Titre: Title:	Évaluation des mouvements rigides et non rigides du visage pour les vidéoconférences
Auteur: Author:	Étienne Boutin
Date:	2005
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Boutin, É. (2005). Évaluation des mouvements rigides et non rigides du visage pour les vidéoconférences [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/7526/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/7526/
Directeurs de recherche: Advisors:	Paul Cohen
Programme: Program:	Non spécifié

NOTE TO USERS

UNIVERSITÉ DE MONTRÉAL

ÉVALUATION DES MOUVEMENTS RIGIDES ET NON RIGIDES DU VISAGE POUR LES VIDÉOCONFÉRENCES

ÉTIENNE BOUTIN
DÉPARTEMENT DE GÉNIE ÉLECTRIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION DU
DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES (M. Sc. A.)
(GÉNIE ÉLECTRIQUE)
Mai 2005

Your file Votre référence
ISBN: 0-494-01289-7
Our file Notre référence
ISBN: 0-494-01289-7

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

ÉVALUATION DES MOUVEMENTS RIGIDES ET NON RIGIDES DU VISAGE POUR LES VIDÉOCONFÉRENCES

présenté par : BOUTIN Étienne

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. COHEN Paul, Ph. D., membre et directeur de recherche

Mme. CHÉRIET Farida, Ph. D., membre du jury

M. GOUSSARD Yves, Ph. D., membre du jury

Remerciements

L'auteur tient à remercier M. Paul Cohen (directeur du projet) et M. Niu Kegong (étudiant sur le projet) pour les conseils et l'aide apportés à la réalisation de ce projet. L'auteur tient aussi à remercier l'École Polytechnique, l'aide financière du GRPR ainsi que tous ceux qui ont accepté de participer aux expérimentations.

Résumé

La vidéoconférence a pour but de permettre à deux ou plusieurs individus de communiquer à distance à l'aide d'ordinateurs tout en ayant accès au son et aux images. Pour un système de vidéoconférence de type « tête-épaules » par exemple, les séquences d'images obtenues couvrent principalement le visage et le buste des usagers puisqu'il s'agit de la région du corps principalement observée lors d'une conversation naturelle. En raison de la lourdeur, en terme de débit de transmission, d'une séquence d'images et de l'importante proportion de redondance présente dans le signal, de nombreuses techniques ont été étudiées afin de réduire le taux de transmission. Dans le cas où les images concernent des scènes générales, les techniques consistent souvent à compresser les données, sans utiliser un modèle de la scène réelle. Dans le cas de la vidéoconférence de type « tête-épaules », une importante quantité d'informations a priori de la scène est disponible provenant du fait qu'il s'agit d'un visage et qu'il existe plusieurs caractéristiques anthropométriques plus ou moins respectées par la plupart des individus. Le but de ce projet est de développer un système spécialisé en analyse et traitement de séquences d'images pour lire les mouvements de la tête et du visage d'un usager d'une façon entièrement automatique. Le projet ne traite pas l'élaboration d'un système complet pour la vidéoconférence, étant donné l'ampleur du travail nécessaire. Mais la simple lecture des mouvements a un grand potentiel pour la vidéoconférence étant donné la réduction du taux de transmission entre ordinateurs. Les séquences d'images concernées sont celles normalement rencontrées lors d'une vidéoconférence.

Ce travail est la poursuite de travaux effectués précédemment [84] portant sur la détection de visage, l'estimation du mouvement 3D rigide de la tête et l'adaptation d'un modèle virtuel de la tête aux paramètres de mouvements estimés. Les travaux décrits dans ce mémoire apportent des améliorations à quelques méthodes élaborées dans [84] et portent principalement sur la détection et la paramétrisation des mouvements non rigides du visage. Les mouvements non rigides pertinents pour ce projet sont ceux engendrés par les yeux, les sourcils et la bouche. Pour estimer ces mouvements, plusieurs modèles géométriques sont utilisés. Malgré le nombre important d'opérations pour estimer à la fois les mouvements rigides et non rigides, le système serait apte à un fonctionnement en temps réel sur un ordinateur assez puissant, disponible avec la technologie actuelle, et grâce à l'optimisation de certains algorithmes. Pour l'instant, sur un ordinateur de type PC 200 MHz, le taux est d'environ d'une image par seconde. Le taux de réussite du système a été jugé acceptable, malgré la difficulté du problème, comme le démontrent les résultats expérimentaux.

Abstract

The goal of videoconference is to allow two or many users to communicate far away from each other by using computer to have sounds and images. For a « head-and-shoulders » videoconference system by example, the images sequences obtained are mainly on the face and bust user since that's the region on the body which is mainly observed in a natural conversation. Since there is a lot of data in an images sequence and redundancy in the signal, many techniques have been studied to reduce the transmission rate. When the images are concerned in general scenes, the techniques often consist to compress data, without using any real scene model. In the case of a « head-and-shoulders » videoconference, a lot of information from the scene is available because that's a face user and many anthropometrics constraint are nearly constants for most of the population. So the goal of this project is to define an images sequences specialized system to read the human head and face motions in an entirely automatic way. This project doesn't define a complete videoconference system, since it requires an amount of work. But only reading the motions has a great potential for videoconference by the data rate reduction between computers. The images sequences concerned in this project are the ones generally meet in a videoconference. This work resume the work already made by [84] which were limited at the face user detection, 3D rigid head motion estimation and virtual head model adaptation using the estimated parameters. The works describe in this memory bring improvement on some methods done by [84] and is mainly on the non-rigid face motions estimation and

parameterization. The non-rigid motions relevant in this project are the ones for the eyes, the eyebrows and the mouth. To estimate these motions, many geometrics models are used. Even if there is a lot of computation for the rigid and non-rigid motions estimation, the system would be suited to work in real-time with a more powerful computer, available with the current technology, and with the optimization of some algorithms. For now, on a PC 200 MHz computer, the rate is close to one image per second. The success rate of the system has been judged acceptable, on the problem difficulties, and will be exposed further in this work with the experimental results.

Table des matières

Remerciements.....IV

Résumé.....V

Abstract.....VII

Table des matières.....IX

Liste des tableaux.....XIV

Liste des figures.....XV

1. INTRODUCTION.....1

2. STRUCTURE GÉNÉRALE DU SYSTEME.....10

2.1. Les contraintes au projet.....10

2.2. Études pour l'évaluation des mouvements rigides.....11

2.2.1. Méthodes cherchant une localisation grossière de la tête.....12

2.2.2. Méthodes comparant la projection du modèle et l'usager.....13

2.2.3. Méthodes cherchant le déplacement d'un objet par flux optique.....14

2.2.4. Méthodes invariantes à l'éclairage.....16

2.2.5. Méthodes basées sur le suivi de points 2D.....17

2.3. Études pour la localisation des éléments non rigides.....18

2.3.1. Méthodes adaptant des modèles géométriques connus.....19

2.3.2. Méthodes adaptant des trajectoires quelconques.....22

2.3.3. Études sur la détection d'expressions faciales.....24

2.4. Méthodes proposées pour ce projet.....25

2.4.1. L'estimation des mouvements rigides.....25

2.4.2. La localisation des éléments non rigides.....28

2.5. Les modules du système proposé.....30

2.5.1. Détection du visage et adaptation du modèle.....31

2.5.2. Estimation du mouvement rigide.....31

2.5.3. La localisation des éléments non rigides.....32

2.5.4. Adaptation du modèle virtuel.....32

3. LE SUIVI DES POINTS 2D.....33

3.1. Objectif du suivi.....33

3.2. Les difficultés du suivi.....34

3.3. Fonctionnement général du suivi.....36

3.3.1. Recherche par la SSD.....37

3.3.2. Recherche par la NCC.....41

3.4. Calculs pour la NCC utilisant les modèles.....46

3.4.1. La création des modèles.....47

3.4.2. Utilisation de la mémoire statique.....50

3.4.3. Utilisation de la mémoire des images.....50

3.4.4. Démonstration des modèles utilisés.....53

3.5. Analyse des résultats obtenus.....55

3.6. Améliorations suggérées.....55

4. ESTIMATION DU MOUVEMENT RIGIDE 3D.....56

4.1. Objectif de l'estimation.....56

4.2. Les difficultés de l'estimation.....57

4.3. Fonctionnement général de l'estimation.....57

4.4. La méthode implantée.....61

4.4.1. L'algorithme RANSAC général.....63

4.4.2. La méthode RANSAC pour estimer le mouvement rigide de la tête.....67

4.4.3. Sélection des points 2D dans le visage pour le suivi.....70

4.4.4. Création des ensembles des points.....72

4.4.5. Attribution d'un score pour un mouvement.....73

4.4.6. Estimation finale du mouvement.....77

4.4.7. Relocalisation des points erronés.....77

4.4.8. Les sources d'imprécisions du système.....79

4.5. Analyse des résultats obtenus.....81

4.6. Améliorations suggérées.....81

5. LOCALISATION DES ÉLÉMENTS NON RIGIDES.....83

5.1. Le suivi des yeux.....86

5.1.1. Brève revue de la littérature sur la détection des éléments non rigides.....88

5.1.2. Description générale de la méthode utilisée.....90

5.1.3. Localiser les deux iris potentiels.....91

5.1.4. Estimer le sommet et la base de l'ouverture des yeux.....98

5.1.5. Déterminer si l'oeil est ouvert ou fermé.....101

5.1.5.1. Vérifier s'il y a des vallées vis-à-vis de l'iris.....103

5.1.5.2. Vérifier s'il y a des sommets vis-à-vis du blanc de l'œil.....104

5.1.5.3. Vérifier si la région de l'iris est assez foncée.....107

5.1.6. Adapter le contour de l'oeil 109

5.1.6.1. Premier cas : les deux yeux sont ouverts.....111

5.1.6.2. Deuxième cas : les deux yeux sont fermés.....118

5.1.6.3. Troisième cas : un oeil est ouvert et l'autre est fermé.....122

5.1.7. Système de correction d'erreur.....124

5.1.8. Analyse des résultats obtenus.....127

5.2. Le suivi des sourcils.....129

5.2.1. Description générale de la méthode utilisée.....130

5.2.2. Obtention des données sur les yeux.....132

5.2.3. Positionnement d'un segment de droite sur le centre du sourcil.....133

5.2.4. Positionner deux segments de droite sur le sourcil.....134

5.2.5. Analyse des résultats obtenus.....136

5.3. Le suivi de la bouche.....139
5.3.1. Description générale de la méthode utilisée.....141
5.3.2. Initialisation d'un modèle de la bouche sur l'image initiale.....144
5.3.2.1. Fenêtres de recherche pour les coins de la bouche.....146
5.3.2.2. Localisation des coins extérieurs de la bouche.....149
5.3.2.3. Fermeture de la bouche.....154
5.3.2.4. Contours extérieurs de la bouche.....156
5.3.2.5. Informations nécessaires au suivi.....158
5.3.3. Réajustement du modèle sur l'image courante.....160
5.3.3.1. Adaptation des coins extérieurs de la bouche.....161
5.3.3.2. Adaptation du centre des courbes extérieures de la bouche.....163
5.3.3.3. Adaptation du centre des courbes intérieures de la bouche.....164
5.3.3.3.1. Premier cas : la bouche est ouverte.....165
5.3.3.3.2 Deuxième cas : la bouche est fermée.....167
5.3.4. Analyse des résultats obtenus.....167
5.4. Intégration et analyse des résultats du suivi.....170
5.5. Conclusion.....172
5.6. Améliorations suggérées.....173
6. CONCLUSION.....174
6.1. Améliorations suggérées au projet.....175
RÉFÉRENCES.....177
Annexe I : Analyse des résultats du suivi de points 2D.....190
Annexe II : Analyse des résultats de l'estimation du mouvement rigide.....197
Annexe III : Analyse des résultats de la localisation des éléments non rigides.....205

Annexe IV : Intégration numérique de courbes ou surfaces sur les images.....226

Annexe V : Calculer les images des vallées, des sommets et des arêtes.....236

Annexe VI : Étude sur l'anthropométrie du visage.....238

Liste des tableaux

Tableau 4.1 Évaluation de la profondeur selon le modèle virtuel.

Tableau 4.2. Combinaisons possibles de points pour créer les ensembles.

Tableau 4.3. Valeurs maximales permises pour augmenter le score.

Tableau 5.1. Caractéristiques propres aux coins d'un œil.

Tableau 5.2. Caractéristiques propres à l'ouverture d'un œil.

Tableau 5.3. Caractéristiques propres à l'iris.

Tableau 5.4. Caractéristiques sur l'intensité des pixels.

Tableau 5.5. Caractéristiques sur le mouvement.

Tableau 5.6. Caractéristiques sur la géométrie.

Tableau 5.7. Caractéristiques selon la géométrie.

Tableau 5.8. Caractéristiques sur l'intensité des pixels.

Tableau 5.9. Caractéristiques sur le mouvement.

Tableau A.1. Résultats de la première séquence avec la SSD simple.

Tableau A.2. Résultats de le première séquence avec la SSD et la NCC.

Tableau A.3. Résultats de la deuxième séquence avec la SSD simple.

Tableau A.4. Résultats de la deuxième séquence avec la SSD et la NCC.

Tableau A.5. Paramètres obtenus pour la première séquence.

Tableau A.6. Paramètres obtenus pour la deuxième séquence.

Tableau A.7. Paramètres obtenus pour la troisième séquence.

Tableau A.8. Paramètres obtenus pour la quatrième séquence.

Liste des figures

Figure 1.1. Exemple de vidéoconférence sans traitement d'images.

Figure 1.2. Exemple de vidéoconférence où un modèle virtuel est utilisé.

Figure 1.3. Exemple de l'obtention du modèle virtuel selon [84].

Figure 2.1. Correspondance entre des points 2D sur l'image et 3D sur le modèle.

Figure 2.2. Exemple d'estimation des mouvements rigides.

Figure 2.3. Exemple de localisation des éléments non rigides.

Figure 2.4. Diagramme démontrant les modules nécessaires pour l'ensemble du projet.

Figure 3.1. Exemple du suivi de N points ($p_0, p_1, \ldots p_{N-1}$).

Figure 3.2. Diverses façons de parcourir la fenêtre de recherche R_{SSD}.

Figure 3.3. Recherche de la position de p_i.

Figure 3.4. Géométrie utilisée pour obtenir un modèle m_k.

Figure 3.5. Limites de l'orientation de la tête.

Figure 3.6. Exemple des M régions retenues à partir d'un point initial p_i sur I_0.

Figure 3.7. Calculer la valeur d'un pixel à une certaine orientation.

Figure 3.8. Obtenir l'intensité désirée d'un pixel en fonction des voisins.

Figure 3.9. Utilisation des modèles pour le suivi.

Figure 4.1. Exemple du mouvement rigide recherché.

Figure 4.2. Exemple où il faut recueillir les bons points du suivi pour estimer le mouvement rigide.

Figure 4.3. Exemple de la recherche d'une droite parmi des points cohérents et incohérents.

Figure 4.4. Divers essais effectués avec RANSAC pour trouver la meilleure droite.

Figure 4.5 Raffinement de la droite avec l'ensemble p.

Figure 4.6. Schéma des étapes de la méthode d'estimation des mouvements rigides.

Figure 4.7. Positions potentielles des points du suivi.

Figure 4.8. Positions utilisées des points du suivi après avoir effectué des tests de stabilité.

Figure 4.9. Refus d'un mouvement trop brusque.

Figure 4.10. Démonstration des distances entre les points P_i et p_i.

Figure 4.11. Exemple du calcul de M_f avec les points récupérés avec m_p.

Figure 4.12. Exemple de relocalisation d'un point erroné.

Figure 5.1. Schéma du fonctionnement géréral du suivi des éléments non rigides.

Figure 5.2. Localisation des yeux ouverts et fermés.

Figure 5.1. La détection des groupes d'iris potentiels G_{IG} et G_{ID}.

Figure 5.4. Création du modèle M_I pour détecter les iris potentiels.

Figure 5.5. Exemple d'utilisation des modèles M_I et M_{grad}.

Figure 5.6. L'obtention des hauteurs h_{haut} et h_{bas} pour l'ouverture de l'oeil.

Figure 5.7. Les images des vallées et des sommets pour des yeux ouverts et fermés.

Figure 5.8. L'obtention de la concentration des vallées à l'intérieur de la région visible de l'iris.

Figure 5.9. Modèle géométrique pour obtenir les régions de recherches R_G et R_D.

Figure 5.10. Quelques exemples où des sommets indésirés sont rencontrés.

Figure 5.11. Quelques exemples où des reflets d'un blanc très vif sont rencontrés.

Figure 5.12. Exemples de localisation de l'iris.

Figure 5.13. Le modèle géométrique pour les deux yeux ouverts.

Figure 5.14. Paramètres nécessaires pour définir une parabole P.

Figure 5.15. La représentation géométrique des informations pour un oeil.

Figure 5.16. Exemple d'adaptation du modèle géométrique du contour de l'oeil.

Figure 5.17. Des cas rencontrés où la localisation a mal été effectuée.

Figure 5.18. Les paramètres nécessaires pour construire une parabole P_f.

Figure 5.19. Exemple de l'adaptation des paraboles pour la fermeture des yeux.

Figure 5.20. Des cas où la localisation a été mal effectuée.

Figure 5.21. Correction de la localisation dans le cas où il y a un œil ouvert et l'autre fermé.

Figure 5.22. Exemples de la correction d'erreur.

Figure 5.23. Exemple du suivi des yeux à partir de la détection initiale.

Figure 5.24. Exemple d'image des sourcils.

Figure 5.25. Mesures utilisées pour trouver la position du centre du sourcil.

Figure 5.26. Positionnement des segments de droite S_g et S_d.

Figure 5.27. Les étapes pour trouver les sourcils à l'aide de l'image des vallées.

Figure 5.28. Exemple du suivi des sourcils sur une séquence d'images.

Figure 5.29. Quelques exemples où les sourcils sont mal localisés.

Figure 5.30. Modèle géométrique pour une bouche ouverte.

Figure 5.31. Modèle géométrique pour une bouche fermée.

Figure 5.32. Quelques tentatives d'extration des lèvres en utilisant la couleur.

Figure 5.33. Géométrie utilisée à l'aide de l'anthropométrie du visage pour obtenir C_b et R_{med}.

Figure 5.34. Démonstration de l'obtention de C_{med} à l'aide d'histogrammes.

Figure 5.35. Démonstration pour obtenir les fenêtres de recherche R_g et R_d.

Figure 5.36. Un exemple du test effectué pour éliminer des pixels.

Figure 5.37. Deux exemples de la distribution des gradients.

Figure 5.38. Le modèle M_c de distribution des gradients.

Figure 5.39. Des exemples où les coins de la bouche ont été détectés.

Figure 5.40. Localisation de la parabole P_c de la fermeture à l'intérieur de l'intervalle défini.

Figure 5.41. Deux exemples de la distribution des gradients sur le contour de la bouche.

Figure 5.42. Quatre régions d'image prélevées sur I_0 à l'aide d'un modèle géométrique.

Figure 5.43. Quelques exemples de variations rencontrées pour les coins d'une même bouche.

Figure 5.44. Quelques exemples d'instabilités de positions qui ont été corrigées.

Figure 5.45. Quelques exemples d'adaptation de la bouche pour diverses situations.

Figure 5.46. Des exemples de l'adaptation sur des images floues.

Figure 5.47. Des exemples où la bouche est mal localisée.

Figure A.1. Première utilisation simple du SSD.

Figure A.2. Première utilisation de l'algorithme proposé utilisant le SSD et le NCC.

Figure A.3. Deuxième utilisation simple du SSD.

Figure A.4. Deuxième utilisation de l'algorithme proposé utilisant le SSD et le NCC.

Figure A.5. Résultats du mouvement rigide pour la 1ère séquence.

Figure A.6. Résultats du mouvement rigide pour la 2e séquence.

Figure A.7. Résultats du mouvement rigide pour la 3e séquence.

Figure A.8. Résultats du mouvement rigide pour la 4e séquence.

Figure A.9. Résultats sur la première séquence.

Figure A.10. Résultats sur la deuxième séquence.

Figure A.11. Résultats sur la troisième séquence.

Figure A.12. Résultats sur la quatrième séquence.

Figure A.13. Résultats sur la cinquième séquence.

Figure A.14. Certaines proportions idéales selon [84].

Figure A.15. Proportions du visage utilisées selon [1].

Figure A.16. Géométrie de quelques éléments du visage.

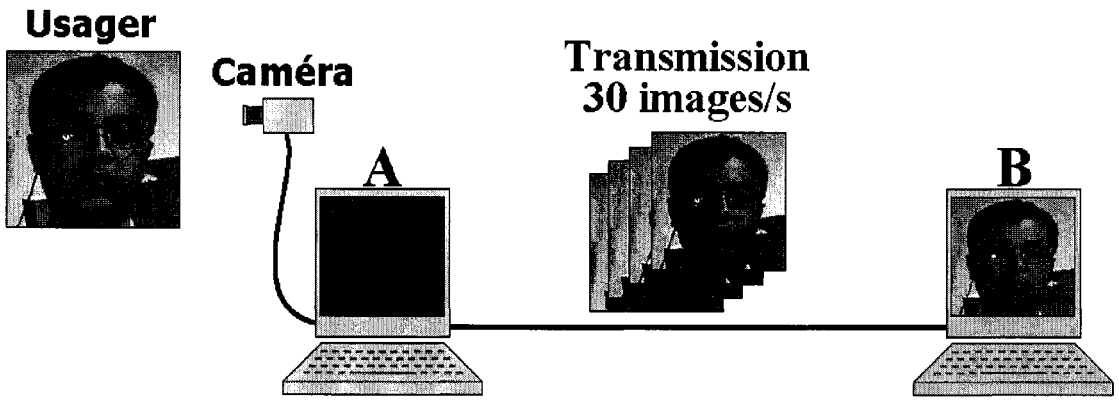
CHAPITRE 1

INTRODUCTION

Ce projet porte sur un système de traitement de séquences d'images pour des applications à la vidéoconférence de type « tête-épaules ». Le but est la diminution éventuelle du taux de transmission d'informations engendré lors du transfert des séquences d'images d'un ordinateur à l'autre. Au lieu d'utiliser des méthodes classiques de compression d'images basées uniquement sur le traitement des pixels, certaines informations intrinsèques de la scène réelle sont plutôt recherchées, tels que les mouvements rigides de la tête de l'usager et les mouvements non rigides du visage. Pour la vidéoconférence, un modèle virtuel serait utilisé afin de reproduire l'image de l'usager à la réception. Ce modèle devrait donc avoir une apparence physique semblable à celle de l'usager. Tout au cours de la vidéoconférence, les mouvements de l'usager doivent être recueillis à l'aide d'opérations de traitements d'images et, lorsque ces données sont transmises, les mouvements peuvent être imposés au modèle virtuel, ceci permet de représenter l'usager d'une façon assez réaliste tout en ayant un très faible débit de transmission. Il est à noter que le présent projet ne consiste qu'à prélever les mouvements de la tête et localiser des éléments du visage d'un usager. L'élaboration d'un système de vidéoconférence complet n'est pas traitée.

