the field of view.

The camera was turned 90° (portrait) to benefit from the larger horizontal resolution along the height of the face. The higher frequency content in the vertical direction of the face (due to mouth and eyebrows) and the stripes projected parallel to the vertical direction of the camera are also better preserved by the larger horizontal bandwidth of video devices.

The 3D system (camera and projector) was rotated of about 40°. Indeed, horizontal stripes on the face induce detection problems in most eyebrows, eyes, noses and mouths. And vertical stripes imply a large difference in the stripe density between the left and right parts of the face (Figure 62a). Keeping a balanced resolution between left and right looked to us important to take advantage of the vertical symmetry of the head. As the head is roughly oval, it has more space to fit rotated in the rectangular area of capture.

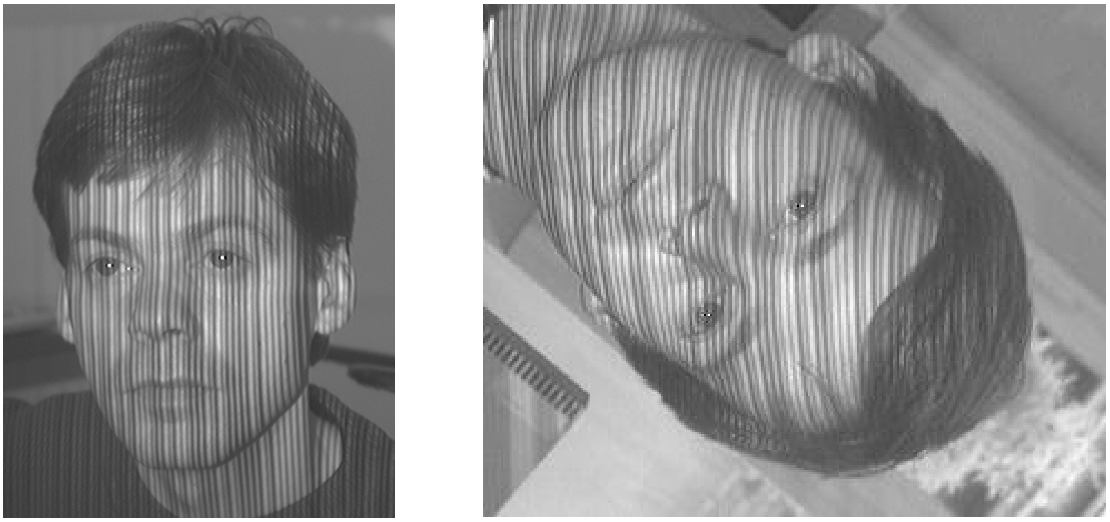


Figure 62: a) Image with vertical striping b) Image with diagonal striping

C.3 *Image to 3D conversion*

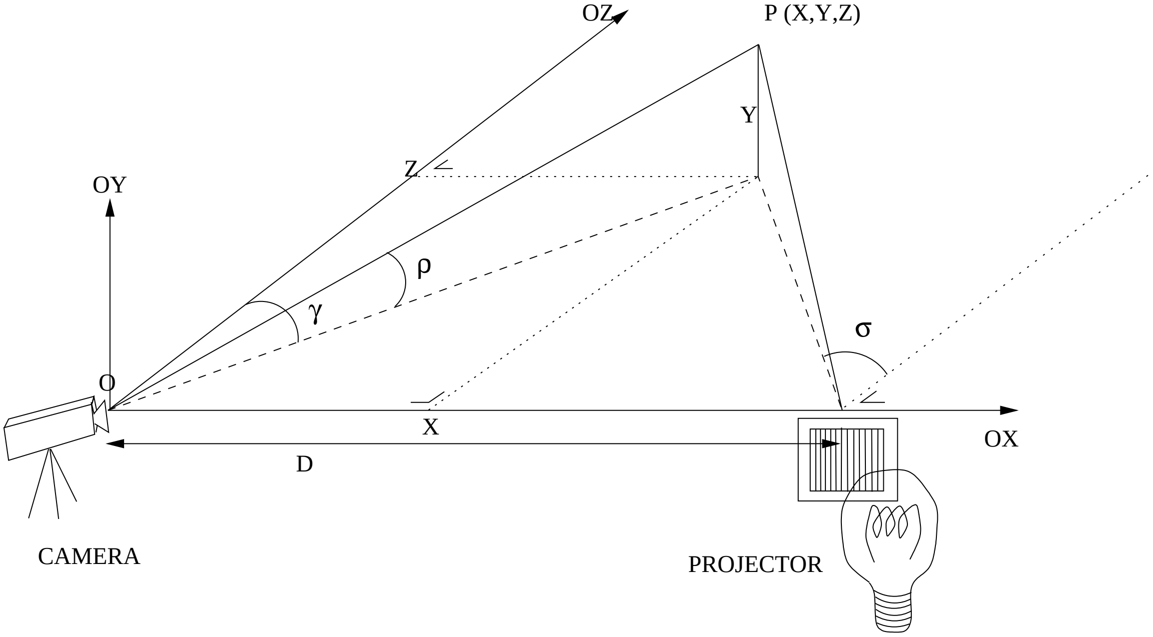


Figure 63: Axis system for structured light

The camera and the projector have been fixed on a rail so that their optical axes are coplanar (plane Y=0). Axis X passes through the camera and projector focal points. Axis Z is perpendicular to X and in plane Y=0. In the case of coplanar optical axes, the number of parameters is reduced and we have:

$$\left.\begin{aligned} Z &= D / (\tan(\gamma) + \tan(\sigma)) \\ X &= Z * \tan(\gamma) \\ Y &= Z * \tan(\rho) / \cos(\gamma) \end{aligned}\right\} \qquad \textbf{Equations 3}$$

With

$$\gamma = \gamma_0 + \mathrm{atan}(x / f_{cx})$$
$$\rho = \rho_0 + \mathrm{atan}(y / f_{cy})$$
$$\sigma = \sigma_0 + \mathrm{atan}(s / f_{px})$$

and

$$x = c - c_0$$
$$y = r - r_0$$
$$s = stripe - s_0$$

where

D is the distance between the camera and the projector focal point,
ρ, γ, σ are angles as depicted on Figure 63,
ρ_0, γ_0 identify the direction of the camera optical axis,
σ_0 identifies the horizontal direction of the projector optical axis,
f_{cx} (f_{cy}) is the “horizontal (vertical) camera focal length” in pixel unit,
f_{px} is the “horizontal projector focal length” in stripe unit,
c, r are horizontal and vertical coordinates relative to the upper left image corner,
x, y are image coordinates relative to the image centre (c_0, r_0),
stripe is the stripe index relative to the leftmost stripe of the pattern,
and s is the stripe index relative to the slide centre (s_0).

We refined the model to take optical distortion into account. Following the hypothesis of an isotropic (circular) deformation, the image coordinates x, y of Equations 3 were modified:

$x' = (c - c_0) * (1 + K_{dc} * dist^2)$
$y' = (r - r_0) * (1+ K_{dc} * dist^2)$

where ***dist*** is the distance of the point to the image centre :

$dist^2 = (c\text{- } c_0)^2 + (r\text{- } r_0)^2$

These equations are valid for a stripe pattern perpendicular to the axis plane. Stripes parallel to that plane do not bring depth information, but would reduce the quantity of transmitted light and the visibility (and facility of detection) of other stripes.

C.4 *Calibration*

C.4.1 Rough measurements

The first step of calibration consists in measuring distances between the camera, the projector and an object point with focused stripes. This allows to estimate the distance between the camera and the projector (*D*) and the angles ($\gamma_{0,}$ σ_0) of the optical axis of the camera and the projector in the axis plane (see Figure 63).

The pixel to angle conversion factor of the camera, dependent on the CCD, the digitisation board and the camera lens, is modelled by 2 parameters (f_{cx} horizontally and f_{cy} vertically). The stripe index to angle conversion factor is modelled by parameter f_{px} and depends on the slide and the projector lens. Focusing the camera or the projector thus needs to recalibrate the system. Calibration is performed for a specific distance of work. At other distances, stripes get blurred and cannot be detected properly.