La figure 1.1. illustre une situation de vidéoconférence où les images sont envoyées sans traitement, ceci engendre un débit de transmission élevé. La figure 1.2.

illustre le cas où ce débit est considérablement diminué grâce à l'utilisation d'un modèle virtuel.



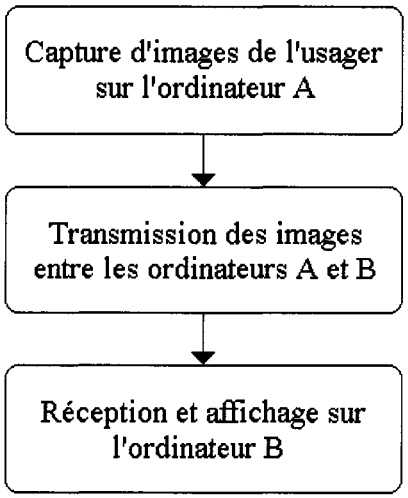
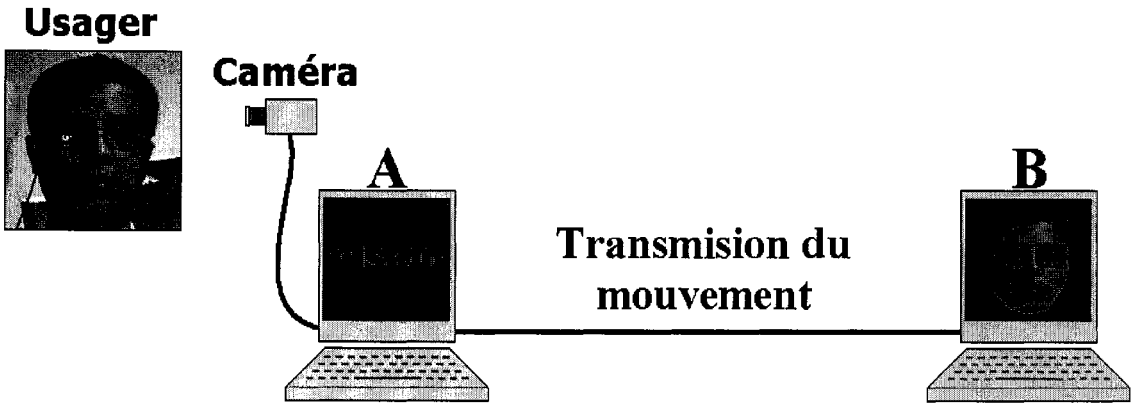


Figure 1.1. Exemple de vidéoconférence sans traitement d'images.



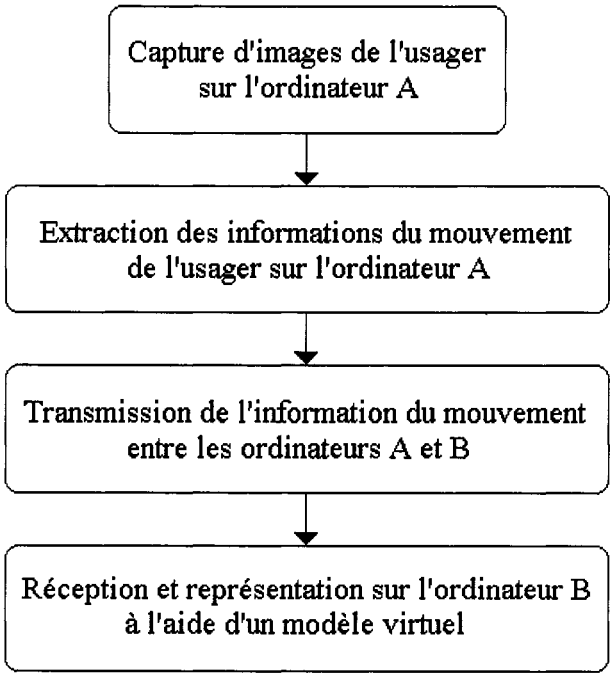


Figure 1.2. Exemple de vidéoconférence où un modèle virtuel est utilisé.

La difficulté de ce projet provient du fait que le système n'a pas de connaissance à priori de l'usager. Peu de contraintes sont imposées sur ce dernier, à part qu'il doit être bien visible et avoir une anatomie semblable à la majorité de la population. Pour un même individu, les images peuvent grandement varier en fonction de la position, de l'expression, de l'éclairage, etc. De plus, les individus peuvent être très différents les uns des autres. Afin de pouvoir estimer de façon automatique les mouvements rigides et non rigides d'un visage à partir d'une séquence vidéo, il est important de pouvoir extraire certaines contraintes peu variantes. L'anthropométrie du visage a donc été étudié ainsi que plusieurs techniques exploitant les proportions du visage. Afin que le système soit efficace, les divers traitements d'images nécessaires doivent idéalement être effectués en temps réel, soit 30 images par seconde.

Plusieurs travaux existants s'intéressent à cette problématique. Afin d'obtenir un modèle virtuel de l'usager, certains partent d'un modèle grossier constitué, par exemple, d'une simple sphère ou d'un ellipsoïde. D'autres modèles beaucoup plus précis peuvent être obtenus avec plusieurs images de l'usager et avec un arrière-plan contrôlé ou bien avec une intervention manuelle.

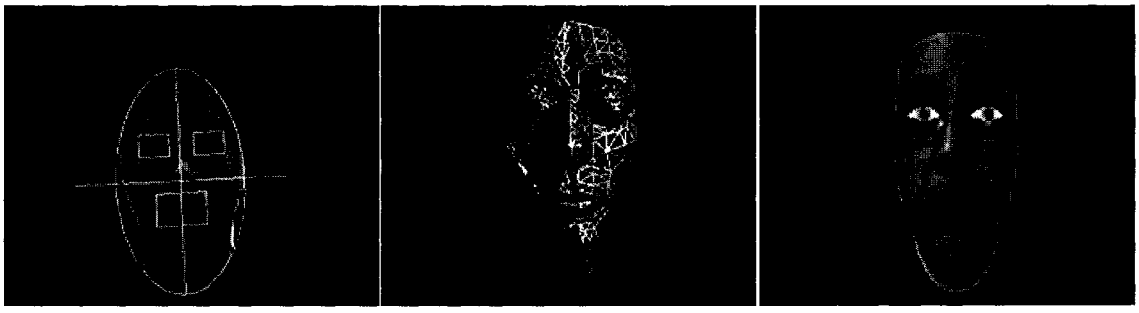
Pour estimer les mouvements rigides, plusieurs techniques effectuent un suivi de caractéristiques faciales d'image en image. Ce suivi peut concerner des points en 2D de l'image qui ont une correspondance en 3D sur le modèle virtuel. Le suivi peut aussi être effectué sur l'ensemble de l'image : le modèle virtuel doit donc être ajusté afin de diminuer la différence entre l'image de la projection du modèle et l'image réelle.

Pour estimer les mouvements non rigides, plusieurs techniques utilisent des modèles géométriques qui doivent s'adapter en subissant des déformations contrôlables. Ces modèles sont souvent utilisés pour se positionner sur les yeux, les sourcils et la bouche. D'autres techniques essaient plutôt d'analyser le visage en général afin de détecter l'expression faciale.

Le présent ouvrage constitue la suite des travaux effectués par [84] et pour des explications détaillées, le lecteur est invité à les consulter. En général, ces travaux consistaient à construire un modèle virtuel à partir de l'image initiale de l'usager, à estimer les mouvements rigides de la tête au cours d'une séquence et ensuite à représenter ces mouvements à l'aide du modèle virtuel (figure 1.3). La localisation des éléments non rigides était traitée grossièrement pour initialiser le système mais pas pour lire les mouvements au cours de la séquence. Sur la première image, la couleur peau du visage était extraite et une ellipse était localisée pour englober cette couleur. Ceci permettait d'avoir la hauteur et la largeur du visage et seuls ces deux paramètres étaient utilisés pour adapter la géométrie du modèle. Ensuite, certains éléments du visage, tels les yeux et la bouche, étaient recherchés pour permettre la segmentation du visage en régions. Ces régions permettaient d'adapter la texture du visage au modèle pour en augmenter le réalisme. Avant de commencer l'estimation du mouvement rigide, des points étaient sélectionnés sur le visage selon les endroits riches en texture. Ces points avaient ainsi une correspondance tridimensionnelle sur le modèle. Durant la séquence d'images, les points du visage étaient suivis en comparant les régions autour de ces points entre deux images successives. La méthode de comparaison était la SSD ou

« Sum of Squares Differences », ceci sera modifié au cours du présent ouvrage afin d'améliorer les résultats. Tous les points du suivi étaient utilisés pour estimer le mouvement rigide et les résultats étaient ensuite filtrés avec un filtre Kalman afin d'adoucir le mouvement. Finalement, le mouvement estimé était représenté par le modèle virtuel.

La suite de ces travaux, abordée dans ce mémoire, consiste d'une part à apporter des améliorations aux travaux de [84]. Ces améliorations concernent la précision à long terme du suivi de points 2D sur le visage, ceci est nécessaire pour l'estimation du mouvement rigide. De plus, ce ne sont plus tous les points du suivi qui seront utilisés pour estimer le mouvement, ceci afin d'éviter ceux erronés qui peuvent engendrer des mouvements très imprécis. Les résultats du mouvement rigide ne sont cependant pas comparés. Une méthode est finalement proposée pour l'évaluation des mouvements non rigides.



a) b) c)

Figure 1.3. Exemple de l'obtention du modèle virtuel selon [84]. En a), des mesures sont recueillies sur l'image initiale. En b), les dimensions d'un modèle 3D sont ajustées. En c), la texture du visage est appliquée au modèle.

L'évaluation des mouvements rigides sera effectuée grâce au suivi de points sur les images. Ces points auront une correspondance sur le modèle virtuel, ceci permettra d'estimer la localisation et l'orientation en 3D de la tête de l'usager sur chaque image.

Pour les mouvements non rigides du visage, des modèles géométriques déformables représentant les yeux, les sourcils et la bouche seront positionnés sur chaque image. Les paramètres ainsi obtenus à partir de la superposition des modèles sur les images permettront d'évaluer les mouvements.

Lorsque les mouvements rigides et non rigides sont obtenus, le modèle virtuel peut ainsi être éventuellement adapté pour représenter l'usager réel. Contrairement à plusieurs techniques déjà étudiées, le système de ce projet n'a aucune connaissance à priori de l'usager et une seule caméra, de qualité amateur, est utilisée. De plus, il n'y a aucun contrôle de l'arrière-plan et le système doit pouvoir fonctionner pour une grande variété d'individus.

Voici donc un résumé des objectifs à atteindre dans ce projet :

1. Développer une méthode permettant le suivi de points 2D sur le visage. Cette méthode doit être plus précise à long terme que celle utilisée dans [84];
2. Développer une nouvelle méthode permettant d'estimer le mouvement rigide de la tête. Contrairement à [84], cette méthode ne devra utiliser que les bons points du suivi et avoir un mécanisme de correction pour pouvoir continuer l'estimation à long terme;
3. Développer des méthodes permettant de localiser les yeux, les sourcils et la bouche d'un usager.

Et voici les méthodes de validation respectives qui seront utilisées pour vérifier l'atteinte des objectifs fixés :

1. En utilisant les mêmes séquences d'images avec les mêmes points initiaux pour le suivi, comparer la précision obtenue entre la méthode de [84] et la nouvelle méthode proposée;
2. En appliquant l'estimation du mouvement rigide sur plusieurs séquences d'images, analyser les résultats obtenus d'une façon qualitative;
3. En appliquant la localisation des éléments non rigides sur plusieurs séquences d'images, analyser les résultats obtenus d'une façon qualitative.

Le chapitre 2 décrit les contraintes du projet ainsi qu'une étude sur les mouvements rigides et non rigides. Une description générale des modules nécessaires au projet est fournie. Une distinction est effectuée entre les modules effectués par [84] et ceux réalisés dans ce mémoire.

Dans le chapitre 3, le suivi de points 2D sur les images de l'usager afin d'estimer le mouvement rigide est étudié. Une technique est proposée et une comparaison est effectuée avec les résultats expérimentaux obtenus par [84].

Dans le chapitre 4, l'estimation des mouvements rigides 3D de la tête de l'usager est étudiée. Bien que cette estimation ait déjà été effectuée par [84], une autre méthode est proposée afin d'être plus robuste aux erreurs engendrées par le suivi des points 2D. Une analyse qualitative est effectuée mais la précision des résultats entre les deux approches n'est cependant pas comparée.

Le chapitre 5 concerne le suivi d'éléments non rigides tels que les yeux, les sourcils et la bouche. Ce suivi est effectué à l'aide de modèles géométriques déformables représentant des yeux ouverts ou fermés, des sourcils de différentes positions et une bouche ouverte ou fermée. La méthode développée permet une très grande liberté sur les mouvements. Une analyse qualitative est effectuée.

Ce mémoire se termine avec une conclusion sur le projet développé.

CHAPITRE 2

STRUCTURE GÉNÉRALE DU SYSTEME

Ce chapitre fournit la structure générale du système de ce projet. D'après les contraintes imposées et des études effectuées sur d'autres travaux existants, des méthodes ont pu être élaborées afin d'évaluer les mouvements rigides et localiser les éléments non rigides. Une étude sur les méthodes existantes ainsi qu'une brève description des modules utilisés sont également fournies.

2.1. Les contraintes au projet

Les principales contraintes technologiques sont les suivantes :

- Le système doit fonctionner en temps réel sur un ordinateur personnel technologie courante ;
- Les images utilisées sont produites par des caméras numériques non-professionnelles (du type « web-cam ») ;
- La résolution des images est de 320 x 240 pixels sur 24 bits de couleur en mode RGB. Cela permet d'avoir une bonne portabilité avec les caméras « web-cam » et avoir une bonne rapidité d'exécution.

Les principales contraintes d'environnement sont les suivantes :

- L'usager doit adopter une position initiale pour la première image : être situé environ à 50 centimètres de la caméra pour permettre un bonne visibilité, être de

position frontale avec le visage au repos et permettre une bonne visibilité de la tête entière ;

- L'éclairage de la scène n'est pas contrôlé de façon précise, mais doit permettre une assez bonne visibilité de l'usager ;
- L'arrière-plan de la scène n'est pas contrôlé mais un seul usager doit être présent dans la scène.

Les principales contraintes de robustesse sont les suivantes :

- Le système doit pouvoir être en mesure de traiter une assez grande population d'usagers ;
- Le système doit pouvoir être en mesure de traiter les divers mouvements de la tête et du visage au cours d'une vidéoconférence.

2.2. Études pour l'évaluation des mouvements rigides

Beaucoup d'articles ont été étudiés pour évaluer les mouvements rigides de la tête d'un usager. En plus de la précision des mesures, la stabilité à long terme est très importante afin de suivre l'usager le plus longtemps possible. L'évaluation du mouvement consiste à prélever, sur chaque image d'une séquence, les trois translations et rotations autour des axes X, Y et Z, selon une référence tridimensionnelle. Lorsqu'il s'agit d'une localisation grossière de la tête, ce ne sont habituellement que les translations selon les axes X et Y qui sont recherchées. Les mouvements sont relatifs à la position initiale de l'usager sur la première image. L'une des principales difficultés est qu'une seule caméra doit être utilisée pour observer l'usager en mouvement.

2.2.1. Méthodes cherchant une localisation grossière de la tête

Ces méthodes peuvent être utilisées afin de délimiter une région de l'image pour les traitements afin de diminuer l'effort de calcul des traitements subséquents. Seuls les translations selon les axes X et Y peuvent être obtenues. Ceci est incomplet pour les objectifs à atteindre dans ce projet.

Dans [18], une ellipse est positionnée au passage de forts gradients de l'image. Ces forts gradients correspondent au contour de la tête. L'ellipse est verticale et les proportions restent fixes, il y a donc trois paramètres variables pour la géométrie : la translation horizontale et verticale et l'homothétie. La couleur peau du visage est d'ailleurs recherchée et celle-ci doit se situer à l'intérieur de l'ellipse.

Dans [71], les modèles déformables sont utilisés pour localiser plusieurs parties du corps (tête, mains, etc). Par exemple, le modèle de la tête est constitué d'une trajectoire fermée, plus ou moins circulaire, composée de plusieurs segments de droite connectés entre eux à leurs extrémités. Plusieurs paramètres sont donc nécessaires pour configurer ces modèles. Pour simplifier les recherches, les gradients de l'image sont grandement utilisés pour la convergence des paramètres vers les valeurs optimales. Des contraintes sont d'ailleurs imposées à la trajectoire afin d'avoir une forme semblable à une tête.

Avantages de ces méthodes :

- Les traitements sont généralement rapides ;
- L'effort de calcul des traitements ultérieurs est minimisé car une position approximative est obtenue.

Inconvénients de ces méthodes :

- Les positions obtenues sont grossières ;
- La rotation autour des axes et la translation en profondeur ne sont pas obtenues.

2.2.2. Méthodes comparant la projection du modèle et l'usager

Ces méthodes obtiennent les paramètres du mouvement 3D en comparant la projection du modèle de l'usager avec l'image courante. Les paramètres du mouvement permettant d'avoir la plus faible différence entre la projection et l'image sont donc ceux recherchés. La minimisation par moindres carrés est généralement utilisée pour quantifier cette différence. Il est à noter que l'image donnée par la projection du modèle doit être très semblable à l'image de l'usager pour pouvoir utiliser de telles méthodes. En effet, la minimisation par moindres carrés est peu robuste aux changements de contraste et d'intensité. Du matériel supplémentaire, telle une carte pour le rendu graphique 3D, est généralement utilisé.

Dans [9], une ellipsoïde est utilisée pour le modèle. Celle-ci est initialement déformée afin d'englober la tête de l'usager selon des poses spécifiques sur trois images initiales de l'usager. Le visage de l'usager est ensuite appliqué sur l'ellipsoïde pour permettre les comparaisons.

Dans [10][11], une image initiale de l'usager est utilisée afin de projeter le visage sur une sphère ou une structure formée de polygones, ceci constitue le modèle. Le système suppose cependant que le visage sur l'image initiale ait été récupéré manuellement.

Avantages de ces méthodes :

- L'image entière de la tête est utilisée pour positionner le modèle, ceci donne une bonne robustesse même si de petites régions du visage sont mal représentées par le modèle.

Inconvénients de ces méthodes :

- Au départ, il faut s'assurer que le modèle représente fidèlement l'usager afin de bien effectuer la comparaison entre la projection du modèle et l'usager ;
- Le rendu 3D du modèle est lourd en calcul parce qu'en plus de traiter la géométrie, il faut traiter la texture et tenir compte de l'éclairage ;
- Puisque plusieurs paramètres sont nécessaires pour représenter les rotations autour des axes et les translations, la comparaison entre la projection du modèle et l'image de l'usager doit s'effectuer par une méthode de minimisation par descente du gradient, ceci peut converger vers de faux minimums si les mouvements sont rapides ;
- Il n'est pas prévu que l'usager puisse effectuer des mouvements non rigides.

2.2.3. Méthodes cherchant le déplacement d'un objet par flux optique

Le flux optique permet d'obtenir un champ de vecteurs dans le but de représenter le déplacement d'objet entre des images. Le flux optique est parfois utilisé pour déterminer les changements de position de la tête de l'usager, soit la translation horizontale et verticale. Après avoir estimé la nouvelle position, un raffinement doit généralement être effectué pour en augmenter la précision.

Dans [6], une représentation multi résolutions de l'image de l'usager est obtenue, ceci permet de chercher la tête avec beaucoup d'essais sur des images de moins haute résolution afin de gagner en temps de calcul. Les informations obtenues par le flux optique sont utilisées pour mieux localiser les recherches.

Dans [9], le flux optique est utilisé pour estimer les mouvements rigides et permet d'interpoler des images entre celles fournies. Il y a ainsi plus d'images à traiter mais les différences entres celles-ci sont plus faibles. La recherche de régions semblables est appliquée entre 2 images consécutives de la nouvelle séquence ainsi qu'entre les images et la projection en 2D du modèle 3D. Le modèle est donc positionné en essayant de minimiser la différence des régions d'image. Le flux optique est d'ailleurs encore une fois utilisé pour renforcer l'algorithme en permettant de mieux localiser les recherches.

Avantages de ces méthodes :

- Le flux optique fournit une bonne indication de la direction du mouvement d'un objet rigide comme la tête ;
- La tête entière est recherchée à l'aide du flux optique, ceci permet d'avoir une bonne robustesse si certaines régions de l'image ne sont pas fiables.

Inconvénients de ces méthodes :

- Le flux optique peut être très erroné si les mouvements rigides sont de grande amplitude entre deux images successives ;
- Le bruit dans les images et les mouvements non rigides peut donner un flux optique indésiré.

2.2.4. Méthodes invariantes à l'éclairage

Si l'on suppose que l'éclairage sur l'usager peut varier beaucoup à cause des mouvements rigides, certaines techniques permettent de faire des initialisations de l'usager sous divers éclairages afin de rendre le suivi plus robuste dans de telles conditions. Cette initialisation consiste à prendre plusieurs images de l'usager en variant la source d'éclairage de façon spécifique. Les diverses images retenues sont ensuite analysées pour former une base. Les images sous divers éclairages peuvent ensuite être reproduites plus ou moins fidèlement en utilisant cette base et variant certains paramètres. Pour retrouver une région d'image ayant subi des variations d'éclairage, il faut donc trouver les bons paramètres de la base utilisée.

Dans [23], des régions du visage de l'usager sont utilisées pour faire un suivi de points, ceci permettra d'évaluer le mouvement rigide. Le suivi est plus robuste à l'éclairage grâce aux régions qui ont été initialisées.

Dans [27], un demi-cylindre est utilisé pour estimer la forme 3D du visage. L'image du visage est projetée initialement sur ce demi-cylindre et une initialisation à l'éclairage est d'ailleurs effectuée. Le mouvement rigide est estimé en trouvant la bonne projection du demi-cylindre sur l'image de l'usager, en tenant compte de l'éclairage.

Avantages de ces méthodes :

- Une grande flexibilité concernant l'éclairage peut être appliquée sur l'usager.

Inconvénients de ces méthodes :

- Une initialisation sous divers éclairages doit d'abord être effectuée ;

- Dans le cas où les variations d'éclairage sur l'usager sont négligeables, beaucoup de traitements seront effectués inutilement.

2.2.5. Méthodes basées sur le suivi de points 2D

Ces méthodes effectuent sur l'image le suivi de points 2D qui ont une correspondance sur un modèle 3D connu [12][22]. Pour simplifier les traitements, la scène est considérée comme une projection orthographique. Un minimum de trois points sur l'usager est nécessaire mais plus de points permettent une meilleure stabilité. La position des points sur l'image permet donc d'estimer la position 3D du modèle. Pour effectuer le suivi de points, il est important que ces derniers soient situés à des endroits riches en texture pour ne pas être confondus avec d'autres éléments du voisinage [81][83]. Il est d'ailleurs important que les positions 3D soient assez fidèles au vrai visage de l'usager.

Avantages de ces méthodes :

- Une modèle virtuel détaillé de l'usager n'est pas nécessaire, seulement quelques points de référence en 3D ;
- Il est plus simple de faire le suivi de plusieurs petites régions d'image que de suivre la tête entière ;
- En effectuant le suivi sur des points stratégiques, les mouvements non rigides n'auront pas d'influence sur l'évaluation de mouvements rigides.

Inconvénients de ces méthodes :

- Il est difficile de trouver les points stratégiques sur l'image pour effectuer le suivi ;
- Pour chaque point du suivi, la correspondance 3D sur le modèle doit être précise afin d'éviter d'étranges configurations des mouvements ;
- Si l'usager effectue des mouvements de grandes amplitudes, des occlusions peuvent survenir pour les points suivis.

2.3. Études pour la localisation des éléments non rigides

Bien que l'importance de la localisation des éléments non rigides pour vidéoconférence ait été mentionnée par [84], elle n'avait pas été étudiée en profondeur. Beaucoup d'articles ont donc été étudiés pour localiser ces éléments. La détection des expressions faciales a aussi été étudiée. La localisation consiste à obtenir le contour des éléments sur l'image. Par exemple, la localisation de l'iris consisterait en un cercle de même rayon et localiser vis-à-vis du centre de l'iris. Dans ce projet, la localisation des yeux, des sourcils et de la bouche est recherchée. Le choix des modèles utilisés pour la localisation n'est pas spécifiques et sera établie en fonction des difficultés rencontrées. Par exemple, il peut être plus simple de localiser un cercle sur l'iris plutôt que de localiser une forme quelconque sur le contour réel de l'iris. Cependant, les modèles utilisés devront être très représentatifs des éléments recherchés.

2.3.1. Méthodes adaptant des modèles géométriques connus

Ces méthodes utilisent des modèles géométriques déformables pour s'adapter sur les éléments du visage tels les yeux, les sourcils et la bouche. Un œil, par exemple, peut être représenté par deux paraboles (le contour) et un cercle (l'iris). Les modèles sont donc connus et sont caractérisés par des paramètres. Ces derniers doivent être estimés afin d'obtenir les déformations désirées pour les modèles. Par exemple, si un cercle est utilisé comme modèle, les paramètres constituent le centre et le rayon. Ces paramètres pourraient être ajustés afin de maximiser le passage sur des arêtes par exemple, à l'aide d'une image traitée représentant les arêtes incluant celles autour de l'iris. Souvent les modèles sont cependant plus complexes et plus de paramètres doivent être déterminés. Pour estimer les paramètres optimaux, des techniques de minimisation, comme la descente du gradient, doivent être appliquées. Bien que des problèmes de convergence soient souvent rencontrés, d'autres techniques visent à contourner ces problèmes [10][78].

Dans [10], les modèles sont utilisés pour s'adapter aux sourcils, aux yeux et à la bouche. D'autres traitements sont effectués afin d'éviter les problèmes de convergence : afin d'obtenir des points de repères, les coins des yeux, de la bouche et de l'intérieur des sourcils sont suivis d'image en image, l'ouverture des yeux est calculée à l'aide de la variation d'intensité, etc.

Dans [63], la couleur peau ainsi que les gradients sont exploités afin de déterminer le contour extérieur des lèvres. Puisque les lèvres ne doivent pas être de

couleur peau, celles-ci se distinguent du visage grâce à la couleur. Une trajectoire fermée composée de B-splines est utilisée pour représenter la bouche. Cette trajectoire doit maximiser un passage sur de forts gradients tout en englobant la couleur rouge.

Dans [72], les modèles sont utilisés pour s'adapter aux yeux et à la bouche. Pour un œil, le modèle est constitué de deux paraboles (pour le contour) et d'un cercle (pour l'iris). Pour une bouche ouverte, le modèle est constitué de 5 paraboles (dont deux pour le contour extérieur du haut). 4 paraboles sont utilisées pour une bouche fermée. Pour converger sur l'image, ces modèles doivent minimiser une certaine fonction d'énergie à l'aide d'une méthode de descente du gradient. Ces fonctions sont principalement basées sur le passage sur les arêtes, les vallées et les sommets de l'image.

Dans [76], un modèle est utilisé pour s'adapter à la bouche. Le modèle est constitué de 4 paraboles pour une bouche ouverte et de 3 pour une bouche fermée. Les coins de la bouche, les arêtes des contours des lèvres ainsi que la couleur rouge de la bouche sont utilisées pour adapter le modèle.

Dans [77], un modèle de la bouche doit s'adapter en maximisant le passage sur des arêtes de fortes amplitudes, qui constituent le contour de la bouche. Seules les arêtes horizontales sont recherchées à l'aide d'un opérateur simple. Les coins de la bouche sont d'abord localisés pour faciliter la recherche des arêtes.

Dans [78], un modèle constitué de 4 paraboles est utilisé pour la bouche. Ce modèle doit principalement englober la couleur rouge de la bouche et passer par de forts

gradients. Un réseau de neurones est entraîné à l'aide des expressions connues. Ce réseau doit classifier les trois expressions « sourire », « neutre » et « triste ».

Avantages de ces méthodes :

- En utilisant ces modèles, la liberté est limitée, en ce qui concerne la géométrie possible pour un élément du visage, ceci évite d'avoir d'étranges modèles ne ressemblant plus du tout à l'élément recherché ;
- En plus de la géométrie du modèle, plusieurs caractéristiques de l'image peuvent être utilisées à la fois, tels les arêtes, les vallées et les sommets de l'intensité, etc., ceci permet une recherche basée sur un ensemble de caractéristiques communes à l'élément recherché.

Inconvénients de ces méthodes :

- Puisqu'il faut beaucoup de paramètres pour représenter la géométrie d'un modèle, l'univers des possibilités de ces paramètres ne peut être essayé pour trouver les meilleurs, c'est-à-dire ceux qui minimisent une certaine fonction d'énergie. La recherche de ce minimum, souvent effectué par une descente du gradient, peut souvent conduire à d'autres minimums locaux non désirés, ceci donne une mauvaise adaptation du modèle ;
- Puisque les modèles doivent être représentés avec le moins de paramètres possibles, les formes géométriques doivent être simples, ceci empêche parfois une adaptation précise sur l'élément recherché.

2.3.2. Méthodes adaptant des trajectoires quelconques

Ces méthodes utilisent des trajectoires qui doivent se déformer afin de minimiser une certaine fonction d'énergie. Généralement, il s'agit de plusieurs segments de droite connectés entre eux à leurs extrémités. La fonction à minimiser, par exemple, peut consister à favoriser le passage d'une trajectoire sur des arêtes dans une image tout en imposant des contraintes géométriques. Pour un œil, par exemple, une forme plutôt elliptique de la trajectoire peut être imposée. Contrairement à l'utilisation de modèles géométriques connus, les méthodes suivantes essaient plutôt de paramétrer les trajectoires en positionnant plusieurs points de repère.

Dans [37], les trajectoires sont utilisées pour estimer la contraction de certains muscles du visage. La contraction estimée est fonction de la position de la trajectoire concernée. Ceci permet ensuite de configurer les muscles d'un modèle virtuel. Ces trajectoires sont situées sur les sourcils, le contour extérieur de la bouche, le bas du menton et l'intérieur des joues. Les gradients dans l'image sont principalement utilisés pour faire converger les trajectoires.

Dans [46], le contour des yeux est d'abord extrait en utilisant deux trajectoires fermées devant maximiser le passage sur les arêtes entre les coins des yeux. Le suivi est ensuite effectué en adaptant d'autres trajectoires sur ces contours. Celles-ci doivent minimiser la différence entre l'image courante et la précédente tout en respectant la géométrie du contour des yeux.

Dans [76][78], la couleur rouge de la bouche est recherchée. Des trajectoires fermées représentant la bouche sont adaptées. Ces trajectoires doivent englober la couleur rouge tout en passant par des arêtes ou des gradients. La couleur rouge spécifique des lèvres est cependant parfois difficile à obtenir.

Avantages de ces méthodes :

- Ces trajectoires n'ont pas de forme géométrique définie, elles peuvent donc s'adapter à un grand nombre de formes différentes ;
- Plusieurs caractéristiques de l'image peuvent être utilisées à la fois, tels les arêtes, les vallées et les sommets de l'intensité, etc. Ceci permet une recherche basée sur un ensemble de caractéristiques communes à l'élément recherché.

Inconvénients de ces méthodes :

- Puisqu'il n'y a pas de forme géométrique définie, la forme obtenue de la trajectoire peut être totalement différente de la forme de l'élément recherché. L'information géométrique n'est pas utilisée ;
- Il y a souvent trop de paramètres à contrôler simultanément. Il est donc souvent impossible d'essayer toutes les possibilités, même avec des paramètres discrétisés, pour trouver la meilleure configuration. La recherche du minimum de la fonction d'énergie établie est donc souvent effectuée par une descente du gradient, ceci peut conduire à des minimums locaux non désirés.

2.3.3. Études sur la détection d'expressions faciales

Certaines techniques recherchent des expressions faciales à l'aide d'une banque d'expressions. D'après les résultats de détection obtenus sur l'usager, une expression est ensuite recherchée dans la banque pour que cette dernière soit appliquée au modèle virtuel de l'usager. Plusieurs techniques se réfèrent à la FACS (Facial Action Coding System) [37].

Dans [29], les lèvres sont adaptées grâce à l'analyse du son de la voix. Selon les syllabes prononcées, des configurations de la bouche sont imposées grâce à la recherche dans une banque.

Dans [37], les contractions musculaires sont estimées à l'aide de trajectoires. Grâce à l'analyse de ces contractions, une expression faciale est déterminée à l'aide de FACS.

Dans [47], le flux optique est analysé entre une expression de l'usager et l'image du visage au repos. Selon le mouvement obtenu par le flux optique, une expression est ensuite recherchée à l'aide de la FACS.

Dans [51], plusieurs modèles géométriques sont utilisés pour représenter le front, les sourcils, les yeux, le haut des joues, l'intérieur des joues et la bouche. Un réseau de neurones est ensuite entraîné à l'aide des modèles obtenus en fonction de plusieurs expressions connues. La FACS est utilisée pour déterminer l'expression.

Avantages de ces méthodes :

- Puisque qu'une banque d'expressions est utilisée, l'expression détectée est facilement applicable pour un modèle virtuel de l'usager ;
- Au lieu de localiser séparément les divers éléments non rigides du visage, c'est l'ensemble qui est traité simultanément, ceci est plus robuste si certains éléments sont trop difficiles à traiter.

Inconvénients de ces méthodes :

- Ces méthodes ne prévoient pas les mouvements rigides que peut effectuer l'usager. Ces mouvements peuvent donc fausser les résultats ;
- La flexibilité des mouvements non rigides est diminuée.

2.4. Méthodes proposées pour ce projet

Après avoir consulté plusieurs articles, une méthode a été proposée pour estimer les mouvements rigides ainsi que pour localiser les éléments non rigides.

2.4.1. L'estimation des mouvements rigides

L'évaluation des mouvements rigides sera effectuée à l'aide du suivi de points 2D initialement localisés sur la première image. Ces points ont une correspondance 3D sur un modèle virtuel, ceci permettra d'évaluer les mouvements rigides représentés par 3 translations et 3 rotations autour des axes X, Y et Z. Cela est illustré à la figure 2.1. La nouvelle méthode a pour but d'obtenir une meilleure stabilité à long terme comparativement à [84], c'est-à-dire sur une longue séquence d'images.

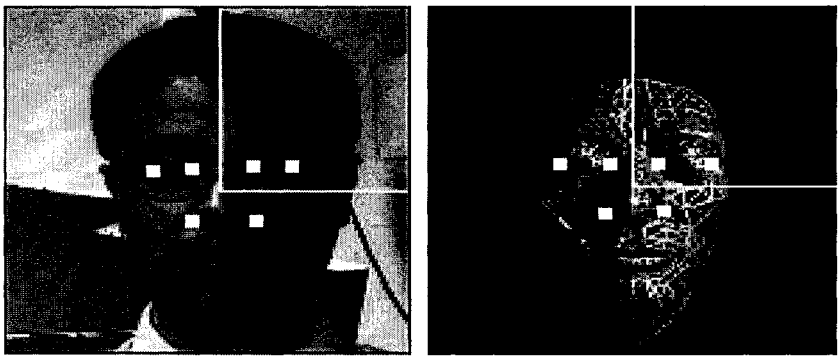


Figure 2.1. Correspondance entre des points 2D sur l'image et 3D sur le modèle.

Le modèle virtuel utilisé provient de [84] et consiste en un modèle formé de polygones représentant le visage d'un usager (figure 2.1). La géométrie du modèle est très peu flexible : les proportions des éléments à l'intérieur du visage sont fixes. D'après la détection du visage supposée déjà acquise, seules la hauteur et la largeur de la tête de l'usager sont adaptées au modèle. Dans ce présent ouvrage, le modèle virtuel est seulement utilisé pour prélever quelques points 3D permettant l'estimation du mouvement rigide. Les détails de ce modèle seraient nécessaires si la représentation virtuelle de l'usager était effectuée dans cet ouvrage, ce qui n'est pas le cas.

Puisque le modèle virtuel utilisé est assez grossier, il a été jugé préférable d'effectuer le suivi de quelques points 2D en correspondance sur le modèle au lieu de se fier à l'ensemble de la tête. Les méthodes comparant la projection du modèle et l'usager, par exemple, ne peuvent donc pas être utilisées car la comparaison serait trop mauvaise. Et puisqu'il est facile d'effectuer le suivi de points 2D, d'autres méthodes de localisation grossière de la tête ne sont pas nécessaires. De plus, une initialisation pour l'invariance

aux changements d'éclairage sur une séquence d'images semble inutile car ces changements sont négligeables. Avec la méthode proposée, il est cependant important d'avoir un bon suivi de points 2D situés sur des endroits stratégiques. Voici les contraintes imposées sur ces endroits afin d'augmenter la précision des résultats :

- Être bien visible malgré les mouvements rigides ;
- Être le moins sensible possible aux mouvements non rigides ;
- Avoir une correspondance précise sur le modèle virtuel de l'usager ;
- Être assez dispersés sur le visage.

Un modèle approximatif très grossier était proposé dans [84] et n'a pas été amélioré. Avec un tel modèle, les correspondances 3D des points 2D ont aussi une imprécision. L'imprécision du modèle virtuel utilisé provient principalement des raisons suivantes :

- Des études sur l'anthropométrie démontrent qu'il faut beaucoup de données afin de reconstruire un modèle virtuel [3][5] et que les proportions du visage varient beaucoup d'un individu à l'autre [2] ;
- Des imprécisions sont obtenues sur le peu de mesures prises de l'usager (les frontières de la tête, par exemple, sont mal délimitées) ;
- Un seul plan de vue de l'usager (de face) est utilisé pour recueillir les informations, les profondeurs sont donc inconnues.

Beaucoup d'effort a donc été mis à améliorer la précision du suivi de points 2D et l'estimation des mouvements rigides doit utiliser plusieurs points du suivi tout en sachant ignorer ceux étant trop imprécis. Ces derniers seront corrigés automatiquement pour continuer le suivi.

La figure 2.2 illustre un exemple de ce qui doit être obtenu pour les mouvements rigides. Sur la première image, l'usager a une pose initiale et des points sont placés pour le suivi. Grâce au suivi et à la correspondance 3D des points, les translations et rotations de la tête peuvent être estimées, ceci est montré sur la deuxième image. Le petit carré noir près du nez à gauche représente un point incohérent qui ne devrait pas être utilisé pour estimer le mouvement. Sur la troisième figure, un modèle virtuel utilise le mouvement estimé pour s'adapter à la pose réelle de l'usager.

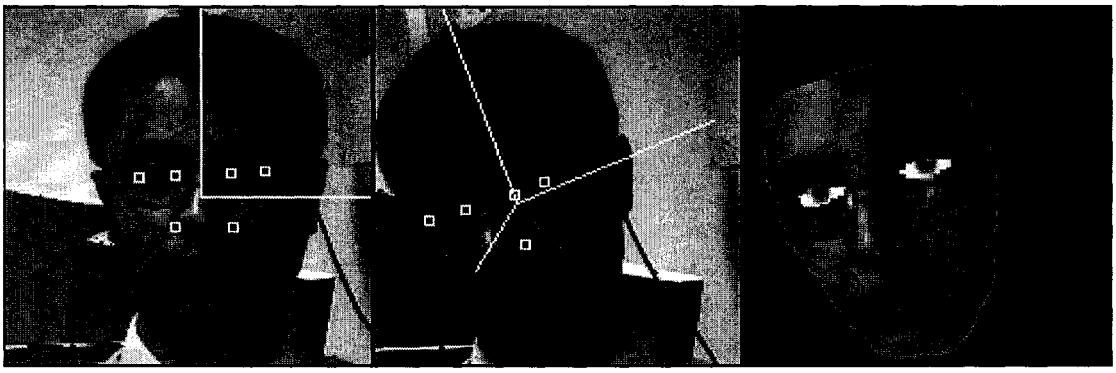


Figure 2.2. Exemple d'estimation des mouvements rigides. En a), la pose initiale de l'usager avec les points pour le suivi. En b), le suivi des points avec l'estimation du mouvement. En c), l'application du mouvement au modèle virtuel.

2.4.2. La localisation des éléments non rigides

L'adaptation de modèles géométriques a été utilisée pour localiser les éléments non rigides car ces modèles permettent beaucoup de liberté de mouvement, même lorsqu'il y a des mouvements rigides de la tête. Les trajectoires fermées ne sont pas utilisées car celles-ci semblent plus favorables dans les cas où la géométrie des éléments recherchés est inconnue, ce qui n'est pas notre cas. De plus, la détection d'expressions

faciales à partir d'une banque semble inappropriée pour ce projet puisque les mouvements rigides ne sont pas négligeables.

Les éléments non rigides recherchés se limitent aux deux yeux, aux deux sourcils et à la bouche car ceux-ci semblent suffisants pour exprimer les expressions du visage. Pour un œil, le modèle géométrique utilisé est constitué de deux paraboles et d'un cercle pour un œil ouvert et d'une parabole pour un œil fermé. Pour un sourcil, il s'agit de deux droites en forme de « Λ ». Pour la bouche, il s'agit de quatre paraboles pour une bouche ouverte et de 3 paraboles pour une bouche fermée. Les modèles utilisés sont à la fois simples et permettent de bien représenter les éléments recherchés. Seules les configurations probables des éléments non rigides seront traitées. On ne tolère pas de grimace par exemple.

D'après les paramètres obtenus de ces modèles, le modèle virtuel pourrait ensuite être adapté pour représenter les mouvements non rigides de l'usager. Ce n'est cependant pas couvert dans ce mémoire.

La figure 2.3 illustre un exemple de ce qui doit être obtenu.

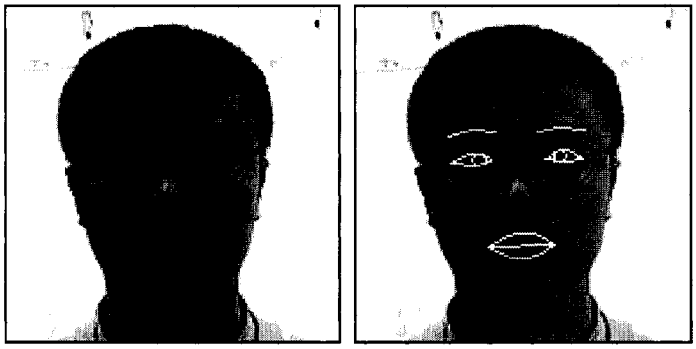


Figure 2.3. Exemple de localisation des éléments non rigides. En a), l'image originale de l'usager. En b), les modèles géométriques des sourcils, des yeux et de la bouche sont adaptés.

2.5. Les modules du système proposé

La figure 2.4 illustre un diagramme démontrant les relations entre les modules utilisés pour le système. Les modules en gris n'ont pas été conçus lors de ce projet mais devraient cependant être inclus dans l'ensemble du projet de la vidéoconférence. Une brève description de chaque module est d'ailleurs fournie mais la description plus détaillée sera donnée plus loin dans ce mémoire.

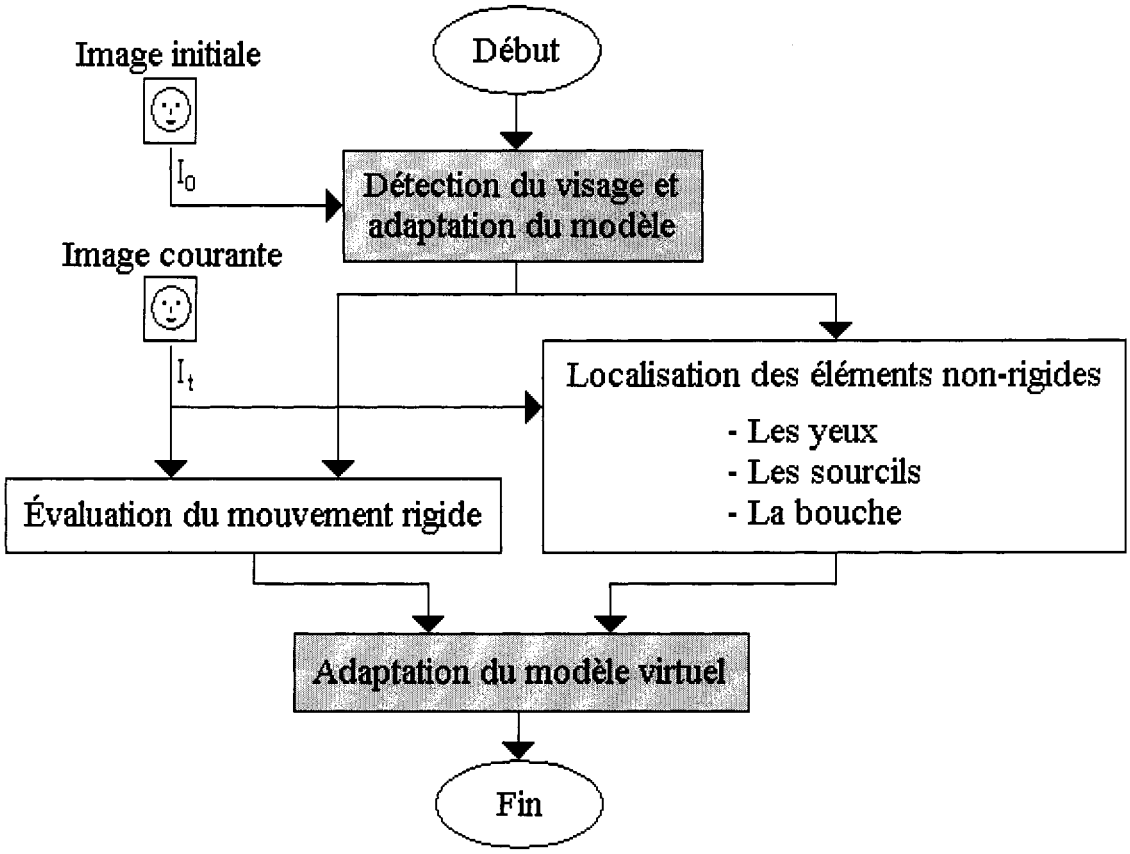


Figure 2.4. Diagramme démontrant les modules nécessaires pour l'ensemble du projet.