Simple trigonometry leads to the following procedure for the initial estimate of the parameters:

1. O_o is an object point near optical axes (lit by a central stripe and projected near the image centre), O_c is the optical centre of the camera and O_p is the optical centre of the projector. Each side of the triangle $O_oO_cO_p$ is measured by a ruler, which gives γ_0, σ_0 and D (Pythagorus):

$$D = \mathrm{d}(O_c,O_p)$$
$$\cos(\gamma_0) = (D^2 + \mathrm{d}^2(O_o,O_c) - \mathrm{d}^2(O_o,O_p)\ )\ \ /\ \ (2 * D * \mathrm{d}(O_o,O_c)\ )$$
$$\cos(\sigma_0) = (\ \mathrm{D}^2 + \mathrm{d}^2(O_o,O_p) - \mathrm{d}^2(O_o,O_c)\ \ /\ \ (2 * D * \mathrm{d}(O_o,O_p)\ )$$

2. The screen rectangle (size *W* and *H*) seen by the camera from a known distance *d1* lets determine f_{cx} and f_{cy}. If *WID* and *HEI* are the size (in pixel) of the digitised image:

$$f_{cx} = (d1 * \mathrm{WID}) / \mathrm{W}$$
$$f_{cy} = (d1 * \mathrm{HEI}) / \mathrm{H}$$

(CCDs with square pixels normally imply $f_{cx} = f_{cy}$, if the acquisition board also has square pixels).

3. f_{px} is obtained by projecting the slide on a screen at a known distance. We count the number of visible stripes *N_STRP* and we measure the distance *d2* from the screen to the projector and the width *W'* of the projected pattern.

$f_{px} = (d2 * N\_STRP) / \mathrm{W'}$

C.4.2 Refined calibration

The initial parameter values obtained according to the previous section are refined to measure distances as accurately as possible on a reference object (see Figure 64). We realised a white square of 150 x 150 mm as the reference object from which to obtain points (the square corners) with known relative distances.

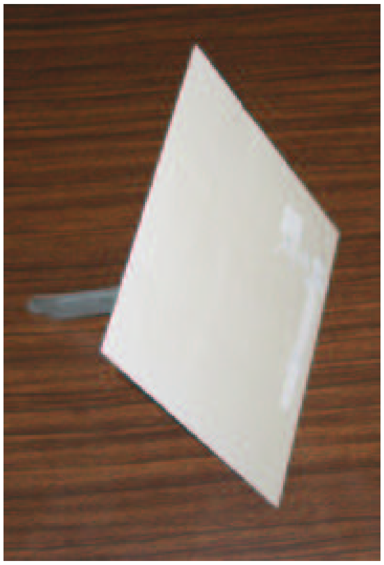


Figure 64: Reference object

The white colour of the square provides a large contrast to ease automatic border detection from edge detection. The four corners are localised as the intersections of the borders (see Figure 65). The rather ideal light pattern seen on the square eases the labelling of the stripes.

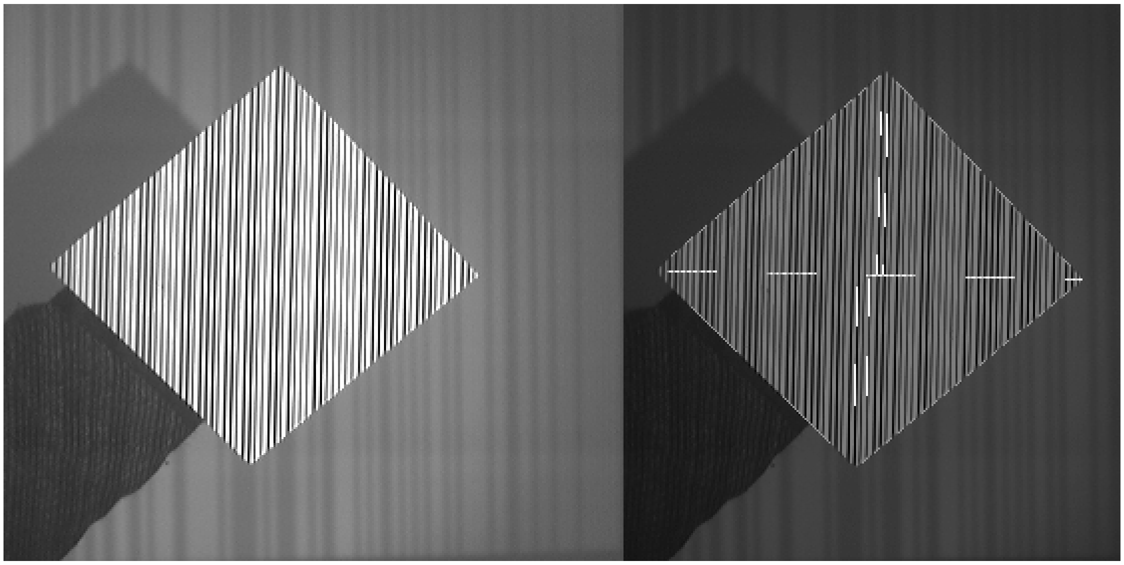


Figure 65: Reference object capture and corner detection

The basically linear dependence between the horizontal stripe position and the stripe index (for instance along the line joining the left and right corners) allows interpolation to get the decimal stripe index of the left and right corners. For the top and bottom corners, stripe index determination requires the vertical orientation of the stripes, what is estimated by following left and right stripes around the top and bottom corners. The output of corner localisation is four sets of (x,y) positions with their decimal stripe index.

Several shots with different plane orientations and positions in the field of view are necessary to have a better estimation of the parameters and to better approximate the possible distortion spread in the field of view. The refinement procedure tries combinations of parameter values (starting from the initial estimates) to minimise a distance error and a planarity error for each shot separately. Absolute positioning is not optimised (we have no clear and precise reference), only the metric is made more homogeneous, for better (relative) distance measurements. The distance error is the mean of the difference between the length of the sides and of the diagonals (divided by $\sqrt{2}$) and *150 mm*. The planarity measure is the distance from a corner to the plane defined by the three other corners.

The optimisation program starts from initial guesses of γ_0, f_{cx}, f_{cy} and f_{px}. To keep the number of parameters small, ρ_0 has been frozen to 0.0, because it mainly corresponds to a translation in y, which does not influence the error measure significantly. σ_0 has been linked to γ_0 to avoid divergence because $\sigma_0+\gamma_0$ constant also mainly implies a translation. Finally, D only introduces a scale factor to measure real centimetres, and does not intervene in the optimisation loop.

The optimisation loop considers ranges of about 20% for each parameter value. For the lowest error, a new iteration is started from the best values and with search ranges reduced by 2. The process stops when the error does not decrease significantly anymore.

Mention that f_{cx}, f_{cy}, f_{px} are intrinsic parameters, bound to the camera/digitiser or slide/projector. γ_0, σ_0, ρ_0 and D are extrinsic parameters, bound to the camera/projector arrangement. If the lenses are not modified (defocused), intrinsic parameters should be constant, what can help calibration.

C.4.3 Calibration results

Table 19 shows measured distances (in mm) after refined calibration, corresponding to sides and diagonals (divided by √2) of 7 shots of the square reference object. The *'Mean dist'* is the average difference of the 6 distances with 150 mm. The '*Planarity*' is the signed distance (convex/concave) of a square corner to the plane defined by the three other corners. The *'Max dist'* is the maximal deviation of the square sides and diagonals (divided by √2) with 150 mm.

	side1 [mm]	side2 [mm]	side3 [mm]	side4 [mm]	diag1/√2 [mm]	diag2/√2 [mm]	Mean dist [mm]	Plan-arity [mm]	Max Dist [mm]
Img0	150.1	152.9	150.1	151.6	151.4	151.0	1.52	0.3	2.9
Img1	151.6	149.8	151.9	148.7	150.3	150.7	1.21	0.0	1.9
Img2	149.5	153.9	147.9	149.8	150.9	149.7	1.85	-3.0	3.9
Img3	150.3	151.5	148.8	148.7	149.4	150.1	0.98	-9.7	1.5
Img4	149.7	149.5	149.8	147.7	148.5	149.9	1.14	0.0	2.3
Img5	147.0	148.8	151.2	149.2	149.5	148.4	1.60	-10.4	3.0
Img6	151.2	146.9	149.7	146.2	149.3	147.7	2.29	0.0	3.8
Global							1.57	3.4	

Table 19: Calibration results on 7 shots of the reference object

	side1 [mm]	side2 [mm]	side3 [mm]	side4 [mm]	diag1/√2 [mm]	diag2/√2 [mm]	Mean dist [mm]	Plan-arity [mm]	Max Dist [mm]
Img0	149.4	152.6	149.6	151.5	150.1	151.4	1.37	0.3	2.5
Img1	151.0	149.5	151.0	148.4	149.7	150.2	0.89	-0.3	1.5
Img2	150.0	153.1	148.3	149.2	149.8	150.4	1.48	-3.2	3.0
Img4	150.0	150.0	150.2	147.9	148.8	150.2	0.99	-1.4	2.1
Img6	151.6	147.7	150.4	147.2	148.6	149.8	1.74	0.0	2.8
Global							1.33	1.0	

Table 20: Calibration results on the 5 best shots of the reference object

As depicted in Table 19, optimisation is sensitive to particular images (for instance, see the planarity of shot 3 and 5). This is due to the large influence of the horizontal localisation on depth estimation (for the current setting, 1 pixel of horizontal displacement means 5 mm displacement in depth). The horizontal localisation of stripes depends on corner and stripe edge localisation, both accounting for position errors. We found experimentally that optimising globally on 5 or more shots provides a good solution.