2.5.1. Détection du visage et adaptation du modèle

Ce module a été repris des travaux de [84]. Il s'agit de localiser la tête de l'usager sur l'image initiale. Ensuite certains éléments du visage, tels les yeux et la bouche, sont recherchés. Ces positions ne sont cependant qu'approximatives. Avec les mesures recueillies, un modèle virtuel est ensuite adapté pour représenter l'usager. Ce module n'est pas étudié de nouveau dans ce mémoire mais est nécessaire pour l'ensemble du projet. Il est d'ailleurs conseillé que, dans l'avenir, ce module soit remplacé par un autre permettant une plus grande précision sur les résultats.

2.5.2. Estimation du mouvement rigide

Ce module permet d'estimer le mouvement rigide à l'aide du suivi de points 2D. Ces points ont une correspondance 3D sur le modèle virtuel de l'usager. Le suivi de points 2D avait déjà été réalisé par [84] mais a considérablement été amélioré, comme décrit en détails au chapitre 3.

Le mouvement rigide est constitué de translations et de rotations selon les axes X, Y et Z. Le suivi est corrigé si la configuration obtenue par l'ensemble des points est peu probable, c'est-à-dire s'il y a des points 2D dont la correspondance 3D avec la position du modèle est mauvaise. L'estimation du mouvement rigide avait déjà été réalisée par [84] mais n'avait pas de système de correction, ceci pouvait engendrer une grande instabilité des résultats pour une longue séquence d'images. La nouvelle méthode proposée est décrite en détails au chapitre 4.

2.5.3. La localisation des éléments non rigides

Dans ce module, le contour des yeux, de l'iris, des sourcils ainsi que de la bouche sont localisés sur la séquence d'images grâce à l'adaptation de modèles géométriques déformables. Pour faciliter le suivi, la géométrie utilisée pour un modèle est basée sur l'anthropométrie du visage.

Les yeux sont localisés en premier et serviront ensuite de points de référence pour la détection des sourcils et de la bouche. Le modèle varie selon qu'un oeil est ouvert ou fermé. Juste au-dessus des yeux détectés, un modèle est ensuite adapté pour chaque sourcil. Dans un intervalle en bas des yeux détectés, les coins de la bouche sont d'abord recherchés et ensuite, le modèle géométrique de la bouche est adapté. Le modèle varie selon que la bouche est ouverte ou fermée. Ceci est décrit en détails au chapitre 5.

2.5.4. Adaptation du modèle virtuel

Ce module n'a pas été développé dans ce travail et consisterait à appliquer les mouvements rigides et non rigides au modèle virtuel de l'usager. La représentation des mouvements rigides a cependant déjà été effectuée par [84], il s'agissait en fait de positionner le modèle en fonction des translations et rotations estimées. L'adaptation du modèle constitue plutôt un travail technique qu'une recherche scientifique. Pour adapter les mouvements non rigides au modèle, il faudrait alors élaborer une méthode permettant de configurer les yeux, les sourcils et la bouche en fonction des paramètres estimés de la localisation sur l'usager réel.

CHAPITRE 3

LE SUIVI DES POINTS 2D

3.1. Objectif du suivi

Soit une suite d'images 2D (I_0, I_1, ..., I_{t-1}, I_t) obtenues de la projection d'une scène 3D au cours du temps (à chaque 30 ms). Un suivi est utilisé dans le but de poursuivre, au cours du temps, un point p_i en 2D provenant de la projection d'un même point 3D de la scène. En utilisant un nombre suffisant de N points (p_0, p_1, ... p_{N-1}) obtenus de la projection d'un objet connu et en supposant l'objet rigide, les déplacements obtenus par le suivi permettront d'évaluer la position 3D de cet objet dans la scène. Dans ce projet, un suivi est donc utilisé pour évaluer la position de la tête d'un usager à tout instant.

Figure 3.1. Exemple du suivi de N points (p_0, p_1, ... p_{N-1}). Ces points proviennent de la projection d'un même objet au cours du temps.

La figure 3.1 illustre un exemple où le suivi est effectué sur 6 points sur le visage de l'usager. Bien que les positions des points soient différentes entre les 2 images, la position sur l'objet de la scène (la tête de l'usager) doit demeurer la même.

3.2. Les difficultés du suivi

Lors de l'initialisation du suivi, plusieurs points p_i (i = 0, 1, .., N-1) sont localisés sur la première image au temps t_0. Pendant le suivi entre l'image I_t et I_{t-1}, la relocalisation d'un point p_i consiste à retrouver un nouvel endroit, avec un faible déplacement, dans I_t dont le voisinage est très semblable à celui dans I_{t-1}. Cette correspondance se fait donc d'une image à l'autre. Au cours du temps, cette approche peut provoquer une erreur cumulative [84] causée par les petites erreurs de localisation générées tout au long de la séquence d'images. Afin d'éviter ce problème, l'information originale obtenue de la première image I_0 devra être utilisée afin de relocaliser p_i. Ceci signifie qu'il faudra comparer une région r_t sur I_t avec la région initiale r_0 sur I_0. Voici quelques contraintes à considérer avant d'effectuer une telle comparaison :

1. D'importantes translations ou rotations peuvent être obtenues entre r_0 et r_t ;
2. De faibles translations ou rotations sont obtenues entre r_{t-1} et r_t ;
3. Le contraste et la luminosité peuvent grandement varier entre r_o et r_t ;
4. Le contraste et la luminosité varient très peu entre r_{t-1} et r_t;
5. La région r_t doit être localisée très rapidement.

Les contraintes 2, 4 et 5 peuvent être respectées grâce à l'utilisation d'une procédure SSD "Sum of Squares Differences". Les contraintes 2, 3, 4 et même 5 peuvent être respectées grâce à une procédure NCC "Normalised Cross-Correlation". Cependant, les SSD et NCC ne peuvent pas respecter la première contrainte. Et si la NCC est effectuée sous plusieurs rotations, alors ce sera la contrainte 5 qui sera difficilement respectée. Une nouvelle approche a donc été élaborée en utilisant à la fois le SSD et le NCC pour respecter ces 5 contraintes.

Dans [36], le suivi est effectué à l'aide de la NCC entre r_t et r_0, ceci est robuste au changement de contraste et de luminosité mais pas aux rotations ou autres transformations affines (sauf la translation).

Dans [79][81], le suivi est effectué en supposant que des transformations affines peuvent être effectuées entre les régions r_t et r_0 et que ces transformations sont très faibles entre r_t et r_{t-1}. Une approche de minimisation par descente du gradient est donc utilisée pour trouver la meilleure transformation affine de r_0 pour obtenir r_t en tenant compte aussi des variations de contraste et de luminosité. Puisque cette transformation nécessite beaucoup de paramètres, des minimums erronés sont souvent atteints s'il existe d'assez grandes varitations entre r_t et r_{t-1}.

Dans [23], la robustesse du suivi est augmentée en tenant compte des divers éclairages possibles que peut rencontrer la région r_t. Il faut cependant initialiser r_0 sous divers éclairages auparavant avant d'utiliser une telle technique.

3.3. Fonctionnement général du suivi

Voici les principales étapes constituant le suivi d'un point p_i:

1. Recherche par la SSD : Entre I_{t-1} et I_t, rechercher r_t à partir de r_{t-1} en appliquant la SSD. Cette recherche est effectuée dans une région R_{SSD} et permettra d'avoir une position approximative de p_i, soit q_i ;
2. Recherche par la NCC : À partir de q_i, rechercher r_t à partir de r_0 en appliquant la NCC sous plusieurs orientations. Cela permettra de réduire l'erreur cumulative en évitant de seulement comparer deux images successives. De plus, la comparaison sera indépendante de l'intensité moyenne et la variance qui peuvent varier entre r_t et r_0. Cette recherche est effectuée dans une très petite région R_{NCC} autour de q_i et permettra d'obtenir la position de p_i .

La première étape permet d'obtenir l'estimation de p_i mais peut introduire une petite erreur de positionnement étant donné que le SSD ne se fait plus avec r_0 mais plutôt avec r_{t-1} pour I_t (où $t > 1$). Sur une longue séquence d'images, cette erreur risque donc de grandir et le suivi sera alors inefficace après seulement quelques secondes d'animation. La deuxième étape permet d'éliminer cette erreur cumulative en comparant de nouveau r_t à r_0 sous diverses orientations. Bien que la première étape soit assez rapide, une attention particulière doit être apportée à la deuxième afin d'éviter les traitements redondants. La dimension de R_{SSD} peut varier selon la vitesse maximale du déplacement de l'usager entre I_{t-1} et I_t mais la dimension de R_{NCC} reste fixe et peut n'être constituée que de pixels voisins de q_i par exemple.

3.3.1. Recherche par la SSD

En supposant que l'intensité moyenne ainsi que la variance, pour une même région, varient très peu entre deux images successives, la SSD est utilisée pour quantifier la ressemblance entre 2 images, ceci est démontré par l'équation 3.1 :

$$SSD\ (R(x_t, y_t, I_t), R(x_{t-1}, y_{t-1}, I_{t-1})) = \sum_{i=-L_1/2}^{i=L_1/2} \sum_{j=-L_2/2}^{j=L_2/2} \left(I_t(x_t + i, y_t + j) - I_{t-1}(x_{t-1} + i, y_{t-1} + j)\right)^2 \qquad (3.1)$$

où R(x,y,*I*) représente une région rectangulaire L_1 x L_2 centrée en *(x,y)* sur une image *I*. Il est préférable d'avoir une grande région lors du calcul de la SSD afin que de grandes portions de la figure soient comparées : si p_i se situe sur un oeil ouvert dans I_{t-1}, par exemple, et que cet oeil est fermé dans I_t, alors de grandes régions de l'image incluront aussi le sourcil et une partie du nez. La SSD a donc plus de chance de bien localiser p_i malgré l'occlusion engendrée par la paupière dans I_t. Cependant, l'utilisation d'une trop grande région est risquée lors d'un mouvement rapide de la tête si cette région inclut aussi l'arrière-plan qui est immobile. Étant donné que les points p_i seront surtout situés loin des limites de la tête pour être près des éléments du visage, une région R_{SSD} rectangulaire étendue sur l'horizontale permettra à la fois de couvrir une grande région tout en diminuant les probabilités d'inclure l'arrière-plan, ceci permet donc une meilleure robustesse aux petites occlusions. De plus, une grande région est préférable lorsque l'information de l'image n'est pas assez consistante : sur une surface lisse, sur une surface contenant d'autres petites régions très semblables, etc.

Une grande région engendre normalement plus de calculs. Une façon simple de contourner ce problème consiste à sous-échantillonner la région R_{SSD}. Ainsi, une région de $2 \cdot L1$ x $2 \cdot L2$ échantillonnée à chaque 2 pixels de façon horizontale et verticale prendra le même temps de calculs qu'une région $L1$ x $L2$ échantillonnée à chaque pixel. Il y a évidemment une perte d'information sur une région lorsqu'un tel sous-échantillonnage est appliqué mais cette perte n'est pas vraiment importante : le gain de rapidité l'est beaucoup plus. Et la deuxième étape du suivi permettra d'ailleurs de corriger l'erreur de positionnement engendrée.

La région de recherche de la valeur minimale de la SSD

La SSD est calculée entre I_t et I_{t-1} autour de la position approximisée de q_i dans I_t. Si l'on suppose qu'entre l'instant t et t-1, l'accélération des déplacements des points est nulle, alors l'approximation initiale de la position (x_t,y_t) de q_i dans I_t est obtenue par l'équation 3.2:

$$\begin{aligned} x_t &= x_{t-1} + \Delta x = x_{t-1} + (x_{t-1} - x_{t-2}) = 2 \cdot x_{t-1} - x_{t-2} \\ y_t &= y_{t-1} + \Delta y = y_{t-1} + (y_{t-1} - y_{t-2}) = 2 \cdot y_{t-1} - y_{t-2} \end{aligned} \tag{3.2}$$

Lorsque la position approximative (x_t,y_t) est calculée, il faut ensuite vérifier dans un voisinage l'endroit où la valeur de la SSD est minimale. La figure 3.2 illustre différentes façons de faire la recherche dans le voisinage. La méthode classique consiste à définir une région rectangulaire de recherche W_1xW_2 centrée en (x_t,y_t). Tous les pixels de cette région sont parcourus et le point où la SSD minimale est atteinte constitue la position recherchée de q_i. La recherche s'effectue donc en parcourant une suite de lignes

(ou colonne) S1, S2, S3, etc. Cette méthode est cependant très lente car toute la région est toujours parcourue, même si le minimum est très près de la position préalablement estimée.

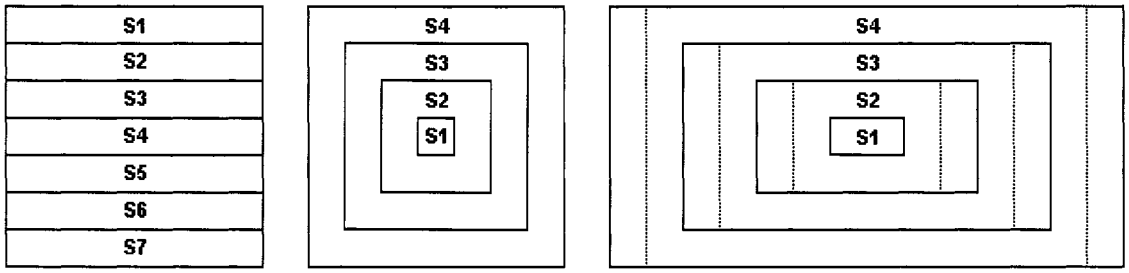


Figure 3.2. Diverses façons de parcourir la fenêtre de recherche R_{SSD}. En a), Une recherche classique ligne par ligne (ou colonne par colonne). En b), Une recherche utilisant une suite de carrés s'éloignant du centre. En c), Une recherche semblabe à b mais où les directions horizontales sont favorisées.

La méthode de recherche utilisée dans ce projet consiste à parcourir une suite de carrés s'éloignant du centre tel que sur la figure 3.2.b). Ces carrés sont vides et d'une bordure d'un pixel. Lorsqu'une nouvelle valeur minimum de la SSD est trouvée, qui est plus basse qu'un certain seuil λ, la recherche se poursuit pour un autre carré afin de vérifier s'il n'y a pas un autre minimum très près. Si aucune autre valeur plus basse n'est trouvée, alors la SSD minimale est considérée comme étant trouvée, évitant ainsi de parcourir le reste de la région. Une autre méthode consiste à avantager la recherche horizontale puisque les mouvements horizontaux semblent les plus fréquents pour la tête. Ainsi, au lieu d'utiliser une suite de carrés, des rectangles (2 pixels de bordure horizontale et 1 pixel de bordure verticale) pourraient permettre de trouver la SSD minimale plus rapidement. D'autres méthodes plus complexes, comme une suite

d'ellipses inclinées, permettraient sûrement de diminuer encore la recherche mais les calculs deviendraient beaucoup plus complexes, l'utilisation de carrés ou rectangles permet d'avoir des calculs très simples et c'est pourquoi ils sont utilisés dans ce projet. Une minimisation utilisant une descente du gradient pourrait aussi être utilisée mais cette dernière ne garantit pas de converger vers le bon minimum.

Le seuil λ est fixé sur la base de la valeur moyenne de la SSD due au bruit entre les images : même si rien ne bouge dans la scène 3D, la SSD calculée à partir des 2 images consécutives ne sera pas nulle à cause du bruit. En supposant que de petits changements d'intensité soient possibles ainsi que de très petites transformations (rotation par exemple), λ peut être quelque peu augmenté. L'équation 3.3 montre donc comment calculer λ :

$$\lambda = L1 \cdot L2 \cdot \sigma_I^2 \tag{3.3}$$

où σ_I signifie la déviation standard, que peut avoir l'intensité d'un pixel, causée à la fois par le bruit dans l'image ainsi que par les très petites transformations dans la scène 3D. Ainsi un minimum de la SSD, parmi plusieurs valeurs inférieures à λ, risque d'être instable car ce minimum peut n'être causé que par un certain état du bruit à cet endroit. Donc pour toutes les positions dont les valeurs de la SSD sont inférieures au seuil λ, la meilleure position est difficilement justifiable et c'est pourquoi une valeur de λ atteinte est jugée acceptable pour arrêter la recherche. Cette nouvelle méthode de recherche est

beaucoup plus rapide que la méthode classique où le seul critère d'arrêt est de parcourir toute la région pour en trouver le minimum. Il est tout de même préférable de chercher une valeur de la SSD qui est la plus basse possible. Donc lorsqu'une valeur inférieure au seuil λ est atteintte, la recherche se poursuit sur les pixels voisins jusqu'à ce qu'un prochain minimum (qu'il soit local ou global) soit trouvé. Pour de bons résultats dans ce projet, les valeurs de *L1*, *L2* et σ^2_I ont été fixées à 15, 15 et 10 respectivement. Ceci permet d'obtenir le seuil $\lambda = 2250$.

3.3.2. Recherche par la NCC

Cette recherche est utilisée pour comparer une région d'image entre I_t et I_0. Puisque de grands mouvements peuvent être effectués par l'usager entre ces images, l'intensité et le contraste peuvent varier pour un même endroit sur le visage. Ces variations peuvent être causées par l'ombrage ou les reflets lumineux et ce, même si l'éclairage demeure constant dans la scène. La NCC semble donc plus adéquate que la SSD pour la comparaison entre I_t et I_0 car elle est justement conçue pour être robuste à ces variations d'intensité, ceci n'est pas le cas de la SSD. Donc en supposant ces variations et que certaines transformations soient négligeables (pas d'étirement ou d'homothétie entre I_t et I_0), la NCC est utilisée pour quantifier la ressemblance de régions entre ces 2 images. Soit R une région d x d d'une image I centrée en (x, y), l'équation 3.4 illustre comment obtenir la NCC entre une région R_t sur I_t et une région R_0 sur I_0 :

$$NCC(R_t, R_0) = \frac{1}{d^2 \sigma_t \sigma_0} \sum_{i=-d/2}^{i=d/2} \sum_{j=-d/2}^{j=d/2} \left(I_t(x_t + i, y_t + j) - \overline{R_t}\right)\left(I_0(x_0 + i, y_0 + j) - \overline{R_0}\right) \quad (3.4)$$

où

$$\overline{R_t} = \frac{1}{d^2} \sum_{i=-d/2}^{i=d/2} \sum_{j=-d/2}^{j=d/2} I_t(x_t + i, y_t + j)$$

$$\overline{R_0} = \frac{1}{d^2} \sum_{i=-d/2}^{i=d/2} \sum_{j=-d/2}^{j=d/2} I_0(x_0 + i, y_0 + j)$$

$$\sigma_t = \frac{1}{d} \sqrt{\sum_{i=-d/2}^{i=d/2} \sum_{j=-d/2}^{j=d/2} \left(I_t(x_t + i, y_t + j) - \overline{R_t}\right)^2}$$

$$\sigma_0 = \frac{1}{d} \sqrt{\sum_{i=-d/2}^{i=d/2} \sum_{j=-d/2}^{j=d/2} \left(I_0(x_0 + i, y_0 + j) - \overline{R_0}\right)^2}$$

La comparaison entre I_t et I_0 est effectuée ici afin d'éliminer l'effet de l'erreur cumulative. À partir de la position (x_t, y_t) de q_i obtenue précédemment, un très petit voisinage R_{NCC} centré en q_i est parcouru. Ce voisinage R_{NCC} peut n'être constitué que du centre et d'un seul pixel dans chaque direction, soit les positions (x_t, y_t), (x_t-1, y_t), (x_t+1, y_t), (x_t, y_t-1) et (x_t, y_t+1). La recherche de la vraie position de p_i est illustrée par la figure 3.3 :

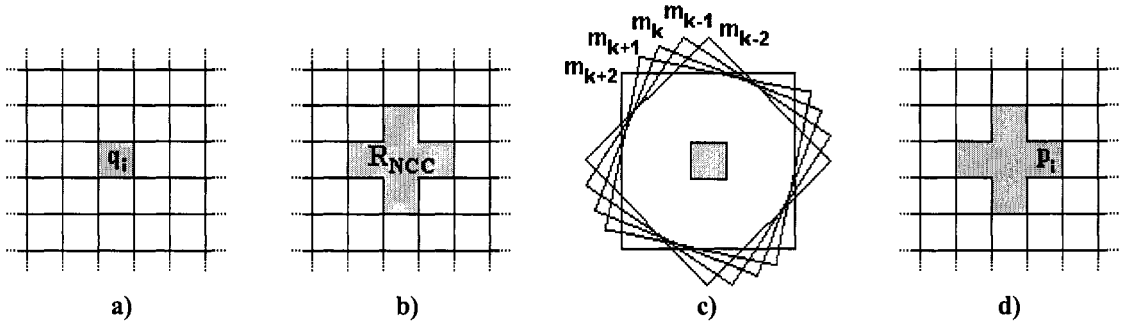


Figure 3.3. Recherche de la position de p_i. En a), la position de q_i est obtenue par la valeur minimale du SSD. En b), un très petit voisinage R_{NCC} centré en q_i sera utilisé pour la NCC. En c), pour chaque pixel de R_{NCC} la NCC est effectuée entre une région carrée centrée en ce pixel et quelques orientations de la région provenant de I_0. En d), le point p_i est finalement localisé là où le maximum de la NCC a été obtenu.

La NCC sera effectuée sur plusieurs positions et orientations. Ces orientations seront centrées selon la dernière orientation θ obtenue sur I_{t-1} et couvriront un total de $2n+1$ angles, où n indique le nombre d'angles parcourus sur un des deux côtés de θ. Pour chaque pixel du voisinage R_{NCC}, la NCC est donc effectuée entre une région carrée d x d centrée en ce pixel et quelques modèles m_k où i = $\theta - n$, $\theta - n + 1$, ..., $\theta - 1$, θ, $\theta + 1$, ..., $\theta + n - 1$, $\theta + n$ qui devront être générés à partir de I_0. Il est à noter que ces modèles sont aussi carrés et ne sont pas inclinés comme sur la figure 3.4.c) : seules les images à l'intérieur le sont. En supposant que R_{NCC} contient v pixels, il y a donc $v \cdot (2 \cdot n+1)$ fois la NCC à effectuer afin de localiser p_i dans I_t là où la valeur maximale de la NCC a été atteinte.

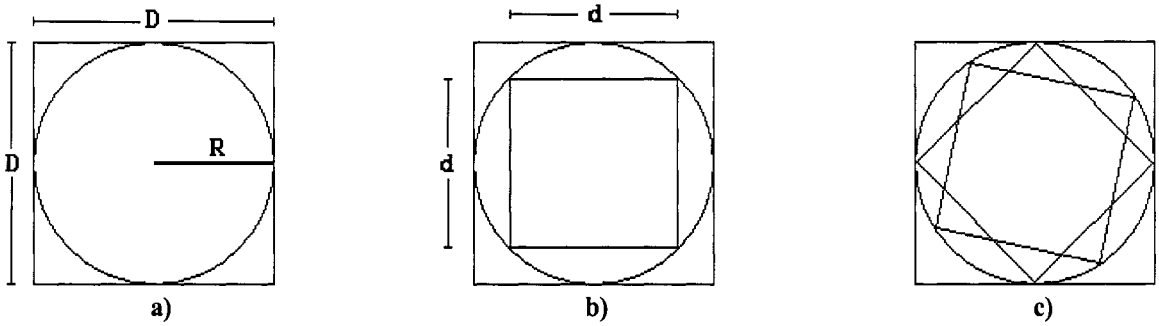


Figure 3.4. Géométrie utilisée pour obtenir un modèle m_k. En a), un carré D x D inclut le cercle de rayon R. En b), le cercle inclut le deuxième carré d x d. En c), plusieurs rotations sont appliquées au deuxième carré afin d'avoir l'information de la région de l'image sous plusieurs orientations.

Puisque la NCC doit être effectuée plusieurs fois pour positionner un point p_i sur I_t, il est avantageux de définir une méthode permettant de diminuer la redondance des calculs afin d'effectuer le suivi plus rapidement. Lorsque p_i est localisé sur I_0 initialement, une région carrée de dimensions D x D de l'image, centrée en p_i, est conservée en mémoire. Ce carré inclut un cercle de rayon R = D/2 et ce dernier inclut un autre carré de dimensions d x d (où $d = R\sqrt{2} = D\sqrt{2}/2$). La figure 3.4.b) illustre le cercle et les 2 carrés. Un nombre M de rotations est appliqué au deuxième carré et, pour chaque rotation, la région de l'image incluse à l'intérieur est retenue en mémoire comme étant un modèle m_k (où k = 0, 1, ..., M-1) de la région r_0: ceci afin de retenir l'information originale du voisinage de p_i sur I_0 sous M orientations possibles. En supposant que la projection en 2D d'une scène 3D peut effectuer un intervalle d'orientation entre θ_{min} et θ_{max}, comme illustré par la figure 3.5, alors le pas angulaire entre chaque modèle sera de $(\theta_{min} - \theta_{max})/M$. Plus M est grand, plus grande sera la

précision mais plus grands seront aussi l'espace mémoire réservé et le temps de calcul pour générer un intervalle d'orientation donnée. La figure 3.6 illustre un exemple des M régions retenues à partir d'un point initial p_i sur I_0. Puisque les M modèles seront utilisés à la deuxième étape du suivi, c'est-à-dire le réajustement de la position du point p_i à l'aide de la NCC, le pas angulaire entre 2 modèles consécutifs doit être assez petit car la corrélation est très peu robuste aux rotations. Bien que tous les modèles puissent être générés dès la première étape, ceux-ci ne seront en fait générés que lorsqu'ils seront nécessaires.

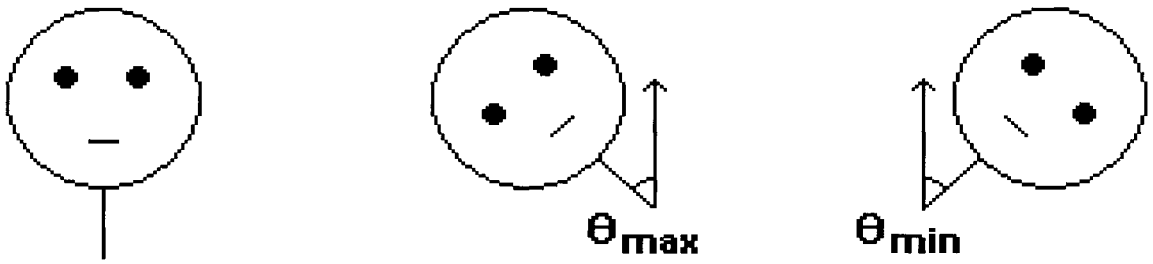


Figure 3.5. Limites de l'orientation de la tête. θ_{max} et θ_{min} représentent les limites de l'orientation que peut avoir la projection en 2D de la scène 3D.

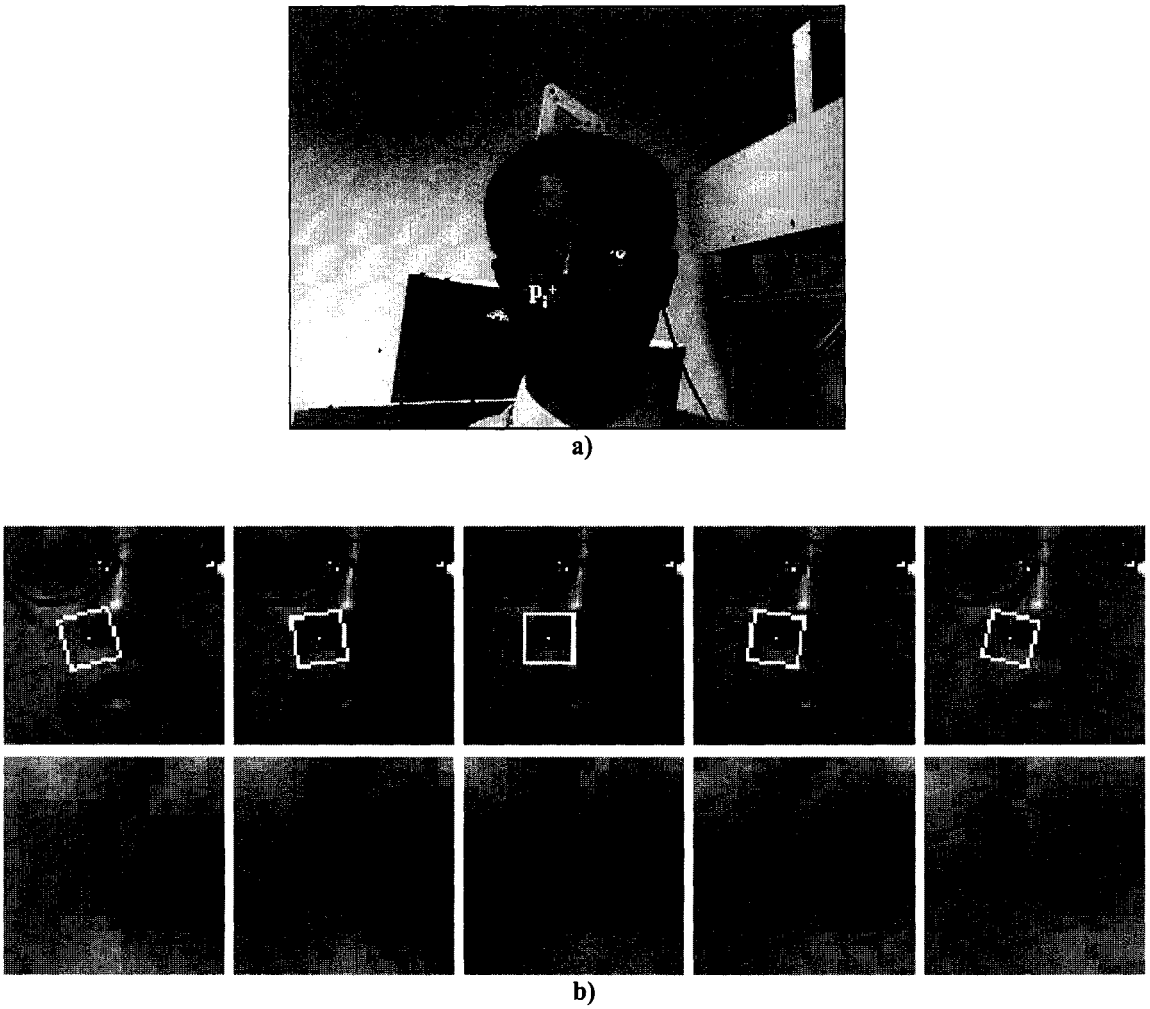


Figure 3.6. Exemple des M régions retenues à partir d'un point initial p_i sur I_0. En a), le point p_i sur l'image I_0. En b) les modèles m_k obtenus en changeant l'orientation de la région sur I_0.

3.4. Calculs pour la NCC utilisant les modèles

Puique les modèles sont récupérés sur I_0, l'utilisation de la NCC demeure justifiée. Par rapport à une autre méthode, comme la SSD par exemple, la NCC est utilisée pour sa robustesse aux variations d'intensités mentionnées précédemment. Soit R_t une région d x d de l'image I_t centrée en (x_t, y_t) et soit m_k un modèle à la k^e orientation

retiré de l'image initiale I_0, l'équation 3.4 est modifiée pour obtenir la NCC en utilisant ce modèle, ceci est montré avec l'équation 3.5. :

$$NCC(R_t, m_k) = \frac{1}{d^2 \sigma_t \sigma_k} \sum_{i=-d/2}^{i=d/2} \sum_{j=-d/2}^{j=d/2} \left(I_t(x_t + i, y_t + j) - \overline{R_t}\right)\left(m_k(i,j) - \overline{m_k}\right) \tag{3.5}$$

où

$$\overline{R_t} = \frac{1}{d^2} \sum_{i=-d/2}^{i=d/2} \sum_{j=-d/2}^{j=d/2} I_t(x_t + i, y_t + j)$$

$$\overline{m_k} = \frac{1}{d^2} \sum_{i=-d/2}^{i=d/2} \sum_{j=-d/2}^{j=d/2} m_k(i,j)$$

$$\sigma_t = \frac{1}{d} \sqrt{\sum_{i=-d/2}^{i=d/2} \sum_{j=-d/2}^{j=d/2} \left(I_t(x_t + i, y_t + j) - \overline{R_t}\right)^2}$$

$$\sigma_k = \frac{1}{d} \sqrt{\sum_{i=-d/2}^{i=d/2} \sum_{j=-d/2}^{j=d/2} \left(m_k(i,j) - \overline{m_k}\right)^2}$$

3.4.1. La création des modèles

Lorsqu'un modèle m_k est créé (selon une orientation de la région sur I_0), plusieurs coordonnées doivent être calculées afin de tenir compte de la rotation (une coordonnée par pixel). Ces coordonnées sont relatives au centre de la région et il est possible d'effectuer plusieurs prétraitements avant même d'avoir l'image I_0. Ainsi, pour une suite de M orientations données d'une région sur I_0, tous les modèles m_k peuvent bénéficier d'un prétraitement avant même le début du suivi, ceci permettra par la suite d'économiser beaucoup du temps de calcul.

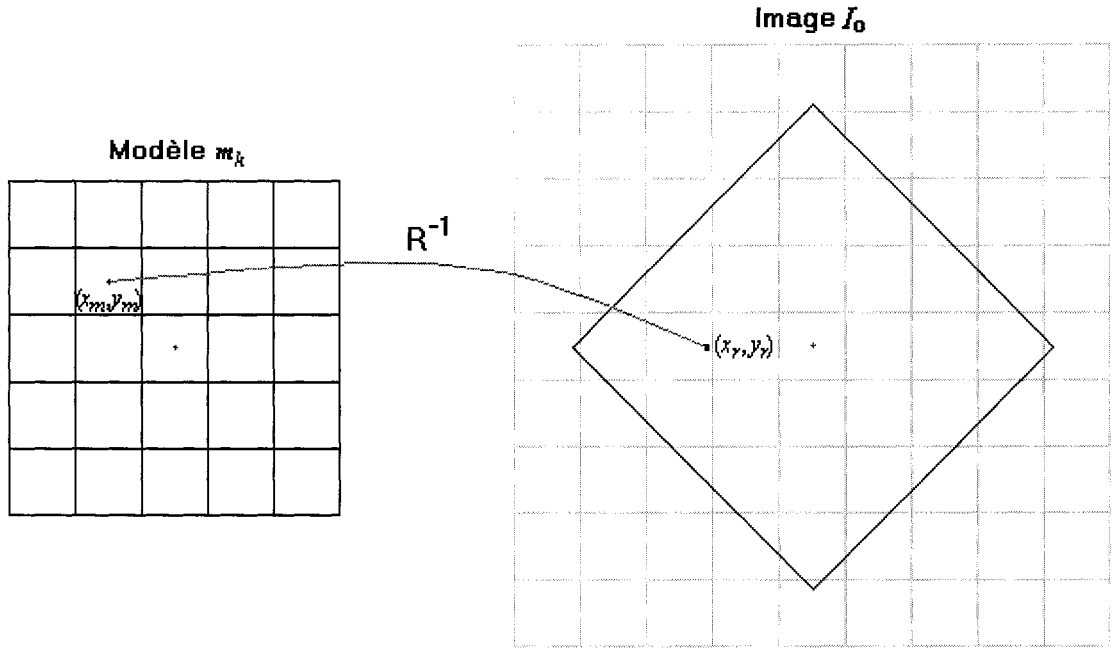


Figure 3.7. Calculer la valeur d'un pixel à une certaine orientation. À gauche, un pixel dans le modèle m_k à une position (x_m,y_m). À droite, une position (x_r,y_r) dans l'image I_0 est la correspondance de (x_m,y_m) après la transformation R^{-1}. Sur cet exemple, $\theta = 45°$.

La figure 3.7 illustre comment les coordonnées sont calculées pour un pixel de m_k à une orientation θ. Ainsi, une position (x_r, y_r) est obtenue à partir de (x_m, y_m) et de θ et (x_m, y_m) comme le montre l'équation 3.6.

$$\begin{bmatrix} x_r \\ y_r \end{bmatrix} = R^{-1} \cdot \begin{bmatrix} x_m \\ y_m \end{bmatrix} \tag{3.6}$$

où $$R^{-1} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

L'intensité de la position intermédiaire (x_r, y_r) peut être estimée à l'aide des intensités des 4 pixels voisins et d'une interpolation bilinaire.

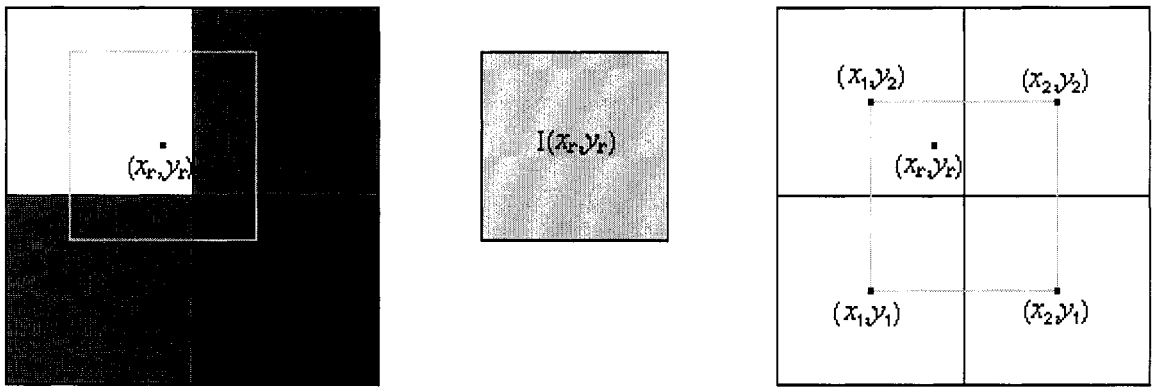


Figure 3.8. Obtenir l'intensité désirée d'un pixel en fonction des voisins. En a), la position (x_r,y_r) sur I_0. En b), l'intensité désirée pour la position (x_r,y_r). En c), les positions entourant (x_r,y_r) qui serviront à calculer $I(x_r,y_r)$.

La figure 3.8 illustre l'intensité désirée $I(x_r, y_r)$ à l'aide des pixels voisins aux positions (x_1, y_1), (x_1, y_2), (x_2, y_1) et (x_2, y_2). L'équation 3.7 montre comment l'interpolation bilinéaire est obtenue.

$$I(x_r,y_r) = I(x_1,y_1) + \Delta x\cdot[I(x_2,y_1) - I(x_1,y_1)] + \Delta y\cdot[I(x_1,y_2) - I(x_1,y_1)] + \Delta x\cdot\Delta y\cdot[I(x_1,y_1) + I(x_2,y_2) - I(x_2,y_1) - I(x_1,y_2)] \quad (3.7)$$

où $\Delta x = x_r - x_1$

$\Delta y = y_r - y_1$

En supposant qu'il y ait M orientations différentes pour un modèle et que ce dernier contienne d x d pixels, le prétraitement consistera donc à calculer $M\cdot d^2\cdot 4$ valeurs. En connaissant chaque orientation et les dimensions de m_k, il est donc possible de récupérer les valeurs désirées à partir d'une mémoire statique.

3.4.2. Utilisation de la mémoire statique

Une mémoire statique, nommée Mem1, est utilisée afin de simplifier les traitements qui devront être effectués très rapidement. Cette mémoire contient les informations nécessaires pour calculer un pixel d'un modèle m_k selon son orientation. En supposant qu'il y ait M orientations possibles et que toutes les coordonnées (x_r, y_r) utilisées pour créer les M modèles se situent dans un cercle de rayon D, alors la dimension de la mémoire nécessaire pour le suivi d'un point p_i sera de D·D·M. Et chaque élément de cette mémoire devra contenir la série suivante :

$$Mem1(x_m, y_m, i) = [x_1, y_1, x_2, y_2, \Delta x, \Delta y] \tag{3.8}$$

où

$$\begin{aligned} x1 &= \lfloor x_r \rfloor \\ y1 &= \lfloor y_r \rfloor \\ x2 &= \lceil x_r \rceil \\ y2 &= \lceil y_r \rceil \\ \Delta x &= x_2 - x_1 \\ \Delta y &= y_2 - y_1 \end{aligned}$$

D'après l'équation 3.8, il n'est pas nécessaire d'avoir I_0 pour remplir la mémoire statique : il suffit de connaître D et les M orientations possibles.

3.4.3. Utilisation de la mémoire des images

Aussitôt que la première image I_0 est connue, une deuxième mémoire, nommée Mem2, est remplie pour éviter les calculs redondant sur les images suivantes lors du calcul de la NCC. Pour déterminer ce que cette mémoire doit contenir, l'équation 3.5 doit être écrite autrement afin d'obtenir l'équation 3.9 :

$$NCC(R_I, R_k) = \frac{1}{\phi_t \phi_k} \sum_{i=-d/2}^{i=d/2} \sum_{j=-d/2}^{j=d/2} R_I(i,j) \cdot R_k(i,j) \tag{3.9}$$

Où

$$R_I(i,j) = I_t(x_t + i, y_t + j) - \overline{R_t}$$
$$R_k(i,j) = m_k(i,j) - \overline{m_k}$$
$$\phi_t = \sigma_t$$
$$\phi_k = d^2 \sigma_k$$

L'équation 3.9 est utilisée dans le projet afin d'accélérer les traitements de la NCC. Lorsqu'un modèle m_k est généré, une valeur $R_k(i, j)$ est aussi calculée et mise en mémoire pour chaque pixel de m_k, cette dernière contient la différence entre un pixel de m_k et l'intensité moyenne de m_k. Lorsqu'un pixel du voisinage R_{NCC} est parcouru, une valeur R_I est créée et celle-ci contient, pour chaque pixel, la différence entre l'intensité du pixel de I_t et l'intensité moyenne de la région d x d centrée sur le pixel concerné de R_{NCC}.

Lorsque la NCC est appliquée sous plusieurs orientations et sur les v pixels de R_{NCC}, alors R_I et ϕ_t ne sont calculés que v fois. Soit M le nombre de modèles utilisés et g le nombre de ces modèles n'ayant pas été déjà générés lors des images précédentes. Pour chaque pixel de R_{NCC}, R_k et ϕ_k ne sont calculés que g fois car les (M-g) autres R_k et ϕ_k sont déjà disponibles en mémoire. Le but de toutes ces manipulations est de diminuer les opérations qui seront les plus fréquentes afin d'avoir un fonctionnement très rapide pour le temps réel : l'équation du NCC est répétée $v \cdot M$ fois pour localiser p_i. Et en utilisant l'équation 3.5, 4 additions et une multiplication doivent être effectuées

$d^2 \cdot v \cdot M$ fois à l'intérieur des sommateurs. En utilisant l'équation 3.9, le terme $R_k(i, j)$ ne peut être calculé qu'une seule fois et être inséré en mémoire puisqu'il est redondant (il ne dépend que de I_0 et non de I_t). Ainsi, le calcul d'au moins une moyenne et une déviation standard sera évité à l'intérieur des sommations. Donc au moins $2 \cdot d^2 \cdot v$ additions, $d^2 \cdot v$ multiplications et v racines carrées seront évitées lors de la recherche de p_i. Voici donc ce que *Mem*2 doit contenir pour chaque oritentation :

$$Mem2(k) = [R_k \quad \phi_k] \tag{3.10}$$

où R_k est l'ensemble des $R_k(i, j)$ où $i, j \in$ [-d/2,d/2]

Voici donc un résumé des étapes à franchir pour faire le suivi d'un point p_i à l'aide des mémoires *Mem*1 et *Mem*2 :

1. Avant le début du suivi, remplir *Mem*1 selon l'équation 3.8 ;
2. À partir de la position de p_i sur I_0, remplir *Mem*2 selon l'équation 3.10 ;
3. Pendant le suivi :
   - En utilisant la SSD, chercher le point q_i sur I_t à l'aide d'un voisinage R_{SDD} autour du point p_i sur I_{t-1} ;
   - En utilisant la NCC de l'équation 3.9, chercher le point p_i sur I_t à l'aide d'un voisinage R_{NCC} autour du point q_i. La recherche est effectuée sous plusieurs orientations et les modèles m_k inexistants seront créés à l'aide de *Mem*1. Les valeurs nécessaires pour la NCC seront récupérées dans *Mem*2.

3.4.4. Démonstration des modèles utilisés

La figure 3.9 illustre quelques exemples des modèles utilisés pour effectuer le suivi. Les positions du point p_i sont situées sur la colonne de gauche. Vis-à-vis de ces positions, une région carrée du voisinage est recueillie et celle-ci est affichée sur la colonne de droite. Il est à noter que cette région n'est pas inclinée. La colonne du centre représente les modèles utilisés pour maximiser la ressemblance avec les régions de la colonne de droite. Les carrés blancs sur la colonne de gauche sont centrés par rapport aux positions de p_i et l'inclinaison de ces carrés représente la meilleure orientation détectée de l'image initiale pour maximiser la ressemblance. Sur 3.9.a), les images sur les colonnes droite et au centre sont les mêmes puisqu'il s'agit de la première image I_0 et que c'est là que les modèles m_i y sont construits. Sur 3.9.b), un oeil fermé sur I_t a dû être comparé à un oeil ouvert puisqu'il n'y a aucun modèle d'oeil fermé. L'algorithme a tout de même choisi un bon modèle grâce à la ressemblance autour de l'oeil qui était conservée. En 3.9.c) et 3.9.d), les yeux sont toujours ouverts mais avec différentes conditions d'éclairage. Là encore, de bons modèles ont été choisis et la ressemblance en est maximisée. Il est à noter que l'orientation (ou le modèle) détectée n'est pas très importante en autant que la position soit précise. Mais dans le cas où l'orientation détectée serait très mauvaise, les chances d'obtenir la bonne position sont très faibles. Heureusement, la SSD effectuée avant la NCC permet d'estimer la position afin d'éviter de grandes erreurs de localisation lors de l'utilisation des modèles.