Table 20, when compared to Table 19, shows the advantage of dropping worst shots (here 3 and 5). We see that the distance and planarity errors can be reduced. The gain obtained for each of these errors cannot be compared because the residual errors depend on the compromise between low planarity and low distance error. But we see that both errors have been reduced.

Considering Table 20, we see that the mean error of distance estimation after calibration with the reference object is on the order of 1% (1.33 mm compared to 150 mm). The maximal error on the estimated distances is about 2% (3.0). The planarity error has a maximum of 2% (3.2) and a mean value of less than 1% (1.0). These error levels are low when we consider the different sources of errors (camera and projector optical distortion, camera and projector alignment, digitisation board distortion, corner localisation, decimal stripe labelling and stripe localisation).

To give a better idea of reconstruction errors than the ones affecting the corners of the calibration object, we considered the reconstruction of a plane spreading over the whole image (section C.5.8).

C.5 *3D extraction from striped image*

C.5.1 Overview

The stripes in the images are first localised thanks to high-pass filtering and contour following. Along each stripe, thickness (thick or thin) is estimated from grey-level and geometry (dark/bright relative size). Labelling along the stripes is obtained from horizontal distribution of thickness of neighbouring stripes. Incorrectly labelled stripes are filtered out thanks to the quasi linear dependence of the horizontal stripe position with stripe index. The same linear dependence is used to smooth the 3D surface by median filtering the horizontal positions of the stripes.

C.5.2 Stripe localisation

Thanks to the vertical stripe orientation, seeds for stripes are detected by horizontal gradient. If the seed belongs to an already localised stripe, the next seed is looked for, scanning the image from left to right, from top to bottom. Else a vertical edge following towards the bottom is initiated, stopping when the horizontal gradient is too weak. If the vertical segment is too short, the stripe hypothesis is abandoned. Else vertical edge following is started from the seed towards the top until a low gradient is encountered, and the new stripe is added to the list.

After all possible seeds have been considered, a list of stripes is available, each one consisting of a list of (x,y) coordinates of points along the stripe. Typical problems in stripe detection are:

- discontinuity of gradient, leading to partial segments for a same projected stripe;
- wrong following due to texture or surface disturbances, possibly connecting partial segments of different stripes;
- undetected stripes by lack of contrast or visibility;
- false stripes corresponding to some vertical features.

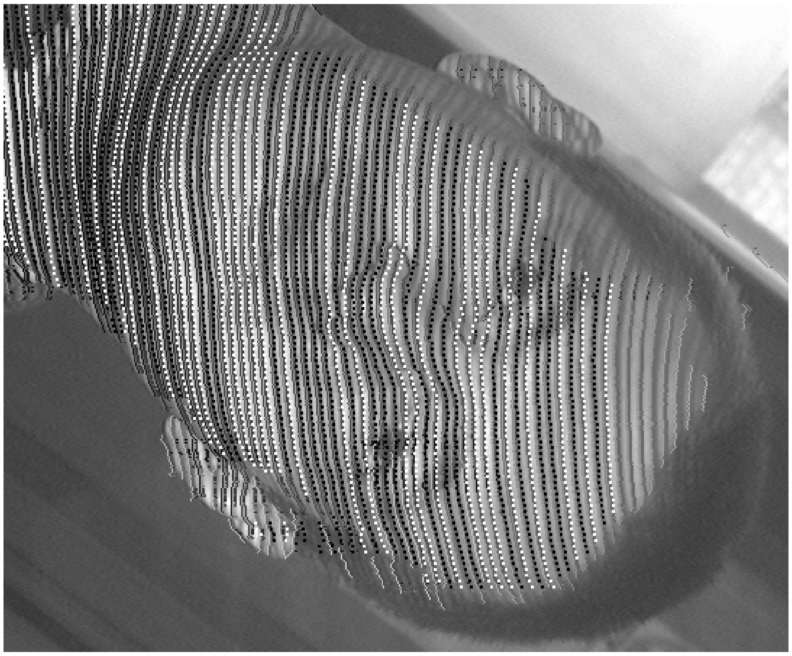


Figure 66: stripe detection

The problem of partial segments is solved by considering vertical local information to achieve stripe labelling and thus segment merging. Incorrect following has been largely reduced by imposing a minimum gradient amplitude during following. Undetected stripes are not addressed, as they can be looked for after a first detection and labelling of detected stripes. Most of the false stripes will be cancelled out during the labelling phase, as no valid thickness pattern will be found for them.

C.5.3 Local stripe thickness

Stripe thickness is evaluated by considering the pixels from the left edge of one stripe till the left edge of its right neighbour. It consists of a cycle of dark and bright grey-levels as depicted in Figure 67. The stripe is considered thick if the dark part is more important than the bright one.

Because the underlying surface has a large influence on the sensed grey level due to texture and surface orientation, the classification into dark and bright parts was not made with a simple threshold. We preferred to use a horizontal lowpass version of the image (with kernel size equivalent to 5 stripes) as local threshold.

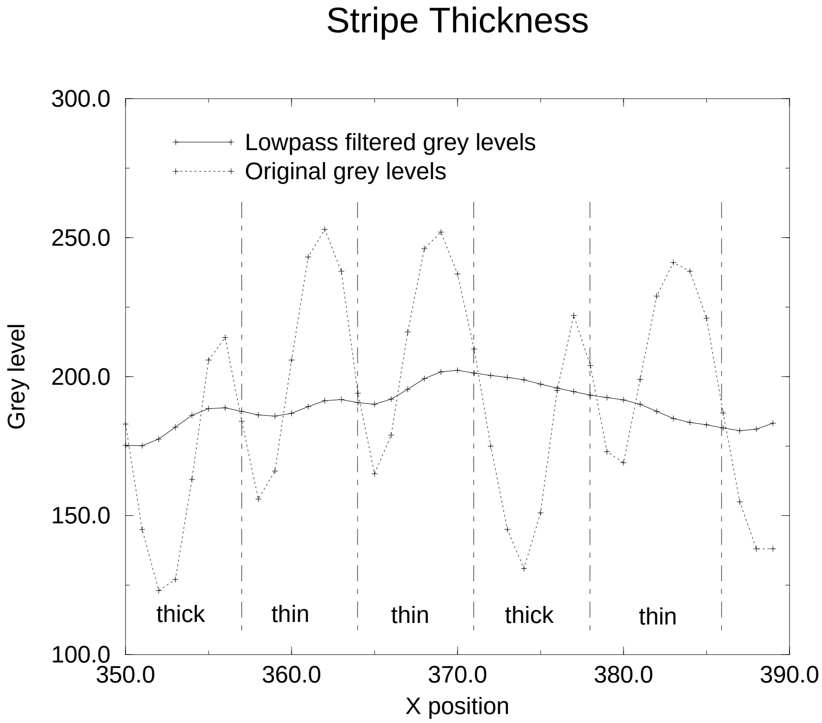


Figure 67: Stripe thickness estimation

The frontier between the dark and bright regions was localised as the highest dark to bright gradient. For both regions, the sum of the grey-level differences between the original and the lowpass image is computed. For both regions, a constant value is added to the sum for each pixel, accounting for the difference in the region size according to the stripe thickness. The largest sum indicates the stripe type. For instance, a thick stripe consists of a wide dark area followed by a narrow bright area. This implies a larger sum for the dark area due to the number of pixels and the lower grey values.

At the end of the process, each point of the stripes is labelled either as thick or thin.

C.5.4 Local stripe identification

Considering horizontal lines along the image, horizontal sequences of thick or thin votes are compared to the reference stripe pattern. For the selected striped pattern, a minimum of 7 thickness values are needed to ensure uniqueness of the corresponding stripe index. Using more thickness values increase robustness relatively to bad thickness estimation or missing stripes. Requiring too many thickness values to match reduces the number of index votes and prevents the identification of small isolated surfaces. We used a matching length of 10 stripes.

At the end of this process, each point of the stripes is either identified by the index obtained from matching with the reference or is left unidentified.

C.5.5 Homogenisation and filtering

To account for noise and problems in the identification process, points of the stripes were identified individually. This step homogenises the votes along stripe segments by picking the most represented index.

Errors in stripe labelling are usually found in regions with high grey-level disturbances (eyes, beard, hair) or little support (ears). To solve this, we exploited the linear dependence of the horizontal position of a stripe with its index (see Figure 68). This is valid for planar surfaces, but the face globally presents a nearly linear relation. Much more, the mean stripe inter-distance is nearly constant as the distance from the object to the projector is nearly proportional to the distance from the object to the camera.