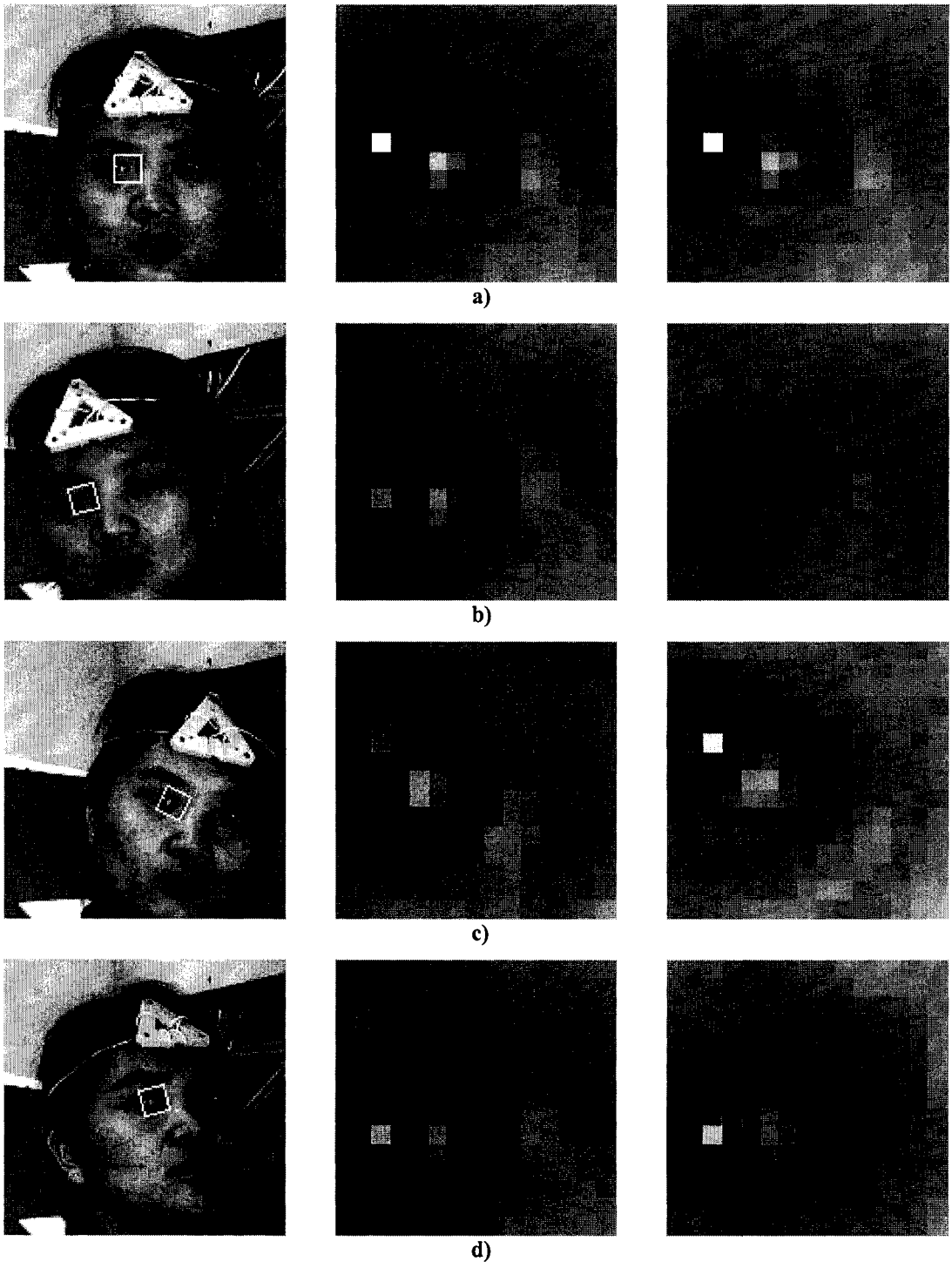


Figure 3.9. Utilisation des modèles pour le suivi. À gauche, les images I_t. Au centre, les modèles m_k. À droite, les régions retrouvées sur I_t.

3.5. Analyse des résultats obtenus

L'analyse des résultats du suivi a été effectuée à l'aide de plusieurs points sur quelques séquences d'images. La précision de la position des points est obtenue et une comparaison est effectuée entre les résultats de la méthode proposée et celle utilisée par [84] qui n'utilisait que l'étape de le SSD. Les détails de cette analyse sont présentés en Annexe I.

3.6. Améliorations suggérées

Le suivi implanté est robuste aux rotations autour de l'axe Z (de profondeur) mais pas autour des axes X et Y (l'horizontale et la verticale respectivement). La méthode pourrait être modifiée pour tolérer ces rotations mais les traitements deviendraient beaucoup plus lourds. Des études plus approfondies pourraient être effectuées afin d'améliorer la robustesse à ces rotations tout en utilisant peu de traitement. De plus, le suivi est peu robuste à l'homothétie. Idéalement, le suivi devrait être robuste aux transformations affines (translation, rotation, homothétie, étirement, etc) afin de tolérer de plus grandes possibiltés de projections de la scène 3D. Certaines méthodes utilisent un suivi robuste aux transformations affines mais seulement pour de très faibles variations entre les images. Des méthodes itératives, comme la descente du gradient, sont utilisées pour maximiser une fonction de ressemblance basée sur la NCC. Pour de grandes variations, la fonction risque donc de converger vers un maximum local aulieu du global. Une méthode hybride entre cette dernière et la méthode implantée dans ce projet pourrait peut-être donner de meilleurs résultats.

CHAPITRE 4

ESTIMATION DU MOUVEMENT RIGIDE 3D

4.1. Objectif de l'estimation

Une estimée des mouvements rigides de la tête de l'usager doit être obtenue afin de positionner le modèle virtuel à tout instant au cours de la séquence d'images. L'estimation consiste à recueillir les trois translations et les trois rotations selon les axes X, Y et Z, c'est-à-dire les axes horizontal, vertical et de la profondeur. Dans la méthode élaborée par [84], il a été constaté que de grandes imprécisions pouvaient être obtenues lorsque quelques mesures utilisées pour l'estimation étaient erronées, ceci ne permettait d'estimer le mouvement que sur une courte séquence d'images. Dans le présent chapitre, une méthode est proposée afin d'apporter une meilleure stabilité sur une longue séquence. La figure 4.1 illustre un exemple du mouvement rigide recherché.

a) b)

Figure 4.1. Exemple du mouvement rigide recherché. En a), la pose initiale de l'usager. En b), le mouvement rigide de l'usager relativement à la pose initiale.

4.2. Les difficultés de l'estimation

La méthode se base sur le suivi de points 2D sur au moins 3 positions différentes sur le visage de l'usager. Lorsque ces positions 2D ont une correspondance 3D unique sur le modèle virtuel, l'estimation du mouvement 3D peut être effectuée. En plus d'avoir des positions imprécises lors du suivi, le modèle virtuel utilisé est lui aussi peu précis, ceci peut être très néfaste pour l'estimation du mouvement. Bien que ces imprécisions puissent être faibles entre deux images successives, des estimations complètement fausses peuvent survenir après seulement quelques secondes, étant donné le nombre d'images utilisé.

4.3. Fonctionnement général de l'estimation

Pour l'estimation des paramètres du mouvement entre deux images consécutives I_t et I_{t-1}, le visage de l'usager peut être perçu approximativement comme un objet rigide par endroits. Le mouvement d'un tel objet a six degrés de liberté et peut être décrit à l'aide de six paramètres, soit la rotation et la translation par rapport aux axes X, Y et Z. Ceci est décrit par l'équation 4.1.

$$p_t = R \cdot p_{t-1} + T \tag{4.1}$$

où p_t : vecteur de la position $(x_t, y_t, z_t)^T$ du point au temps t

T : vecteur de la translation $(t_x, t_y, t_z)^T$ du point entre les temps t-1 et t

R : matrice de rotation définie par l'équation 4.2

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \tag{4.2}$$

où $r_{11} = \cos\theta_z \cdot \cos\theta_y$

$r_{12} = \sin\theta_z \cdot \cos\theta_y$

$r_{13} = -\sin\theta_y$

$r_{21} = -\sin\theta_z \cdot \cos\theta_x + \cos\theta_z \cdot \sin\theta_y \cdot \sin\theta_x$

$r_{22} = \cos\theta_z \cdot \cos\theta_x + \sin\theta_z \cdot \sin\theta_y \cdot \sin\theta_x$

$r_{23} = \cos\theta_y \cdot \sin\theta_x$

$r_{31} = \sin\theta_z \cdot \sin\theta_x + \cos\theta_z \cdot \sin\theta_y \cdot \cos\theta_x$

$r_{32} = -\cos\theta_z \cdot \sin\theta_x + \sin\theta_z \cdot \sin\theta_y \cdot \cos\theta_x$

$r_{33} = \cos\theta_y \cdot \cos\theta_x$

θ_x : angle de rotation autour de l'axe X entre les temps t et t-1

θ_y : angle de rotation autour de l'axe Y entre les temps t et t-1

θ_z : angle de rotation autour de l'axe Z entre les temps t et t-1

Une projection orthographique de la scène est supposée afin de simplifier les traitements. Cela est parfois utilisé lorsque que la profondeur du visage est beaucoup plus petite que la distance entre la tête et la caméra [12][22][84]. Cela engendre cependant des imprécisions sur le mouvement rigide car la translation en profondeur ne pourra jamais être estimée. De plus, des imprécisions peuvent s'introduire dans les autres paramètres du mouvement puisque la ressemblance diminue entre les points 2D

du suivi et la projection de leur correspondance 3D. Par exemple, si la tête s'approche trop de la caméra, la projection des points 3D du modèle virtuel ne chevauchera plus les points 2D du suivi car le visage complet aura grandi mais pas le modèle virtuel. La projection orthographique permet de simplifier les équations 4.1 et 4.2 et d'obtenir l'équation 4.3.

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{pmatrix} \cdot \begin{pmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \tag{4.3}$$

Dans ce cas, les coordonnées (x_t, y_t) du point 3D p_t sont égales aux coordonnées 2D dans l'image au temps t.

Selon [22], si de faibles mouvements ont lieu entre les images I_t et I_{t-1}, des simplifications peuvent être effectuées par l'équation 4.4.

$$\begin{aligned} &sin\ \theta \approx \theta \\ &cos\ \theta \approx 1 \\ &\theta_y \gg \theta_x \cdot \theta_z \end{aligned} \tag{4.4}$$

Ceci permet de simplifier l'équation 4.3 pour obtenir l'équation 4.5.

$$\begin{pmatrix} x_t - x_{t-1} \\ y_t - y_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & -z_{t-1} & y_{t-1} \\ 0 & 1 & z_{t-1} & 0 & -x_{t-1} \end{pmatrix} \cdot \begin{pmatrix} t_x \\ t_y \\ \theta_x \\ \theta_y \\ \theta_z \end{pmatrix} \tag{4.5}$$

Afin de récupérer les paramètres du mouvement $(t_x, t_y, \theta_x, \theta_y, \theta_z)$, les valeurs (x_t, y_t) sont récupérées à l'aide du suivi de points 2D et les valeurs $(x_{t-1}, y_{t-1}, z_{t-1})$ à l'aide du mouvement au temps t-1 du modèle virtuel de l'usager. Malheureusement, t_z n'est pas

disponible et les translations selon l'axe Z doivent être faibles de la part de l'usager. Puisqu'il y a 5 inconnues, le suivi d'un seul point est insuffisant. Des ajouts sont donc faits aux deux matrices à gauche de l'équation 4.5 afin d'inclure au moins 3 points au lieu d'un seul, ceci sera suffisant pour trouver les paramètres du mouvement. L'équation 4.6 illustre comment faire ces ajouts.

$$\begin{pmatrix} x_t^0 - x_{t-1}^0 \\ y_t^0 - y_{t-1}^0 \\ x_t^1 - x_{t-1}^1 \\ y_t^1 - y_{t-1}^1 \\ \vdots \\ \vdots \\ x_t^{n-1} - x_{t-1}^{n-1} \\ y_t^{n-1} - y_{t-1}^{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & -z_{t-1}^0 & y_{t-1}^0 \\ 0 & 1 & z_{t-1}^0 & 0 & -x_{t-1}^0 \\ 1 & 0 & 0 & -z_{t-1}^1 & y_{t-1}^1 \\ 0 & 1 & z_{t-1}^1 & 0 & -x_{t-1}^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & -z_{t-1}^{n-1} & y_{t-1}^{n-1} \\ 0 & 1 & z_{t-1}^{n-1} & 0 & -x_{t-1}^{n-1} \end{pmatrix} \cdot \begin{pmatrix} t_x \\ t_y \\ \theta_x \\ \theta_y \\ \theta_z \end{pmatrix} \tag{4.6}$$

Pour chaque variable, l'indice du haut indique le numéro de point tandis que l'indice du bas indique le temps. Pour un nombre total de N points disponibles, c'est-à-dire ceux imposés pour le suivi, un sous-ensemble de n points peut aussi être utilisé, en autant que $n \geq 3$. En attribuant des noms aux matrices, l'équation 4.6 peut s'écrire par l'équation 4.7 et les paramètres du mouvement M sont obtenus en appliquant l'équation 4.7.

$$T_n = A_n \cdot M \tag{4.6}$$

$$M = \left(A_n{}^T \cdot A_n\right)^{-1} \cdot A_n{}^T \cdot T_n \tag{4.7}$$

Bien que 3 points soient suffisants pour obtenir les paramètres du mouvement, il est préférable d'en utiliser davantage pour compenser les imprécisions. Si, par exemple, seulement 3 points imprécis sont utilisés, alors le mouvement estimé risque d'avoir

d'étranges paramètres pour correspondre parfaitement à ces 3 points. Au contraire, si davantage de points sont utilisés, alors le mouvement estimé aura tendance à correspondre approximativement à l'ensemble des points, ceci permet plus de robustesse.

4.4. La méthode implantée

Puisque certains points du suivi peuvent diverger sur une longue séquence d'images, une modification a été apportée pour tenir compte des points un peu imprécis et ceux totalement erronés. Parmi tous les points du suivi, un ensemble des points assez précis est recherché pour estimer le mouvement rigide tout en ignorant les points erronés. Ces derniers doivent par la suite être corrigés pour revenir sur de meilleurs endroits. Si l'estimation est effectuée en incluant les points erronés, d'étranges configurations du mouvement peuvent être obtenus et le suivi risque de ne plus fonctionner.

La figure 4.2 illustre un exemple où les bons points du suivi doivent être retrouvés pour estimer le mouvement rigide. Initialement, des points 2D sont sélectionnés sur le visage et ont une correspondance 3D sur le modèle virtuel. Il est à noter que seule la coordonnée 3D des points est nécessaire et non le modèle entier. L'obtention de ces coordonnées sera établie plus loin. Pour permettre une estimation précise du mouvement, il faudrait que les points noirs en d) soient identifiés pour les exclure de l'estimation. Cette estimation n'utiliserait donc que les 4 points blancs, ceci est suffisant car il en faut un minimum de 3.

La prochaine section explique le fonctionnement de l'algorithme RANSAC en général. Ceci permettra de mieux comprendre comment cet algorithme peut être utilisé pour estimer des paramètres lorsque de mauvaises données sont présentes dans l'ensemble utilisé.

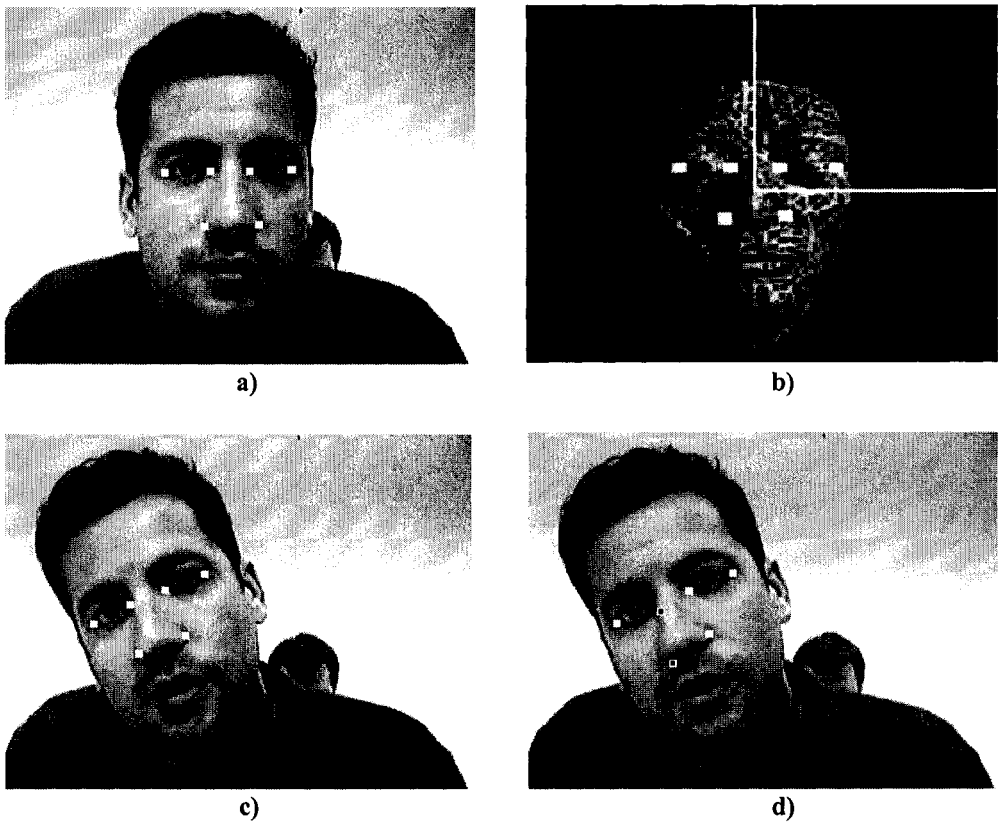


Figure 4.2. Exemple où il faut recueillir les bons points du suivi pour estimer le mouvement rigide. En a), les points sont initialement placés sur le visage. En b), la correspondance 3D sur le modèle des points de a). En c), les positions idéales des points du suivi. En d), les positions réelles des points qui peuvent survenir lors du suivi : les points noirs sont ceux mal localisés et identifiés.

4.4.1. L'algorithme RANSAC général

RANSAC [12] est souvent utilisé en vision artificielle pour estimer les paramètres d'un modèle quelconque lorsque, parmi les données utilisées, certaines sont imprécises et d'autres sont totalement fausses. RANSAC permet une bonne estimation si une grande proportion (plus de 50%) des données sont assez précises.

Pour expliquer le fonctionnement de RANSAC, quelques définitions sont utilisées.

N : l'ensemble total des données disponibles

p : l'ensemble des données jugées assez précises dans N

e : l'ensemble des données jugées fausses dans N

u_n : le $n^{ième}$ ensemble minimal des données pour estimer un modèle

u_p : le meilleur ensemble minimal de données pour estimer le meilleur modèle

m_n : un modèle quelconque généré avec u_n

m_p : le meilleur modèle cohérent à p et généré avec u_p

De façons répétées, un ensemble u_n doit être retiré de N pour estimer un modèle m_n. Selon le u_n choisi, le modèle estimé peut être cohérent ou non à l'ensemble p. Mais p et m_p sont inconnus, RANSAC doit donc pouvoir les récupérer grâce à des tests pour retrouver u_p. Ainsi, le modèle m_p généré avec u_p sera cohérent avec p et incohérent avec e. Pour déterminer quel sera l'ensemble u_p parmi les essais effectués, un score doit être attribué pour chaque essai. Le score maximal obtenu avec l'estimation d'un modèle m_n utilisant l'ensemble u_n donnera en fait m_p et u_p. Le score quantifie la cohérence du modèle estimé et peut être établi de plusieurs façons. Par exemple, pour un ensemble u_n,

il peut simplement être le nombre de données respectant un intervalle de précision avec le modèle m_n.

Afin de faciliter la compréhension, les figures 4.3 et 4.4 illustrent comment utiliser RANSAC pour obtenir une droite cohérente à un ensemble de points lorsque certains mauvais points sont présents. Ce problème tel quel n'est pas à résoudre dans ce projet, il s'agit seulement d'un exemple simple de RANSAC car il sera ensuite utilisé pour résoudre un problème plus complexe, celui de l'estimation des paramètres pour le mouvement rigide.

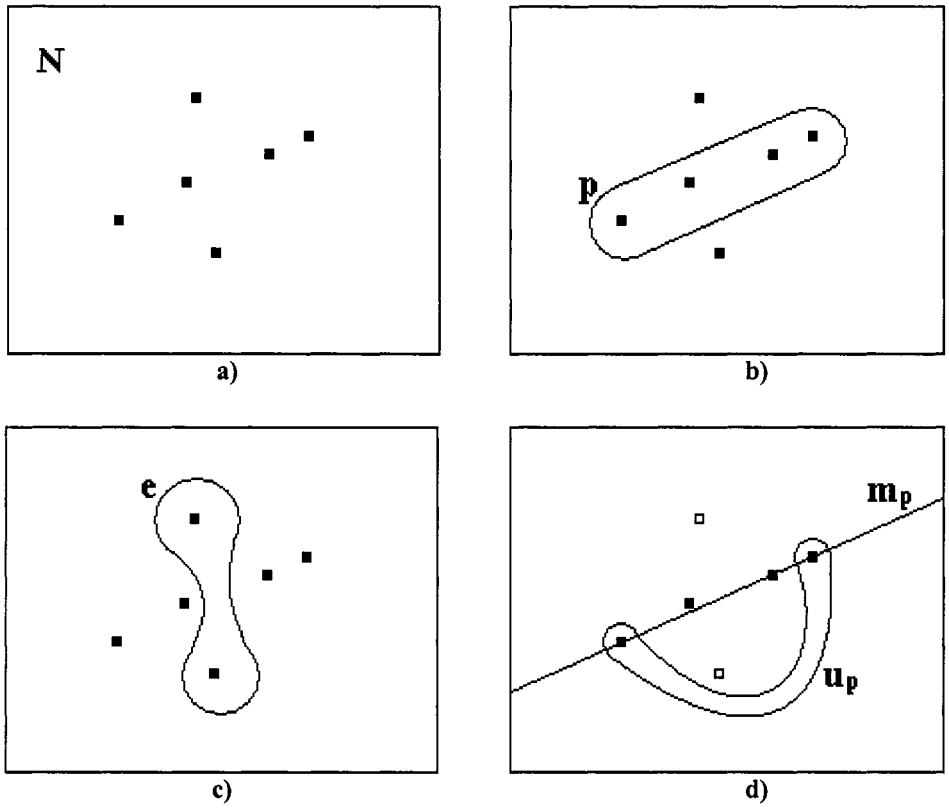


Figure 4.3. Exemple de la recherche d'une droite parmi des points cohérents et incohérents. En a), l'ensemble N de tous les points. En b), l'ensemble p des points cohérents. En c), l'ensemble e des points incohérents. En d), le meilleur ensemble minimal de points u_p permet d'estimer la meilleure droite m_p.