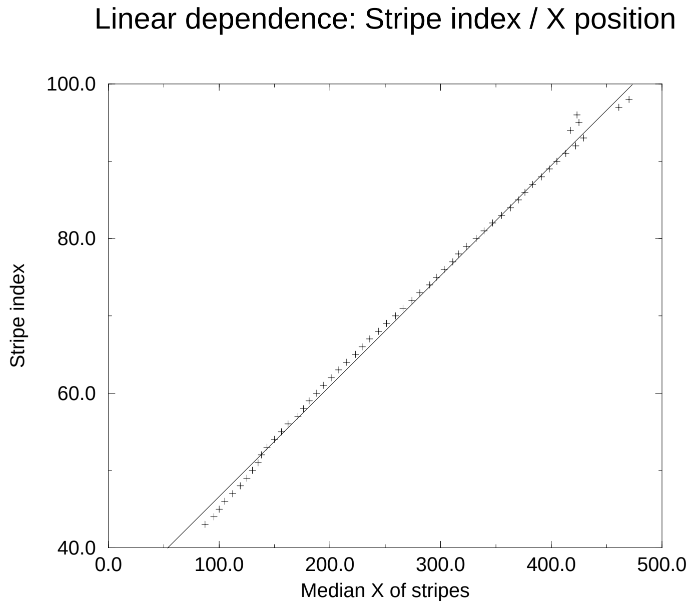


Figure 68: Stripe index versus X position

The linear dependence between the stripe position and its label is estimated in two steps. First, the slope is derived from the maximal occurrence of slope values along the index/position curve (Figure 68). As mentioned in the previous paragraph, this slope is nearly constant (but varies with surface orientation) and can also be estimated from the camera and projector settings. At least, a rough value can be obtained from the striped image, looking at the horizontal distances in pixel separating stripes. Secondly, the offset term of the linear fit is obtained from the largest occurrence of values derived from the index/position pairs, using the slope estimated in the first step.

As a first filtering step, we filter out the stripes that stand too far from the linear approximation. This has the desired advantage to cancel out stripes that induce large range discontinuities for a surface that is normally rather smooth.

As a second filtering step, median filtering is applied locally to the position/index curve, updating the position value, according to the 6 closest neighbours. This smoothes the 3D surface at a finer grain.

C.5.6 Image to 3D conversion

For each stripe having an accepted label (after filtering), the (*x,y*) coordinates of the contained points are converted into (X,Y,Z) coordinates thanks to equations of section C.3.

C.5.7 Output format

The problem of data representation is crucial for the success of an application. On the one hand, the representation should be obtained without too much effort from the acquisition system. On the other hand, other representations, better suited to feature extraction or display, should be easily derived. Considerations such as computation time, memory needs and conversion to standard format are important.

Because the facial geometric information we are looking for lies in the external surface and because this is exactly what the 3D acquisition system delivers, we opted for a 2D arrangement of 3D coordinates. Stripes are considered one by one, by increasing order of index. For each stripe, the (X,Y,Z) coordinates of points along the stripe with increasing Y are stored. Not all points are stored, only one out of four pixels, as x positions along the stripes are highly correlated. A face is typically represented by about 50 stripes of around 60 points. Storing each coordinate on 2-byte integer, a file is typically 20 Kbyte long.

C.5.8 Reconstruction error

To assess the fidelity of the 3D acquisition system, we submitted a planar object spreading over the whole field of view (Figure 69). The surface of the object was captured by the algorithm presented in this section. A plane was then fitted to the obtained surface by minimising the variance of the distances separating the object points to the plane. Only two rotation parameters intervene in this minimisation.

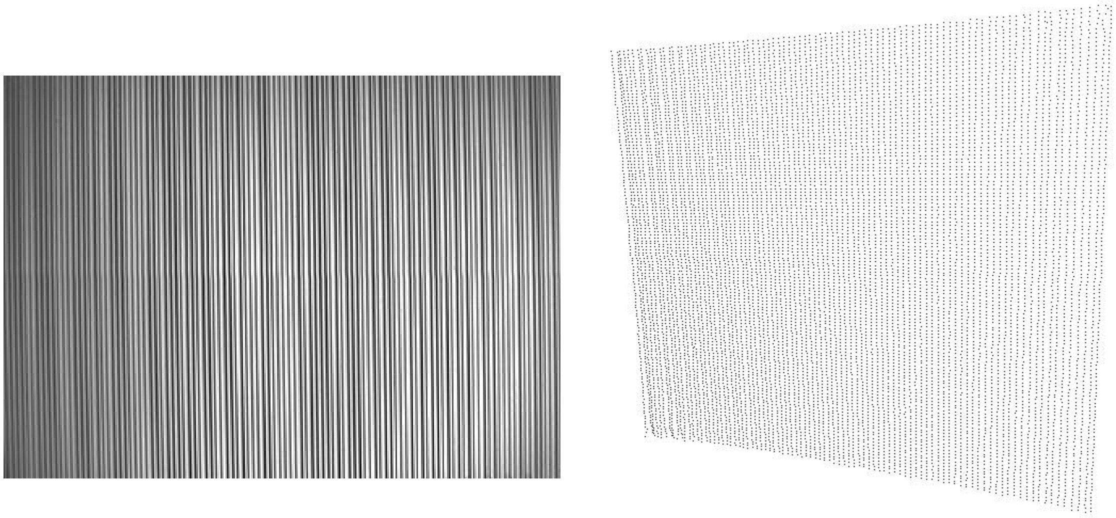


Figure 69: Capture of a planar object and perspective representation

For the best plane orientation, the histogram of the distances of object points to this plane reveals how good the acquisition is. In its cumulated form (Figure 70), this histogram gives the percentage (ordinate) of points at a maximal distance (abscissa) to the fitting plane.

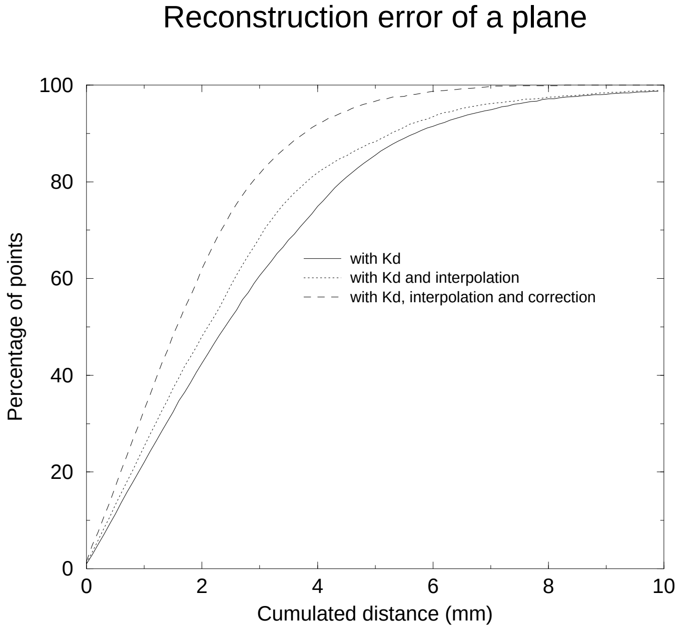


Figure 70: Cumulated histogram of error distances of the planar fit to a planar object

From the histogram results and thanks to the 3D representation of the plane (see Figure 69) several improvements were highlighted. First, as outlined during calibration, the use of the parameter *Kd* taking optical distortion into account reduces deformation at the image borders. Secondly, the sub-pixel localisation of the stripe edges gives smoother reconstructed images, fitting the reconstructed data closer to a plane ('Kd and interpolation' in Figure 70). Thirdly, it appeared that the stripe code in thickness was visible from the reconstructed plane. Looking at a profile through the plane in Figure 71, the ordinate being along the depth (Z axis) and the abscissa across the stripes (X axis), one can notice the high correlation of the Z coordinate with the stripe thickness. We attribute this dependence to the influence of the lowpass filter of the digitisation board that shifts edges in images, according to their amplitude [Dahler87]. Thin and thick stripe edges are shifted with different offsets. An average compensation was brought to the stripe position depending on their thickness type, which largely enhanced the quality of the reconstruction (Figure 70, Kd, interpolation and correction).

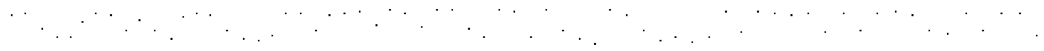


Figure 71: Cut in the planar object across the stripes

Appendix D Equations of colinearity

Lavest, in his work about geometric camera calibration [Lavest98], considers the simultaneous estimation of intrinsic and extrinsic parameters as well as reference point coordinates.

Adopting the pinhole model for the perspective projection of the 3D world into the 2D image, we have:

$$\begin{cases} x = \lambda * [r_{11}\, X + r_{12}\, Y + r_{13}\, Z + T_x] \\ y = \lambda * [r_{21}\, X + r_{22}\, Y + r_{23}\, Z + T_y] \\ z = \lambda * [r_{31}\, X + r_{32}\, Y + r_{33}\, Z + T_z] \end{cases}$$

with

(x,y,z), an image point in the camera coordinate system (z = f, focal length)

λ, a scale factor

(X,Y,Z), the reference point coordinates in the world axis system

(T_x,T_y,T_z), the translation vector

r_{ij}, the elements of the rotation matrix R, expressing 3 Euler angles (α around X axis, β around Y axis and γ around Z axis)

$$R = \begin{pmatrix} \cos\alpha \cos\beta & \cos\gamma \sin\beta \sin\alpha - \sin\gamma \cos\alpha & \cos\gamma \sin\beta \cos\alpha + \sin\gamma \sin\alpha \\ \sin\gamma \cos\beta & \cos\gamma \cos\alpha + \sin\gamma \sin\beta \sin\alpha & \sin\gamma \sin\beta \cos\alpha - \cos\gamma \sin\alpha \\ -\sin\beta & \cos\beta \sin\alpha & \cos\beta \cos\alpha \end{pmatrix}$$

Eliminating λ, we obtain the so called collinear equations

$$\begin{cases} x = f * (r_{11}\, x + r_{12}\, y + r_{13}\, z + T_x) / (r_{31}\, x + r_{32}\, y + r_{33}\, z + T_z) \\ y = f * (r_{21}\, x + r_{22}\, y + r_{23}\, z + T_y) / (r_{31}\, x + r_{32}\, y + r_{33}\, z + T_z) \end{cases}$$

Introducing the pixel coordinates (*c, r*) :

$$\begin{cases} x = (c + \varepsilon_x - c_0) * dx - distor_x \\ y = (r + \varepsilon_y - r_0) * dy - distor_y \end{cases}$$

with

dx, dy, the pixel sizes in both directions;

ε_x, ε_y, the measurement errors in pixel, to be minimised;

$distor_x$, $distor_y$, the distortion components.

Adopting the common forms for radial (dr) and tangential (dt) distortion [Lavest98] of distorx and distory:

$$\begin{cases} dr_x = (c - c_0) * dx * (a1*r^2 + a2*r^4 + a3*r^6) \\ dr_y = (r - r_0) * dy * (a1*r^2 + a2*r^4 + a3*r^6) \end{cases}$$

$$\begin{cases} dt_x = p1 * (r^2 + 2*(c - c_0)^2*dx^2) + 2 * p2 * (c - c_0)*dx*(r - r_0)*dy \\ dt_y = p2 * (r^2 + 2*(r - r_0)^2*dy^2) + 2 * p1 * (c - c_0)*dx*(r - r_0)*dy \end{cases}$$

with

c_0, r_0, the pixel coordinates of the principal point

a1, a2, a3, the polynomial coefficients of radial distortion

p1, p2, the polynomial coefficients of tangential distortion

r = sqrt($(c-c_0)^2 * dx^2 + (r-r_0)^2 * dy^2$), the radial distance from the principal point.

Introducing $f_x = f / dx$ and $f_y = f / dy$, we obtain:

$$\begin{cases} c + \varepsilon_x = c_0 + dr_x + dt_x + f_x*(r_{11}*x+r_{12}*y+r_{13}*z+T_x)/(r_{31}*x+r_{32}*y+r_{33}*z+T_z) = P(\mathbf{\Phi}) \\ r + \varepsilon_y = r_0 + (dr_y+dt_y)*f_x/f_y + f_y*(r_{21}*x+r_{22}*y+r_{23}*z+T_y)/(r_{31}*x+r_{32}*y+r_{33}*z+T_z) = Q(\mathbf{\Phi}) \end{cases}$$

with $\mathbf{\Phi} = [f_x, f_y, c_0, r_0, a1, a2, a3, p1, p2, T_x, T_y, T_z, \alpha, \beta, \gamma]^T$, the vector parameter, so that:

$$\begin{cases} \varepsilon_x = P(\mathbf{\Phi}) - c \\ \varepsilon_y = Q(\mathbf{\Phi}) - r \end{cases}$$

Appendix E Calibration: graphical interface

E.1 Objectives

The main objective of the calibration procedure is to estimate the values of the parameters so that 3D localisation is as accurate as possible.

We developed a graphical interface

- to allow for supervision of corner detection (bad images, wrong corners)
- to initialise parameter values from measures
- to control the optimisation process (parameter values, iterations, tolerance)
- to enable later examination, possibly adding new images

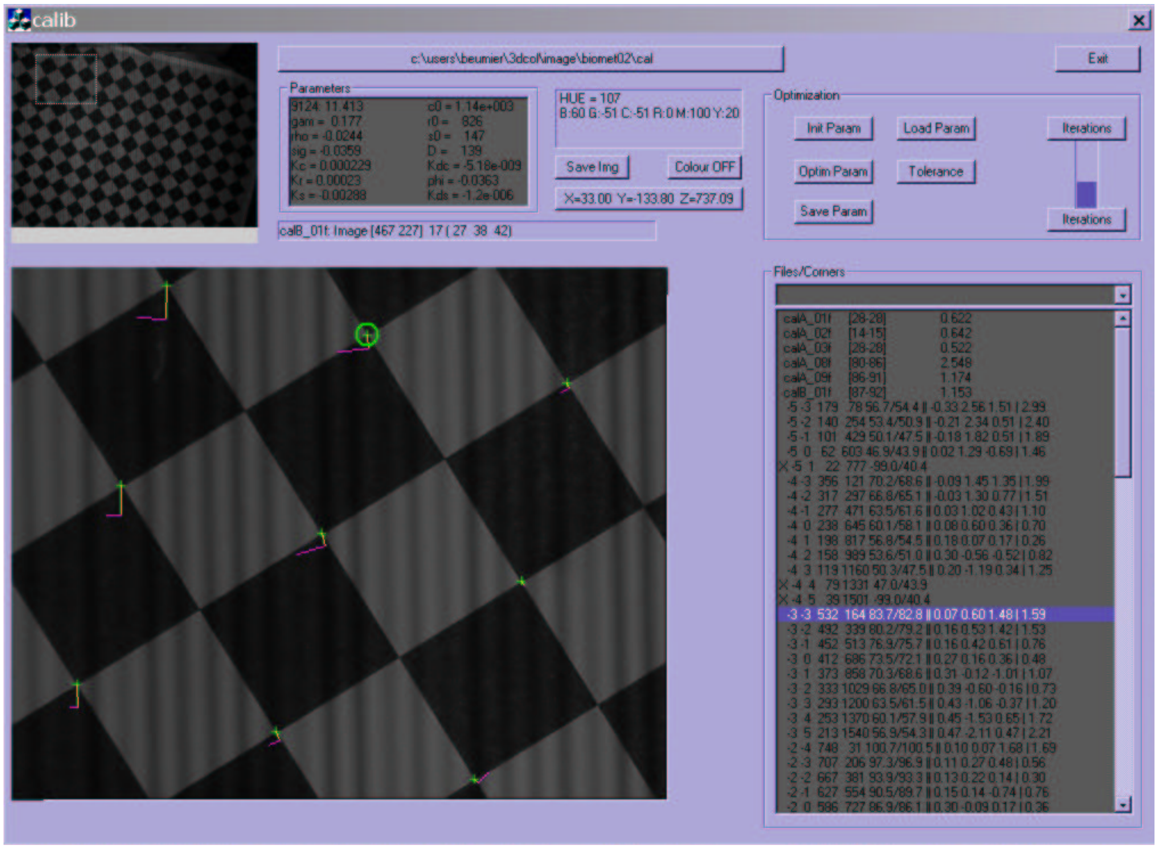
The interaction is visual, intuitive and fast.

E.2 Implementation

The calibration procedure consists in:

- capturing images of the reference object at different positions and orientations;
- localising the reference points (corners) in the image;
- initialising and optimising the parameter values to reduce errors of 3D localisation of the corners.