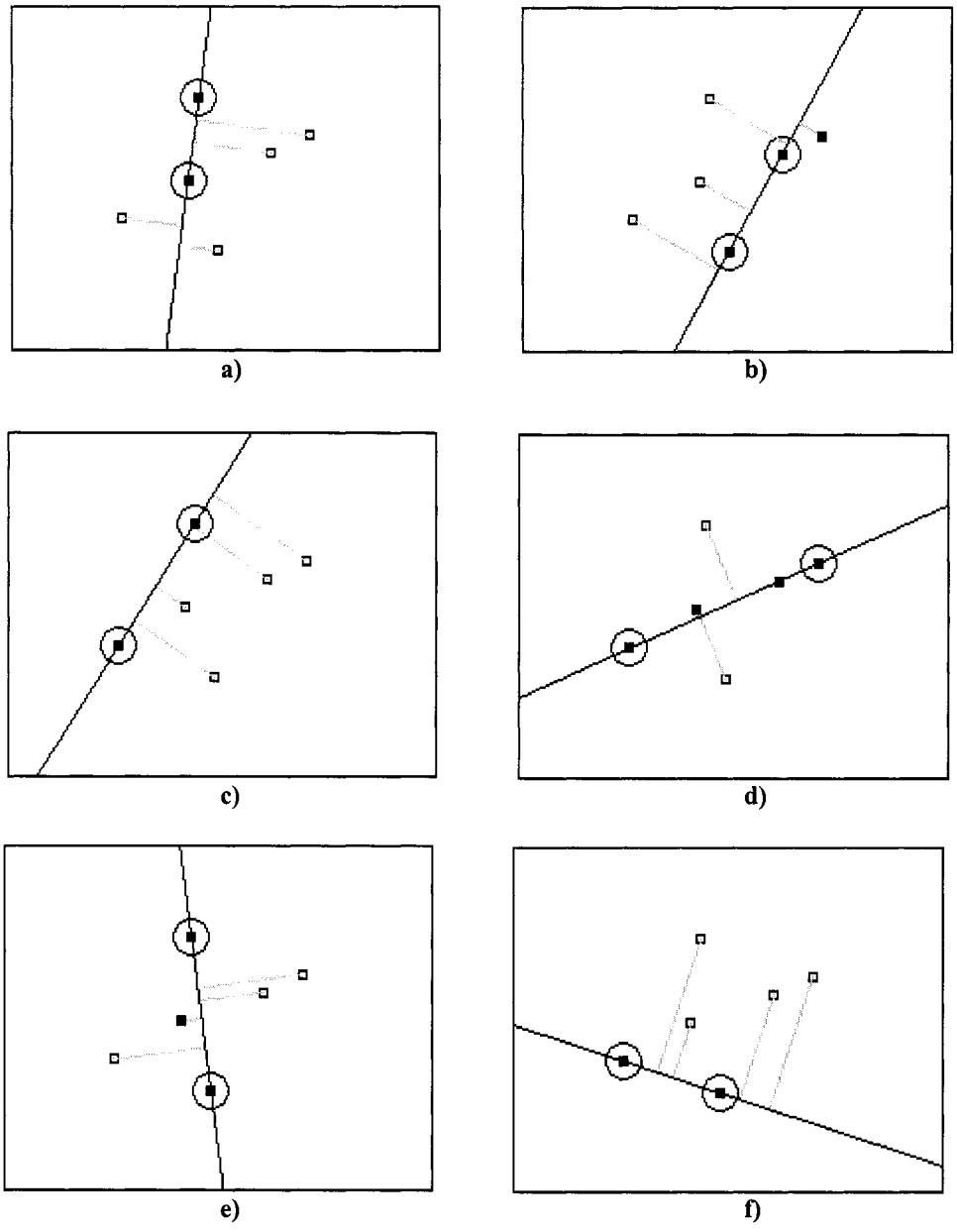


Figure 4.4. Divers essais effectués avec RANSAC pour trouver la meilleure droite. En a), score = 0. En b), score = 1. En c), score = 0. En d), score = 2, c'est la meilleure droite. En e), score = 1. En f), score = 0.

Sur la figure 4.3, les ensembles p et e ne sont pas initialement connus et une droite m_p doit être estimée à l'aide d'un ensemble u_p. Cette droite doit être cohérente au plus grand nombre de point de N. Sur cette figure, les points de p se situent très près de m_p grâce au bon ensemble u_p qui a été sélectionné. Pour tout autre ensemble u_n, moins de points de N se situeraient près de m_n.

Sur la figure 4.4, une suite d'essais est effectuée pour estimer une droite avec des ensembles de 2 points différents : il s'agit du nombre minimal de points nécessaires. Toutes les possibilités ne sont pas montrées mais seulement celles qui permettent de distinguer la bonne droite m_p (en d) des autres. Les points pleins représentent ceux qui sont assez près (d'après un seuil imposé) de la droite estimée et les points vides représentent ceux trop éloignés. En établissant pour score le nombre de points situés assez près de la droite, celle en d) remporte. Lorsque l'ensemble u_p est trouvé, la droite m_p n'est pas nécessairement la droite optimale passant par les points de p. L'ensemble p peut donc être utilisé pour estimer une meilleure droite par une minimisation des moindres carrés par exemple. Ceci est illustré à la figure 4.5 : en a), la droite passe exactement par les deux points extrêmes gauche et droite, sans tenir compte des deux autres points retenus. En b), les 4 points retenus sont considérés pour obtenir la droite optimale.

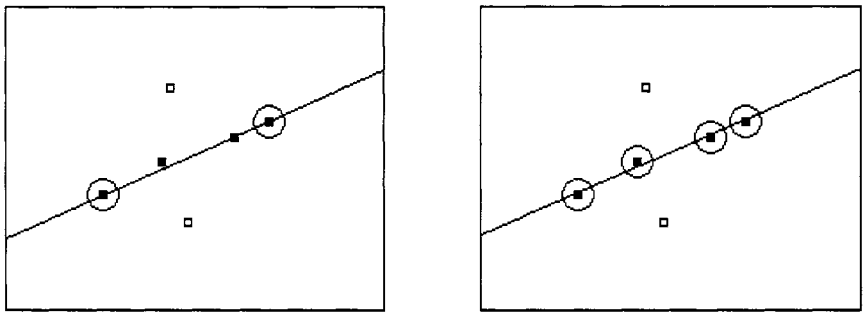


Figure 4.5 Raffinement de la droite avec l'ensemble p. En a), seulement u_p est utilisé. En b), p est utilisé avec une minimisation par moindres carrés

4.4.2. La méthode RANSAC pour estimer le mouvement rigide de la tête

Dans la section précédente, RANSAC était utilisé pour déterminer les paramètres d'une droite : le problème était en 2D et 2 points étaient nécessaires pour générer les paramètres nécessaires. Dans la présente section, RANSAC est utilisé pour déterminer le mouvement rigide 3D [12], c'est-à-dire les translations et rotations autour des axes X, Y et Z. Le problème est cette fois en 3D et un minimum de 3 points est nécessaire pour générer les paramètres. Pour expliquer le fonctionnement, les définitions de la section précédente sont ajustées.

N : l'ensemble total des points 2D du suivi au temps t

p : l'ensemble des points jugés assez précis dans N

e : l'ensemble des points jugés faux dans N

u_n : le $n^{ième}$ ensemble de points pour estimer le mouvement

u_p : le meilleur ensemble de points pour estimer le meilleur mouvement

m_n : un mouvement quelconque généré avec u_n

m_p : le meilleur mouvement cohérent à p et généré avec u_p

Et de nouveaux termes sont ajoutés afin d'expliquer le schéma de la figure 4.6 illustrant les étapes de la méthode d'estimation.

M_0 : l'ensemble des points 3D initiaux sur le visage de l'usager

e_c : l'ensemble des points e qui ont été corrigés grâce à m_p

N_c : le nouvel ensemble total des points 2D formé de p et e_c

Sur l'image initiale, l'ensemble M_0 est recueilli et la projection coïncide avec l'ensemble N. Lorsqu'un mouvement doit être estimé entre deux images consécutives (I_{t-1} et I_t), divers essais sont effectués avec des ensembles de points u_n (section 4.4.4) pour estimer des essais de mouvement m_n. Pour chaque essai, un score est calculé (section 4.4.5). D'après le meilleur score obtenu, le mouvement est utilisé pour récupérer tous les points cohérents, ceci forme l'ensemble p. Avec cet ensemble, qui contient habituellement plus de points que les autres essais, le mouvement m_p (section 4.4.6) est estimé afin d'avoir une meilleure précision. En sachant u_p, N et m_p, l'ensemble e est récupéré et corrigé (section 4.4.7) pour devenir e_c. L'ensemble Nc est ainsi obtenu et le suivi peut se poursuivre à l'image suivante.

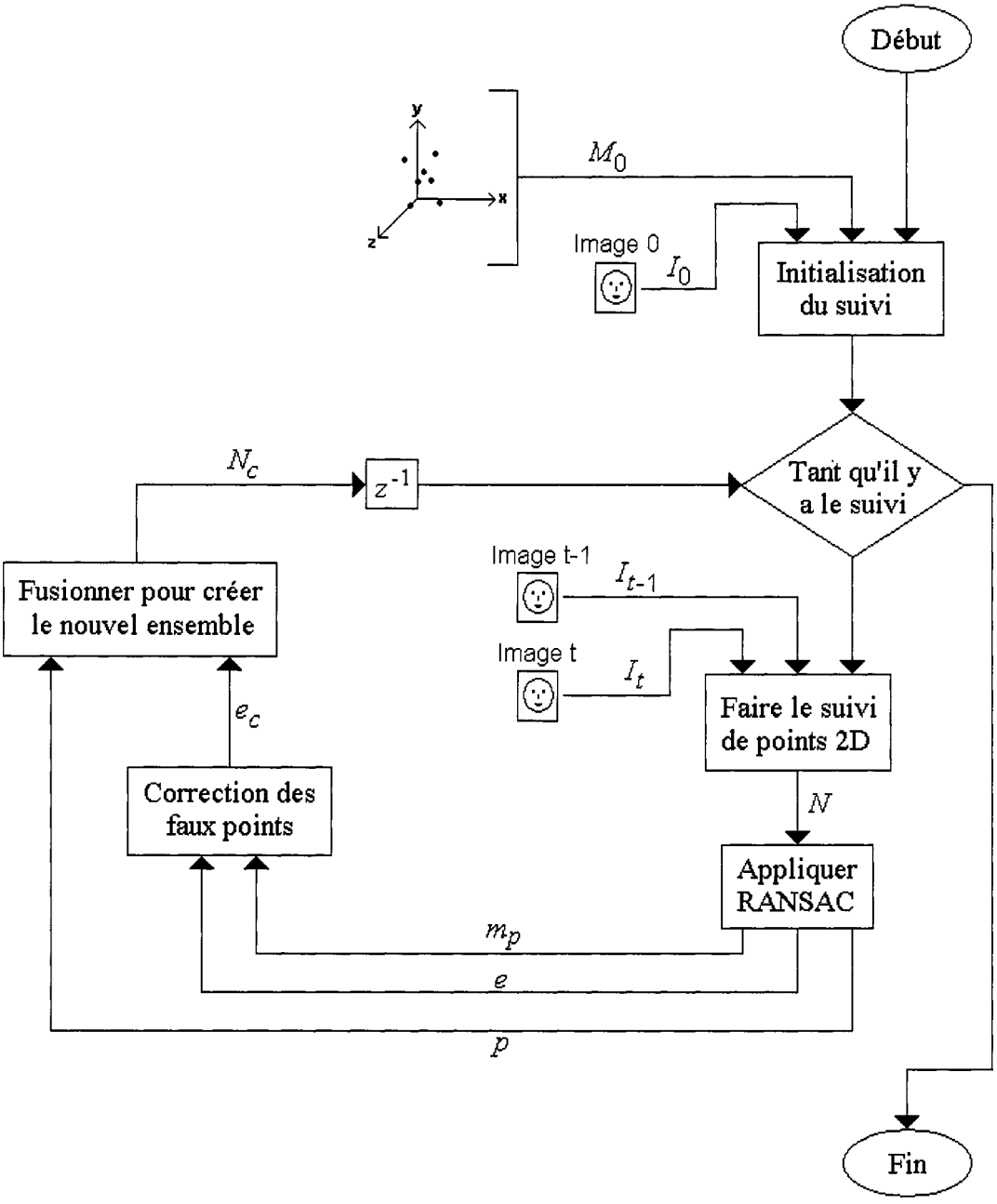


Figure 4.6. Schéma des étapes de la méthode d'estimation des mouvements rigides.

4.4.3. Sélection des points 2D dans le visage pour le suivi

La figure 4.7 illustre des endroits sur le visage où les points du suivi auraient pu être positionnés. Ces points sont situés sur des régions riches en texture et différentes de leur voisinage immédiat, ceci permet une grande stabilité lorsque le suivi s'effectue selon une méthode de recherche de régions similaires d'image en image.

Cependant, avec divers essais effectués avec ces points, certains ont permis d'avoir plus de stabilité que d'autres. Ceci est principalement dû aux changements causés par les mouvements rigides et non rigides. Par exemple, les points 1, 4, 5 et 8 vont se déplacer avec les sourcils, les points 12 et 13 vont se déplacer avec la bouche, les points 2, 3, 6 et 7 risquent d'être difficiles à retrouver s'il y a une fermeture des yeux et le point 10 va se déplacer lorsque l'usager penchera la tête vers l'avant.

Après plusieurs essais, les points 2, 3, 6, 7, 9 et 11 ont été jugés les meilleurs pour le suivi car leur position est principalement liée aux mouvements rigides, subit peu d'influence de la part des mouvements non rigides et est située sur un endroit riche en texture. Cette dernière caractéristique permet un suivi plus stable d'une image à l'autre. La position de ces points est représentée sur la figure 4.8 et les points sont identifiés de nouveau avec les lettres *a*, *b*, *c*, *d*, *e* et *f*.

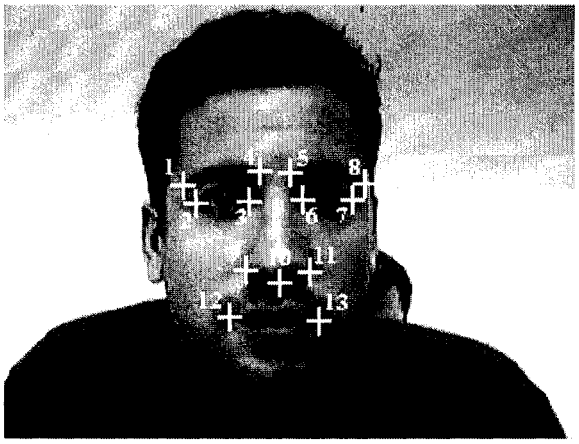


Figure 4.7. Positions potentielles des points du suivi.

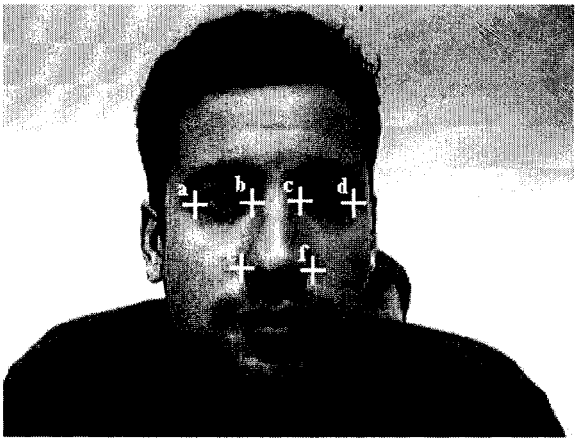


Figure 4.8. Positions utilisées des points du suivi après avoir effectué des tests de stabilité.

Bien que la position 2D de ces points soit directement obtenue sur l'image initiale, la profondeur ne peut être qu'approximative puisque seulement l'image du devant du visage est disponible. Cette approximation de la profondeur est obtenue à l'aide d'un modèle virtuel de l'usager. Ce modèle est récupéré de [84] et est basé sur les proportions moyennes du visage en tenant compte des contraintes anthropométriques. Avec ce modèle, les profondeurs sont estimées proportionnellement à la distance calculée initialement entre le centre des deux yeux de l'usager. Le tableau 4.1 illustre

comment ainsi obtenir la profondeur des points par rapport au centre de la tête. Ces profondeurs ne sont en fait qu'approximatives puisque, d'un individu à l'autre, les proportions peuvent varier en fonction de l'âge, la race, le sexe, etc (voir Annexe III).

Points	Profondeur
A	$0.72 \cdot dist\_yeux$
B	$0.9 \cdot dist\_yeux$
C	$0.9 \cdot dist\_yeux$
D	$0.72 \cdot dist\_yeux$
E	$1.18 \cdot dist\_yeux$
F	$1.18 \cdot dist\_yeux$

Tableau 4.1 Évaluation de la profondeur selon le modèle virtuel.

Où *dist_yeux* : distance entre le centre des yeux.

Grâce à cette estimation, les 6 points sélectionnés sur le visage ont tous des coordonnées en 3 dimensions. Ceci permettra d'estimer un mouvement grâce à l'équation 4.7 établie auparavant.

4.4.4. Création des ensembles des points

Chaque ensemble doit comporter un nombre suffisant de points pour calculer un estimé de mouvement avec l'équation 4.7. Bien que 3 points soient suffisants pour calculer un estimé, 4 points sont utilisés afin d'obtenir plus de stabilité au bruit. De plus, chaque ensemble doit au moins inclure l'un des points près du nez (*e* ou *f*) afin de bien identifier la rotation autour de l'axe X. Les 14 combinaisons possibles d'ensembles sont illustrées au tableau 4.2.

Numéro d'ensemble	Points utilisés
1	*a, b, e, f*
2	*a, c, e, f*
3	*a, d, e, f*
4	*b, c, e, f*
5	*b, d, e ,f*
6	*c, d, e, f*
7	*a, b, c, e*
8	*a, b, d, e*
9	*a, c, d, e*
10	*b, c, d, e*
11	*a, b, c, f*
12	*a, b, d, f*
13	*a, c, d, f*
14	*b, c, d, f*

Tableau 4.2. Combinaisons possibles de points pour créer les ensembles.

4.4.5. Attribution d'un score pour un mouvement

En effectuant le mouvement sur les points 3D initialisés sur le modèle virtuel de l'usager, la projection de ces points sur l'image est comparée avec les points 2D obtenus par suivi. Soit P_i un point 3D du modèle résultant de la projection sur l'image et p_i le point 2D correspondant au suivi. Le score n'a pas pour but de quantifier la précision d'un mouvement mais plutôt de déterminer lequel est le meilleur parmi plusieurs essais. Le mouvement avec le score maximal doit être retenu. Le score sera donc établi avec les informations des points p_i et les points P_i. Pour un mouvement de la tête, il est souhaitable qu'il y ait le plus de points p_i qui se situent près des points P_i. L'algorithme suivant est donc utilisé pour obtenir le score de chaque essai de mouvement:

- Initialiser le *Score* à 0
- Si les valeurs de *param3D* sont inférieures ou égales à celles de *param3D_max* alors
    - Parcourir *i* de 0 à N-1

- Si $dist_i$ est inférieure ou égale à *dist_max* alors augmenter le *Score* de 1

où *param3D* : les paramètres obtenus pour le mouvement

param3D_max : les paramètres maximaux acceptables pour le mouvement

$dist_i$: distance sur l'image entre P_i et p_i

dist_max : distance maximale acceptable entre P_i et p_i

Le tableau 4.3 indique les valeurs maximales fixées au mouvement entre deux images successives. Ces dernières sont peu sévères et permettent une grande liberté de mouvement de l'usager, peu importe sa distance par rapport à la caméra. Une étude plus approfondie pourrait permettre d'obtenir des valeurs plus strictes et de les ajuster en fonction de la distance à la caméra. Il est à noter qu'il n'y a aucune restriction pour la translation en profondeur puisque celle-ci ne peut pas être calculée.

Paramètres	Valeurs
param3D.t_x	20
param3D.t_y	20
param3D.θ_x	$\pi/8$
param3D. θ_y	$\pi/8$
param3D. θ_z	$\pi/8$
dist_max	5

Tableau 4.3. Valeurs maximales permises pour augmenter le score.

Le score de chaque mouvement est donc en fonction de la distance euclidienne entre les points p_i et P_i. Ainsi, plus il y aura de points p_i assez près de leur

correspondance P_i, plus l'estimation du mouvement sera acceptable. Dans le cas idéal, les points p_i et P_i devraient se chevaucher sur l'image. Cependant, ce cas est peu probable étant donné l'imprécision du modèle virtuel de l'usager d'où les points P_i sont récupérés. De plus, même avec un modèle virtuel très précis, la méthode ne serait pas robuste pour de grandes translations en profondeur. Il est à noter que le score d'un essai aurait pu être établi de différentes façons. Dans ce projet, la méthode a été inspirée de [12]. D'autres essais ont été effectués, comme l'analyse de la géométrie des points p_i d'un ensemble par rapport aux autres, mais les résultats n'étaient pas convaincants. Puisque l'algorithme RANSAC utilise peu de points pour chaque essai, des imprécisions sur ceux-ci sont très néfastes pour les résultats. L'imprécision du modèle virtuel est donc très néfaste pour l'estimation du mouvement. Heureusement, la dernière étape de RANSAC consiste à utiliser plus de points pour estimer de nouveau le mouvement, ceci favorise la stabilité du système.

Les mouvements trop brusques sont évités grâce à *param3D* et de trop grandes distances entre les points P_i et p_i sont évitées grâce à *dist_max*. La figure 4.9 illustre un exemple d'un mouvement refusé. Il est à noter que les images a) et b) sont très semblables puisqu'il s'agit de deux images consécutives dans la séquence mais le passage entre a) et b) est tout de même trop brusque. La figure 4.10 illustre les distances entre les points p_i. Sur cette figure, les points P_i sont représentés par des carrés blancs tandis que les points p_i sont représentés par des croix blanches. La distance $dist_a$, par exemple, représente la distance entre la position de a lors du suivi 2D et la projection du point 3D correspondant lors du mouvement rigide essayé.

Figure 4.9. Refus d'un mouvement trop brusque. En a), le mouvement à l'image I_{t-1}. En b), un mouvement trop brusque refusé à l'image I_t.