A graphical interface has been developed to simplify those tasks. Although primarily intended for automatic processing, the final version allows semi-automatic procedures to keep the important supervision of the user. The implied overhead is compensated by the increased reliability of a supervised approach. The whole procedure takes about 15 minutes and must be carried out only when the system setup has been modified.



```
c:\users\beumier\3dcol\image\biomet02\cal

Parameters
9124: 11.413          c0 = 1.14e+003
gam =  0.177          r0 =   826
rho = -0.0244         s0 =   147
sig = -0.0359         D =   139
Kc = 0.000229         Kdc = -5.18e-009
Kr = 0.00023          phi = -0.0363
Ks = -0.00288         Kds = -1.2e-006

HUE = 107
B:60 G:-51 C:-51 R:0 M:100 Y:20

X=33.00  Y=-133.80  Z=737.09

calB_01f: Image [467 227]  17 ( 27  38  42)
```

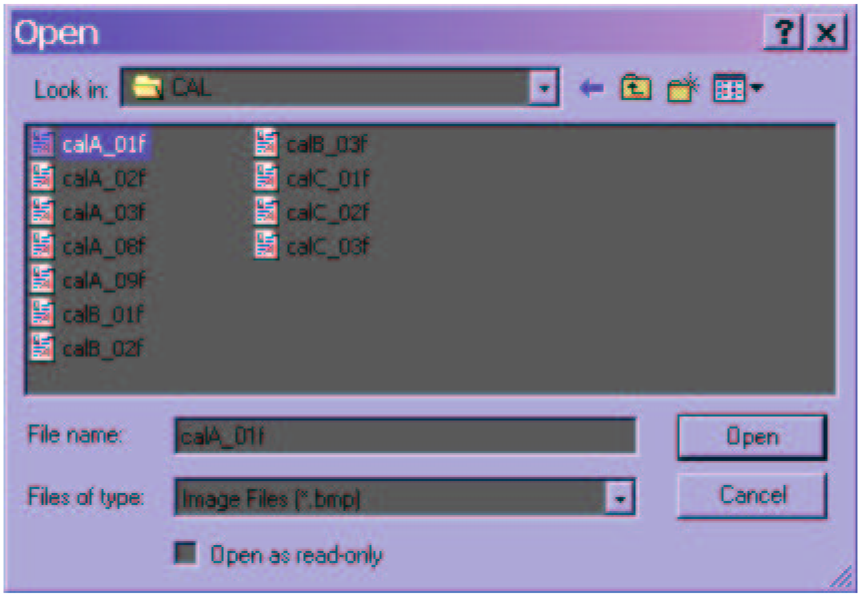
```
calA_01f    [28-28]        0.622
calA_02f    [14-15]        0.642
calA_03f    [28-28]        0.522
calA_08f    [80-86]        2.548
calA_09f    [86-91]        1.174
calB_01f    [87-92]        1.153
 -5 -3  179   78 56.7/54.4 || -0.33 2.56 1.51 | 2.99
 -5 -2  140  254 53.4/50.9 || -0.21 2.34 0.51 | 2.40
 -5 -1  101  429 50.1/47.5 || -0.18 1.82 0.51 | 1.89
 -5  0   62  603 46.9/43.9 || 0.02 1.29 -0.69 | 1.46
X -5  1   22  777 -99.0/40.4
 -4 -3  356  121 70.2/68.6 || -0.09 1.45 1.35 | 1.99
 -4 -2  317  297 66.8/65.1 || -0.03 1.30 0.77 | 1.51
 -4 -1  277  471 63.5/61.6 || 0.03 1.02 0.43 | 1.10
 -4  0  238  645 60.1/58.1 || 0.08 0.60 0.36 | 0.70
 -4  1  198  817 56.8/54.5 || 0.18 0.07 0.17 | 0.26
 -4  2  158  989 53.6/51.0 || 0.30 -0.56 -0.52 | 0.82
 -4  3  119 1160 50.3/47.5 || 0.20 -1.19 0.34 | 1.25
X -4  4   79 1331 47.0/43.9
X -4  5   39 1501 -99.0/40.4
 -3 -3  532  164 83.7/82.8 || 0.07 0.60 1.48 | 1.59
 -3 -2  492  339 80.2/79.2 || 0.16 0.53 1.42 | 1.53
 -3 -1  452  513 76.9/75.7 || 0.16 0.42 0.61 | 0.76
 -3  0  412  686 73.5/72.1 || 0.27 0.16 0.36 | 0.48
 -3  1  373  858 70.3/68.6 || 0.31 -0.12 -1.01 | 1.07
 -3  2  333 1029 66.8/65.0 || 0.39 -0.60 -0.16 | 0.73
 -3  3  293 1200 63.5/61.5 || 0.43 -1.06 -0.37 | 1.20
 -3  4  253 1370 60.1/57.9 || 0.45 -1.53 0.65 | 1.72
 -3  5  213 1540 56.9/54.3 || 0.47 -2.11 0.47 | 2.21
 -2 -4  748   31 100.7/100.5 || 0.10 0.07 1.68 | 1.69
 -2 -3  707  206 97.3/96.9 || 0.11 0.27 0.48 | 0.56
 -2 -2  667  381 93.9/93.3 || 0.13 0.22 0.14 | 0.30
 -2 -1  627  554 90.5/89.7 || 0.15 0.14 -0.74 | 0.76
 -2  0  586  727 86.9/86.1 || 0.30 -0.09 0.17 | 0.36
```

Figure 72: Graphical interface for calibration

The graphical interface has been developed for prototype B. The parts related to image and corner manipulation and detection have been used for prototype C as well. The interface contains four areas:

E.2.1 Image selection

Clicking the button displaying the current directory calls the ***File Dialog Box*** that allows picking one of the BMP files of the current directory. The chosen item replaces the current image that is displayed in the image area, and which is used for interactive corner localisation and display.



Look in: CAL

calA_01f
calA_02f
calA_03f
calA_08f
calA_09f
calB_01f
calB_02f
calB_03f
calC_01f
calC_02f
calC_03f

File name: calA_01f

Files of type: Image Files (*.bmp)

Figure 73: File Selection Box

The File Dialog Box is mainly used to change the current directory. When a change is made, all BMP files are listed in the corner and image list (see E.2.4 below). Clicking an image in this list changes the current image that is displayed in the image area.

At startup, the last current directory and current image as stored in the saved context are loaded. To change these specifications, choose ‘Save’ after pressing ‘Exit’ when exiting.

E.2.2 Corner localisation

Corner localisation can be achieved automatically or manually.

When an image is loaded (selected in the File Dialog Box), the constants of the linear relation, topic of section 2.4.4c), are automatically estimated (Δs_x, s_1 and possibly Δs_y). Should this evaluation fail for one of these parameters, the software looses its ability to check corner position and stripe index for global consistency. Certainly in this case, the image should be investigated to verify its adequacy for corner or stripe detection.

Manual corner localisation is achieved by calling the contextual menu at an expected position on the image and selecting "Add Corner". If a corner is clearly identified from the vertical edges, it is displayed by a 'plus' sign (normally green) on the image and added to the corner list with coordinates and stripe index. Should this index appear too different from the linear guess (estimated at image load), the corner is marked inactive '#' and the plus sign is drawn in red. If no corner is found, the message 'Bad corner detection' is displayed.

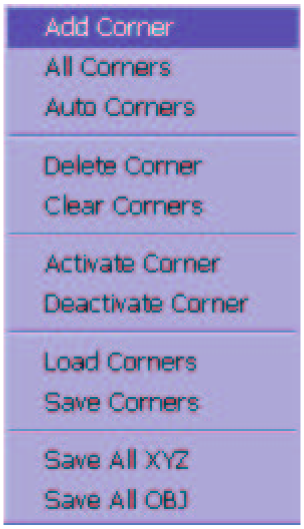


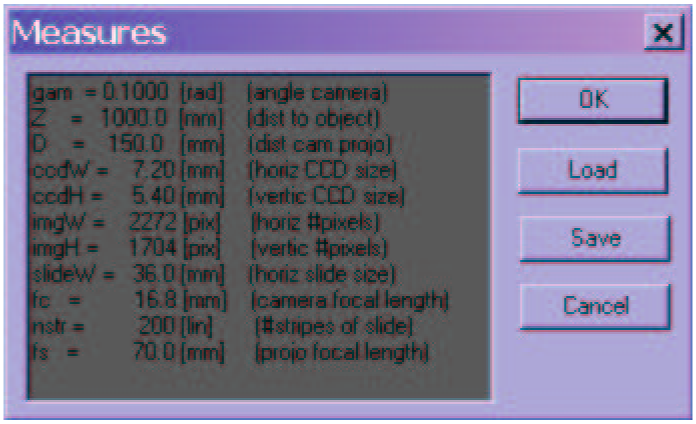
Figure 74: Contextual menu

Up to 3 corners can be manually localised. This forms the basis (symbolic coordinates ‘0 0’ ‘1 0’ ‘0 1’) for the search for the array of corners based on the two main directions. The entry 'All Corners' of the contextual menu implements this function. In case a new basis must be defined, first select 'Clear Corners' from the contextual menu to erase all the corners or select 'Delete Corner' to erase the current corner (selected in the corner list box and circled on the image).

The manual procedure can be ***automated*** (entry 'Auto corners') to look for all the corners from the first three ones obtained automatically near the centre of the image. In this case, a special marker is looked for in the calibration object to find the origin of the symbolic coordinates necessary for coherent indexing between various object presentations.

E.2.3 Parameter optimisation

Parameter optimisation typically starts with the ***initialisation*** of the parameter values. This is performed either by ‘Load Param’ or by ‘Init Param’ buttons in the 'Optimisation' area. ‘Load Param’ loads the last saved parameter set. ‘Init Param’ calls the ‘Measures’ dialog to enter the measures of the system like distances, focal lengths, sensor and image sizes. When ‘OK’ is pressed, those measures are converted into initial values for the parameters.



```
gam  = 0.1000  [rad]   (angle camera)
Z    =  1000.0 [mm]    (dist to object)
D    =   150.0 [mm]    (dist cam projo)
ccdW =    7.20 [mm]    (horiz CCD size)
ccdH =    5.40 [mm]    (vertic CCD size)
imgW =    2272 [pix]   (horiz #pixels)
imgH =    1704 [pix]   (vertic #pixels)
slideW =  36.0 [mm]    (horiz slide size)
fc   =    16.8 [mm]    (camera focal length)
nstr =     200 [lin]   (#stripes of slide)
fs   =    70.0 [mm]    (projo focal length)
```

Figure 75: Measure interface

Parameter optimisation proceeds when the 'Optim Param' button is pressed. The Simplex algorithm [NumRecipes] is launched with current parameter values and a maximal number of iterations. Once launched, the button reverts to 'Stop Optim' to give the possibility to stop optimisation. Otherwise, the optimisation runs until either the tolerance has been reached or the number of iterations has been exhausted. Optimisation is implemented as a separate thread to maintain interaction with the application like stopping optimisation or continuously refreshing the parameter values in the parameters list box.

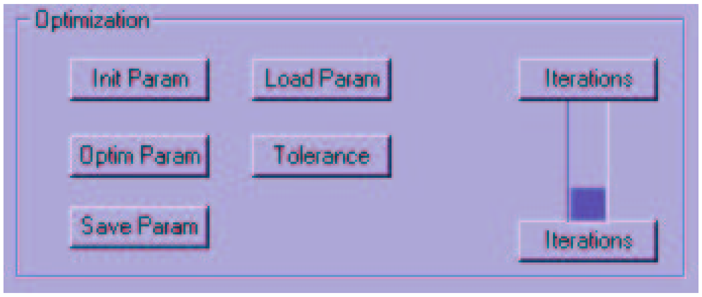


Figure 76: Parameter interface

The chance of success increases with correct initial values and representative data (valid corners from correct images). Another way to investigate convergence when optimising with a set of data is the possibility to activate a subset of parameters while freezing the others. A frozen parameter is not tuned during optimisation.

At any time, the set of parameters can be saved ('Save Param') to file 'param.txt' in the current directory. This is an ASCII file listing the parameter values in the order specified in Table 5.

E.2.4 Corner interaction

For the sake of coherence, the ***corner and the file lists*** are managed through the same list box.

Initially, image files found in the specified directory are listed. The files with corresponding '.cal' files (corners previously detected and saved) are active and other files are inactive ('#' prepended).
Only the current file (the one selected in the list) lists its corners, if any, and displays them (by plus signs) on the image.

List navigation occurs with the mouse or with UP and DOWN arrows. When a file is encountered, it is loaded and displayed. If a corresponding '.cal' file exists, corners are listed below the file in the list and displayed on the image by plus signs. When a corner is encountered, the plus sign in the image is circled, with the same colour as the plus sign (red for bad corners, green for good and active corners and blue for corners unselected by the user).

An inactive corner is not used during optimisation. It is marked with 'X' (red in image). When a corner is detected, its stripe index is estimated and compared to the global linear guess. In case of match, the corner is active (green), otherwise it is inactive ('X', red). The active attribute can be changed (toggled) by clicking the corner in the list. When set inactive by the user, the corner is blue in image and marked '#' in the list. When a file is toggled inactive, none of its corners is used for optimisation.

During optimisation, the error of each active corner of the selected file is displayed in the list (error in x, y, z and 3D). The mean error of the active corners of an active file is displayed respectively to file entries of the list, beside the number of detected and active corners for those files.

After optimisation, larger errors can be tracked, either considering the file errors or the individual corner errors. Corners with significant errors should be examined, possibly set inactive. Most of the time, these correspond to incorrectly localised corners, or with erroneous stripe index. Of course the number of discarded corners should remain limited, otherwise the whole image discard should be considered.

E.3 Output

Three types of file output are possible:

Save to 'param.txt'
The set of parameters is stored in file 'param.txt' in the order specified in Table 5. Other programs use this file to extract 3D data from image measures.

File '.cal'

The list of corners of a file can be saved (image contextual menu ‘Save Corners’) in order to be later retrieved in the graphical interface or processed by another program. The file has the name of the image with extension ‘.cal’. If it exists, the list of corners is automatically loaded when the corresponding image file is loaded.

File ‘.xyz’

The 3D coordinates of the active corners in the current image for the current parameter values can be saved (image contextual menu ‘Save All XYZ’ or ‘Save All OBJ in ‘.obj’ format) for 3D inspection by some other tool.

Appendix F Historic

The successive developments have led to three major prototypes named “B&W thickness” (prototype A), “Colour striping” (prototype B) and “Stripe with dots” (prototype C), in chronological order.

Their main features are summarised in the next table. Refer to Chapter 2 for deeper technical explanations.

Prototype:	**Prototype A**	**Prototype B**	**Prototype C**
Nickname	“B&W thickness”	“Colour striping”	“Stripe with dots”
Date	1999	End 2002	End 2003
Full description	Appendix C	Chapter 2	Chapter 2
Camera			
Type	Panasonic WV-BL600	Canon G2 / colour	Canon G2 / colour
Image size	768x576 (digitiser)	2272x1704	2272x1704
Texture	No	Yes	Yes
Projection			
Type	Projector 150 W	Flash lamp	Video projector
Slide	B&W print on glass	Colour film slide	BMP file
Pattern	Vertical lines	Vertical lines	Vertical lines
Code	Line thickness (2)	Line colour (6)	Dots on line
Calibration			
Object	One large square	Chessboard	Grid of black lines
Optimisation	Semi auto	Simplex	Gradient descent
3D			
Point density on face	50x70	60x80	70x100
Database	2 x 120pers / 3shots	1 x 81pers / 6 shots	None
Output format	Proprietary	‘.obj’	.wrl (VRML), .obj

Table 21: Main characteristics of the three acquisition prototypes

As already explained in Chapter 1 on face recognition, the successful face ***profile recognition*** prototype was naturally followed by a feasibility study about the possibility to build a low cost and rapid 3D acquisition system for face recognition. In late 1994, the first experiments were conducted which gave us the confidence to propose 3D Face Recognition activities in the M2VTS consortium (“Multi-Modal Verification for Telesurveillance and Security Application”, [M2VTS]) that was built up beginning 1995.

F.1 Prototype A

M2VTS actually started at the end of 1995. This meant for us the development of a robust profile identification prototype. 1996 was partially spent to complete the feasibility study about 3D capture and 3D face recognition. The development of Prototype A really started in early 1997, with the mechanical construction, the slide realisation and stripe detection and labelling.

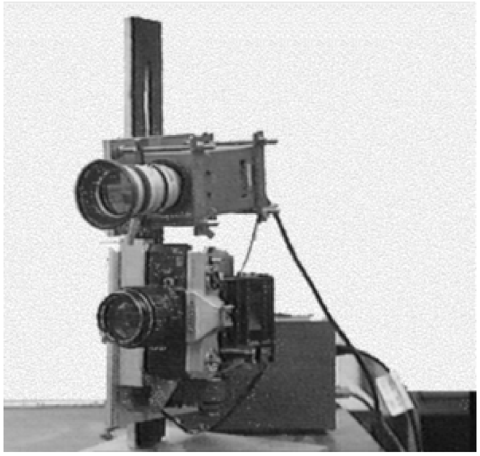


Figure 77: Prototype 'B & W thickness'

By mid 1998, we had realised the complete chain for 3D capture and basic recognition. 3D acquisition needed only half a second to deliver 50x50 3D coordinates. The camera and PCI acquisition board captured the image in 0.1 sec and 0.4 sec were necessary to extract the stripes, label them and filter some rough errors. Calibration was manual, time-consuming and poor. Face recognition was based on the comparison of the central vertical profile.

By mid 1999, a database was acquired and the face recognition approach was refined to include a complete facial surface comparison. A series of recognition experiments were conducted and many publications followed [Beumier99b, Beumier00a, Beumier00c, Beumier00f, Beumier01a, Beumier01b]. At that time, no continuation was foreseen and the 3D face recognition activities stopped.

The 3D Face project was revived in mid 2001, thanks to collaboration within the BIOMET project. In the first place, we wanted to take the opportunity to improve the 3D acquisition system, taking advantage of improved quality digital cameras. At the same time, we refined the mechanical design to build up a lighter and more compact version. Finally, a decent calibration procedure was designed to support the possible quality improvement. This involved the design of a calibration object and the development of an interactive program for reference point detection and selection, and parameter initialisation and optimisation. The final touch was the replacement of the classical incandescent lamp of the projector by a flash to build up a compact system with large depth of focus. The fan cooler required by incandescent light became obsolete.

Ready in Mid 2002, the enhanced prototype A performed poorly during the second BIOMET database acquisition, due to the impossibility to control the flash power from the camera. More than half of the images had blurred stripes.

F.2 Prototype B

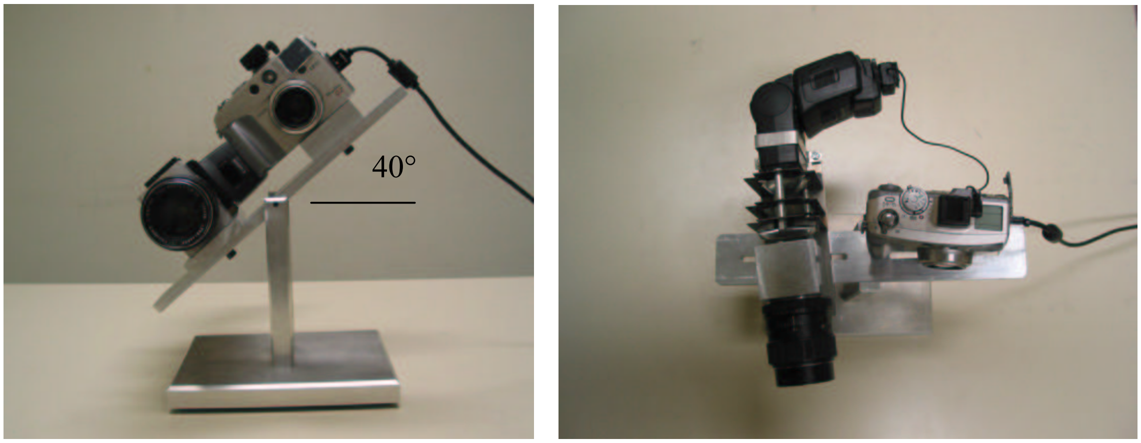


Figure 78: Prototype B 'Colour'

Learning from the poor results caused by the flash, we designed prototype B, based on the same hardware but with a different slide. The main idea was to downscale the stripe resolution so it would be less sensitive to focus.

The colour slide appeared as the perfect solution. Stripes can have constant width, with the possibility to use the edges to have twice as many 3D points. Colour striping allows a more compact labelling. Fewer and larger stripes can then be projected.

Finally, colour appeared as a good opportunity to apprehend texture acquisition from the same image. In spite of the difficulty we experienced in realising a colour slide of good quality, the prototype B was ready by the end of 2002 for the third BIOMET acquisition campaign.

Focusing problems which impaired the 3D acquisition during the second campaign disappeared in the last campaign. But the quality of the colour slide (film) was not as good as the slide realised for prototype A (made of glass). The illumination was rather dark and the colours were weaker. We had to adapt the stripe localisation algorithms, because stripe edges (between colour and white stripes) were not always visible enough. The centres of colour stripes were more easily detected but were only half as numerous as the edges, resulting in a density reduction of the 3D points.

The same campaign highlighted a calibration problem related to the necessity to limit the number of parameters optimised at a time, due to dependencies. Since this led to huge errors in the estimation of the camera focal length, and incidentally to wrong qualitative 3D appearances, we decided to separate the problems of camera and projector calibration and to consider a well established calibration procedure.

F.3 Prototype C

The negative experience regarding colour slide realisation and the image sensitivity to colours such as cyan or yellow revealed a possible weakness of the colour approach. Although we could reduce the set of colours to the most visible ones, we preferred to consider black and white projection to take advantage of good slides, good contrast and accurate image sensing thanks to the coherent R, G, and B signals. It is clear that stripe detection and localisation are the vital steps in the processing. A good positioning is necessary for texture and labelling. Labelling can however be corrected with a priori knowledge regarding the projected pattern and texture capture is not primordial. The idea of thin stripes arose as a way to address a better recovery of texture thanks to the reduced influence of the stripes. We finally found an appropriated solution: labelling the stripes by vertical dot positions.

This solution was further motivated by the possibility to address 3D capture in general, not only for faces and, according to details given before, with a higher resolution.

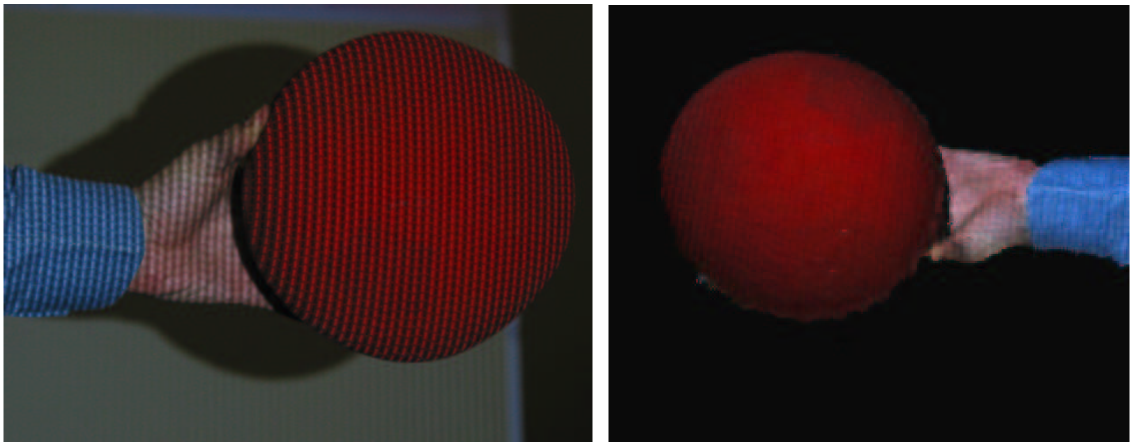


Figure 79: Capture of coloured object and 3D representation

The Prototype C is not considered as finished. A demonstrator first based on a video projector was realised. The calibration procedure, the stripe localisation and labelling, and the texture recovery were successfully developed. In order to palliate the limited depth of field of the video projector, we decided to use the flash. A slide was produced with a photo slide film. Unfortunately, the slide opacity and the limited power of the flash prevented us to capture striped images with good contrast. We plan to produce the slide on glass, as we did for prototype A.

Appendix G Database download

This document is available at
http://www.sic.rma.ac.be/~beumier/DB/3d_rma.html

G.1 Distribution restrictions

The use of the "3D_RMA" database recorded at the SIC is restricted to research purposes. The (re-)distribution of the database - or part of it - either in original form or modified is allowed only with the written permission of the concerned persons of the database. However, any know-how gained thanks to the database may be freely used and transmitted.

SIC distributes the database subject to the acceptance of the above conditions. By reception of the signed agreement, SIC will give access to the files for download (tar.gz 13MB).

G.2 3D acquisition

In the framework of our face recognition activities, we developed at the SIC, a 3D acquisition system based on structured light. With a projector and a camera, the facial surface is acquired as a set of 3D coordinates. The quality and precision seems sufficient for this first prototype. However, glasses and bushy or dark facial hair impair the acquisition.

G.3 3D Database

120 persons were asked to pose twice in front of the system: in Nov 97 (session1) and in January 98 (session2). For each session, 3 shots were recorded with different (but limited) orientations of the head:
frontal / Left or Right / Upward or downward.

Among the 120 people, two thirds consist of students from the same ethnic origins and with nearly the same age. The last third consists of people of the academy, all aged between 20 and 60.

Different problems encountered in the cooperative scenario were taken into account. People sometimes worn their spectacles, sometimes did not. Beards and moustaches were represented. Some people smiled in some shots. Small up/down and left/right rotations of the head were requested. We regret that only a few (14) women were available.

G.4 3D format

Each 3D file, with extension '.xyz', is organised as a set of 3D points along stripes. Each stripe starts with a short integer indicating the number N of points in the stripe. Then N triplets of short (signed 2-byte) integer give the point coordinates. The resolution is 0.1 mm per unit. All short values are stored binary.

Ex: 2 x y z x y z 3 x y z x y z x y z 2 x y z x y z

The head is scanned (stripe order) from chin to forehead (increasing x) and the stripes are scanned left to right (increasing y). Do mention that for technical reasons, the camera/projector head was rotated about 40°, so that the X and Y axes do not correspond to respectively the vertical and horizontal axis.

See readXYZ.c for a read function in C.

G.5 Download

Please address a fax to the attention of

Mr Charles Beumier
ELEC Dpt, SIC, Royal Military Academy Belgium
+32 2 737 64 72

G.6 M2VTS

This database has been acquired in the framework of the M2VTS project.