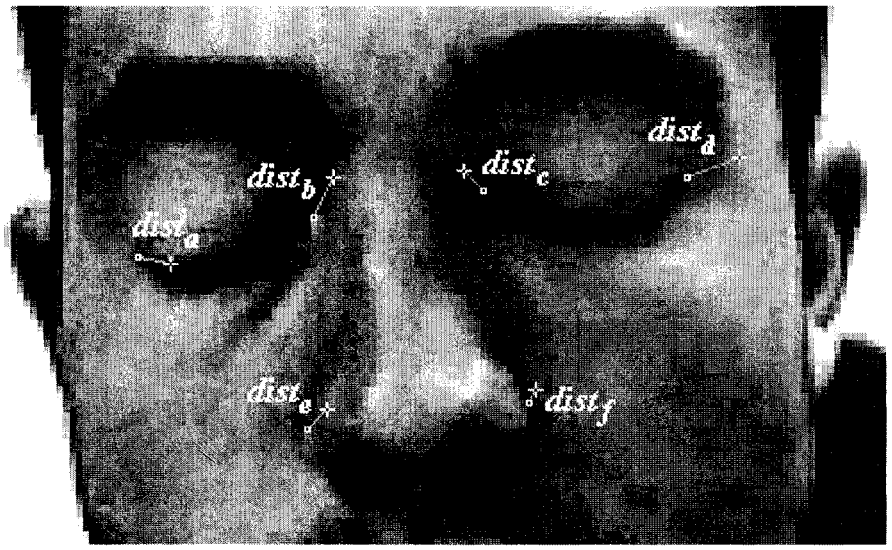


Figure 4.10. Démonstration des distances entre les points P_i et p_i. Les P_i sont les carrés blancs et les p_i les croix blanches.

4.4.6. Estimation finale du mouvement

Lorsque le meilleur essai de mouvement m_p est obtenu (celui avec le meilleur score), les distances $dist_i$ entre les points P_i et les points p_i sont recalculées. Le mouvement final M_f est estimé avec l'équation 4.7 en incluant tous les points p_i dont la distance à P_i est inférieure ou égale à *dist_max*. Plus il y aura de points inclus, meilleure sera la stabilité du résultat (en supposant qu'il n'y ait pas de points totalement erronés dans l'ensemble récupéré). La figure 4.11 illustre un exemple où les bons points p_i sont récupérés après avoir calculé m_p pour ensuite calculer M_f.

a) b)

Figure 4.11. Exemple du calcul de M_f avec les points récupérés avec m_p. En a), m_p a été calculé avec l'un des ensemble de 4 points (les croix blanches). En b), les points cohérents à m_p (les croix blanches) ont été utilisés pour calculer M_f.

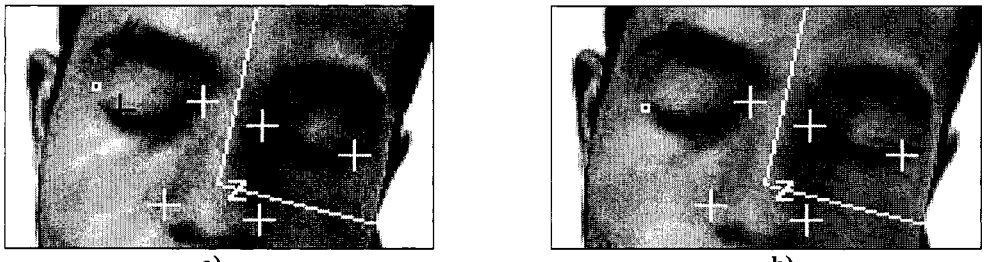
4.4.7. Relocalisation des points erronés

Les points p_i qui ne sont pas utilisés pour calculer M_f sont jugés erronés et doivent être relocalisés pour poursuivre le suivi sur les images suivantes. Il suffit alors de prendre la projection des points P_i correspondants. Puisque M_f n'est qu'une estimation et que les profondeurs des points imposés initialement comportent des

imprécisions, il est important de donner un peu de liberté aux points relocalisés afin qu'ils se situent sur des endroits qui permettront de poursuivre le suivi de points 2D. La projection des points P_i est donc un point de départ pour la relocalisation. Par la suite, une meilleure position est recherchée en comparant la région sur l'image courante avec la région sur l'image initiale en utilisant la SSD et la NCC (voir chapitre 3). L'intervalle de recherche sera cette fois plus grand que *dist_max* pour 2 raisons :

1. Permettre de retrouver la bonne position même si elle est éloignée;
2. Permettre qu'un point soit encore refusé pour le prochain calcul du mouvement s'il converge vers une mauvaise position.

Après plusieurs essais expérimentaux, cet intervalle de recherche a été fixé à 2·dist_max afin de laisser une liberté suffisante. La figure 4.12 illustre un exemple de relocalisation d'un point erroné. Sur cet exemple, le point relocalisé sera probablement encore rejeté sur le prochain calcul du mouvement car il a convergé vers une mauvaise position. Ce point redeviendra peut-être valide lorsque l'usager ouvrira les yeux car la région recherchée sera de nouveau semblable à celle sur l'image initiale.



a) b)

Figure 4.12. Exemple de relocalisation d'un point erroné. En a), le point erroné (la croix noire) est d'abord relocalisé (le carré blanc) à l'aide de la projection du point 3D correspondant et en fonction du mouvement Mf. En b), ce point se relocalise de nouveau en essayantde retrouver une région semblable à celle de l'image initiale.

4.4.8. Les sources d'imprécisions au système

Avec le système proposé, des imprécisions sont inévitables concernant les paramètres du mouvement rigide. Voici quelques sources d'imprécisions avec leurs impacts :

La projection de la scène est considérée orthographique

Comme déjà mentionné, la translation en profondeur de la tête ne peut pas être estimée. Ce paramètre du mouvement est donc directement affecté. De plus, la projection des points 3D du modèle virtuel se superposera difficilement sur les points 2D du suivi lors de grandes translations en profondeur. En effet, la dimension du visage de l'usager peut varier en fonction du temps mais pas la dimension du modèle virtuel. Donc même si de bons points du suivi sont sélectionnés pour estimer le mouvement rigide, les points utilisés peuvent engendrer de faux mouvements de la tête. Il est alors difficile pour le système de choisir le meilleur ensemble de points 2D permettant d'estimer le meilleur mouvement. Lors d'une mauvaise estimation du mouvement rigide, il y a peu de ressemblance entre la projection des points 3D et les points 2D du suivi. De bons points 2D risquent donc d'être relocalisés sur de mauvais endroits par le système de correction, ceci peut affecter tout le reste du suivi. Il est donc difficile d'évaluer avec précision l'erreur introduite par la considération d'une projection orthographique car elle peut se manifester à plusieurs niveaux. Il est cependant nécessaire que peu de translations en profondeur soient rencontrées sur une séquence d'image.

La correspondance des points 3D est imprécise

Un modèle virtuel est supposé être adapté initialement au visage de l'usager. Les seuls paramètres ajustables sont la largeur et la hauteur de la tête. Des coordonnées 3D sur le visage de ce modèle sont utilisées pour estimer le mouvement rigide et celles-ci sont imprécises car les proportions du visage varient beaucoup d'un individu à l'autre (voir Annexe VI). De plus, il peut y avoir des imprécisions sur la taille estimée de la tête de l'usager. Un problème de superposition de la projection des points 3D sur les points 2D du suivi peut donc encore être rencontré.

Les points 2D du suivi affectés par le mouvement non rigide

Sur l'image initiale, les points 2D du suivi ont été sélectionnés afin d'être riches en texture tout en étant invariants aux mouvements non rigides. Malheureusement ces contraintes sont difficiles à respecter. D'après des essais expérimentaux effectués, les points sur les coins des yeux sont généralement très stables. Cependant, les points sur les côtés des narines peuvent être déplacés par les mouvements de la bouche. Ceci est problématique puisque l'estimation du mouvement considère que l'objet concerné est rigide, donc que les points ne bougent pas les uns par rapport aux autres. Ces déplacements cependant sont rarement rencontrés lors de mouvements naturels de la bouche. Dans ce projet, ces imprécisions sont plutôt considérées négligeables.

4.5. Analyse des résultats obtenus

Quatre séquences d'images ont été utilisées pour quantifier les résultats et ensuite en faire une analyse qualitative. Sur ces séquences, les mesures ont été prises lorsque de grandes variations de mouvements étaient visibles, ceci afin de mettre en évidence l'estimation du mouvement rigide. La précision des mesures ne peut cependant pas être obtenue puisque les paramètres du mouvement réel de l'usager sont inconnus. Seuls les paramètres expérimentaux (t_x, t_y, t_z, θ_x, θ_y et θ_z) peuvent être quantifiés. Pour t_z, la valeur est toujours nulle car la projection est supposée orthographique pour la simplification des calculs, ceci ne permet pas d'obtenir la profondeur. Des axes permettent de visualiser les mouvements des quatre séquences sur des figures paramètres sont fournis dans des tableaux respectifs. Les détails de l'analyse sont présentés en Annexe II.

Certaines techniques auraient cependant pu être utilisées pour mesurer la précision des résultats (section 4.6). Étant donné le temps consacré pour le bon fonctionnement du projet, ces techniques de mesures n'ont malheureusement pas été exploitées.

4.6. Améliorations suggérées

Les résultats seraient sans doute meilleurs si un modèle plus précis de l'usager était disponible, cela pourrait être fait à l'aide de pauses initiales de l'usager. Une meilleure détection initiale des points 2D sur le visage, avec un meilleur modèle, permettrait aussi

une meilleure stabilité du suivi. Finalement, les paramètres du mouvement pourraient être estimés à l'aide d'un filtre Kalman [84], ceci permettrait d'adoucir les mouvements et prévenir les paramètres peu probables. Ce filtre utilise les paramètres du mouvement obtenus des images précédentes, les mesures obtenues du suivi de points 2D et des paramètres spécifiques (les déviations standard acceptables pour les paramètres du mouvement). Ce filtre semble être assez robuste pour donner un mouvement stable et continu malgré les erreurs de mesure du suivi, ceci pourrait diminuer les tremblements.

Finalement, il serait plus facile d'analyser les résultats en quantifiant la précision. Pour ce faire, il serait important d'exploiter des techniques permettant d'avoir le mouvement réel de l'usager. Par exemple, des capteurs placés sur la tête de l'usager pour en mesurer le déplacement. Un visage virtuel réaliste généré en infographie aurait pu aussi être utilisé. La manipulation manuelle de l'axe du mouvement ou des points du suivi aurait aussi permis de quantifier la précision des résultats.

CHAPITRE 5

LOCALISATION DES ÉLÉMENTS NON RIGIDES

Introduction

Les mouvements non rigides du visage sont ceux engendrés par les diverses expressions faciales et surviennent inévitablement lors d'une vidéoconférence. Par exemple, des mouvements sont observés sur les yeux, les sourcils, la bouche, le front, etc. L'intérêt de la localisation de ces éléments non rigides est de permettre à un modèle virtuel d'imiter les expressions faciales en plus d'imiter le mouvement de la tête. L'adaptation du modèle virtuel aux mouvements non rigides n'a cependant pas été développée dans ce projet étant donné l'ampleur des travaux. Ces mouvements seraient paramétrisés à l'aide de diverses localisations d'éléments d'intérêt sur le visage. Dans ce chapitre, la localisation est étudiée et concerne celle des yeux, des sourcils et de la bouche. Ces derniers semblent être les plus importants du visage car ils sont très impliqués pour les différentes expressions faciales. La difficulté dans la détection de ces éléments provient principalement du fait qu'ils varient beaucoup d'une personne à l'autre ainsi que pour une même personne selon les expressions. La couleur de l'iris, par exemple, peut varier d'une personne à l'autre et l'œil a plusieurs états (ouverts, fermés, etc). Et pour la bouche, plusieurs configurations très différentes les unes des autres peuvent survenir. Pour localiser les éléments non rigides, des modèles géométriques utilisant des droites, des cercles et des paraboles, seront positionnés aux contours des éléments.

La figure 5.1 illustre un schéma représentant le fonctionnement géréral du suivi des éléments non rigides, c'est-à-dire la localisation des modèles géométriques sur le visage d'image en image. Le centre des yeux et les coins de la bouche sont d'abord obtenus sur la première image. Ces éléments sont recherchés initialement car ils ont semblé les plus faciles à trouver et sont suffisants pour le suivi avec les méthodes qui seront élaborées. Bien que les yeux et la bouche aient des repères initiaux sur l'image initiale, les sourcils n'en ont pas et se réfèrent plutôt à ceux des yeux. En effet, les sourcils sont juste au-dessus des yeux, ceci simplifie beaucoup leur localisation. D'ailleurs, la localisation indépendante des sourcils semblait difficile avec des essais effectués car un sourcil contient peu de caractérisques propres à lui-même et peut ressembler beaucoup à un œil fermé ou à des cheveux par exemple. En ce qui concerne les coins de la bouche, les traitements sont effectués d'une façon indépendante car ces coins peuvent bouger beaucoup sur le visage. Une méthode aurait pu être élaborée permettant l'inter-dépendance de tous les éléments non rigides mais aurait demandé une étude plus approfondie et le suivi aurait sûrement été plus lourd en temps de calcul car beaucoup de paramètres auraient été concernés simultanément. Par exemple, la localisation des yeux serait basée à la fois sur l'image et sur la localisation des autres éléments comme les sourcils et la bouche.

Il est à noter que le suivi des éléments non rigides est aussi effectué indépendamment du mouvement rigide de la tête. Cela aurait pu être contourné en imposant des contraintes sur les positions des éléments mais il a été préférable de ne pas imposer ces contraintes puisque beaucoup d'imprécisions peuvent être rencontrées lors

de l'estimation des mouvements rigides. Finalement, en terme de rapidité de calcul, il est plus simple de traiter les mouvements rigides et non rigides d'une façon indépendante.

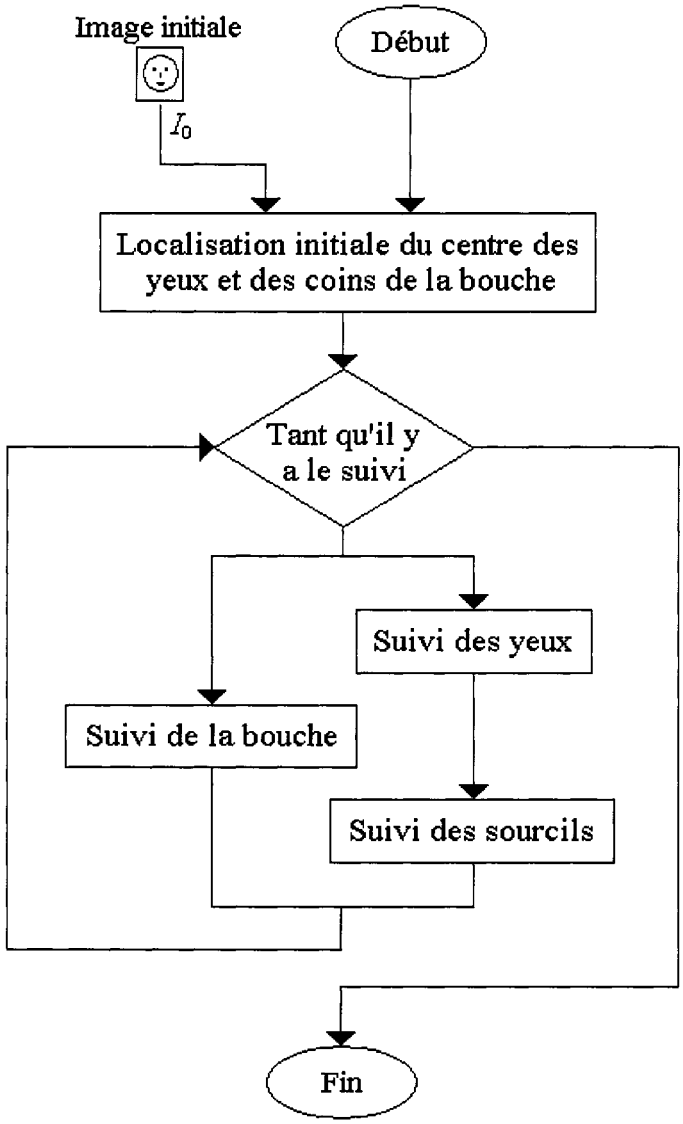


Figure 5.1. Schéma du fonctionnement géréral du suivi des éléments non rigides.

5.1. Le suivi des yeux

Le suivi des yeux est effectué grâce à l'estimation de la position du centre des yeux acquis sur la première image I_0. Comme il sera démontré, un certain intervalle d'erreurs est acceptable pour cette estimation. Le suivi sera effectué grâce à plusieurs caractérisques anthropométriques communes à la plupart des individus, ceci facilite grandement le travail à effectuer. Ce suivi est en fait une tâche très difficile étant donné le nombre d'états différents que peut avoir un oeil. Les images des yeux varient beaucoup d'un individu à un autre en plus de varier pour un même individu, selon l'état de ses yeux. Le suivi des yeux est donc la tâche la plus complexe de ce projet concernant l'estimation des mouvements non rigides. Pour l'estimation des mouvements des yeux, voici les informations jugées pertinentes :

1. Le contour de l'ouverture des yeux ;
2. Le centre des iris ;
3. L'état binaire de l'ouverture : ouvert ou fermé.

a) b)

Figure 5.2. Localisation des yeux ouverts et fermés.

Les tableaux 5.1, 5.2 et 5.3 indiquent plusieurs caractéristiques concernant la plupart des images des yeux.

Les coins d'un œil	
1	Les deux coins de l'oeil se situent aux extrêmes (à droite et à gauche) de l'œil
2	Les quatre coins engendrés par les deux yeux sont presque colinéaires
3	L'espace libre entre les deux yeux est presque de la même largeur qu'un oeil

Tableau 5.1. Caractéristiques propres aux coins d'un œil.

L'ouverture d'un œil	
1	L'ouverture de l'oeil est constituée de deux courbes reliant les deux coins
2	La courbe de la paupière du haut est souvent plus foncée que celle du bas
3	La courbe du haut est concave vers le haut et celle du bas, concave vers le bas
4	Les sommets des concavités sont souvent au centre des courbes
5	Les deux courbes forment de petits angles aux coins à leurs intersections
6	Par rapport à l'axe de la tête, les yeux sont étendus à l'horizontale
7	La largeur des yeux est constante peu importe l'état
8	La courbe de la paupière du haut est souvent plus foncée que celle du bas
9	Un oeil peut être ouvert ou fermé
10	L'ouverture d'un oeil est indépendante de l'ouverture de l'autre
11	Pour un œil ouvert, le blanc d'oeil est toujours visible
12	Pour un œil fermé, le blanc de l'oeil et l'iris sont invisibles
13	Pour un oeil fermé, le contour constitue une seule courbe foncée
14	Pour un oeil fermé, la courbe du contour est très souvent concave vers le bas
15	Les yeux sont situés en bas des sourcils

Tableau 5.2. Caractéristiques propres à l'ouverture d'un œil

L'iris	
1	La pupille est le centre d'une région circulaire très foncée à l'intérieur de l'œil
2	Le rayon de l'iris est constant peu importe l'état
3	La couleur de l'iris varie d'un individu à l'autre mais pas pour un même individu
4	Pour un oeil de couleur foncée, la pupille est rarement visible
5	Pour un oeil de couleur claire, la pupille est très souvent visible
6	Entre l'iris et le reste de l'oeil, le changement de couleur est brusque
7	Le rapport du rayon de l'iris sur la distance entre les yeux est d'environ 1/9
8	Les deux yeux regardent dans la même direction
9	Pour un oeil ouvert, l'iris est visible entièrement ou partiellement
10	Pour un oeil ouvert, le bas de l'iris est très souvent visible

Tableau 5.3. Caractéristiques propres à l'iris.

5.1.1. Brève revue de la littérature sur la détection des éléments non rigides

Dans [46], des masques déformables sont d'abord utilisés pour localiser les coins des yeux. Le contour d'un oeil est retrouvé grâce à la trajectoire empruntée par des arêtes entre les coins. Le suivi de l'ouverture des yeux est effectué avec une méthode semblable aux « snakes ». Des essais expérimentaux avaient été effectués et les résultats étaient décevants. De plus, le suivi ne peut fonctionner que si les images consécutives sont très semblables et que les yeux soient très visibles, ceci n'est pas toujours le cas pour ce projet.

Dans [72], des masques déformables sont utilisés pour localiser les yeux et faire le suivi. Le modèle d'un œil est constitué de 2 paraboles pour les fermetures et d'un cercle pour l'iris. Les images des arêtes, des vallées et des sommets (voir Annexe II) sont

utilisées pour faire converger le modèle. Des essais expérimentaux ont été effectués avec cette technique mais de mauvaises convergences étaient trop souvent observées : cela était dû aux mauvaises positions initiales des modèles. De plus, les images traitées étaient beaucoup plus réalistes pour le projet concerné que celles présentées dans l'article.

Dans [49], les yeux sont détectés à l'aide d'un filtre construit à partir de plusieurs images d'exemples d'œil. La couleur peau est d'ailleurs extraite de l'image pour éliminer les positions inutiles. Lorsqu'une position est estimée grâce au filtre, un modèle d'oeil est utilisé pour augmenter la précision.

Dans [50], les yeux sont détectés et suivis mais la résolution des images doit être beaucoup plus élevée que celle retenue dans ce projet. De plus, l'algorithme ne semble fonctionner que si les mouvements des yeux sont très lents.

Dans [58], les yeux semblent être détectés d'une façon trop simple, ceci ne peut pas être appliqué à ce projet. L'arrière-plan est contrôlé et un seuil est appliqué sur l'image pour ne retenir que les pixels assez foncés. Des contraintes sur les régions retenues sont utilisées ainsi qu'un modèle se basant sur l'image des arêtes pour localiser les yeux. Le suivi n'est pas traité.

Dans [60], les yeux sont détectés en utilisant l'image des arêtes et des modèles d'œil 3D. Ces modèles ne peuvent cependant représenter que quelques états possibles des yeux, soit le regard à gauche, à droite et devant. Le suivi n'est pas traité.

Dans [62], un modèle géométrique est utilisé pour effectuer le suivi des yeux. Ce modèle est constitué de 2 paraboles pour l'ouverture et d'un cercle pour l'iris. La forme

semi-circulaire du bas de l'iris est d'abord recherchée grâce à l'image des arêtes et en vérifiant si la couleur de l'iris détecté est semblable à la couleur initiale. Si l'iris n'est pas détecté correctement, l'œil est considéré fermé. Grâce au suivi des coins intérieurs des yeux, les coins extérieurs peuvent être estimés grâce à quelques contraintes anthropométriques et finalement le contour de l'ouverture est adapté. Bien que cette méthode semble peu robuste pour de grands mouvements et qu'il faille fournir des données intiales, plusieurs modules semblent très intéressants pour le projet concerné.

Dans [67], les yeux sont retrouvés à l'aide des tâches foncées de l'image et d'un modèle permettant d'accepter ou de rejeter les « blobs ». Cette technique semble cependant peu robuste car un seuil sur l'intensité est nécessaire ainsi qu'une dimension approximative des yeux à retrouver. Le suivi n'est pas traité.

5.1.2. Description générale de la méthode utilisée

L'image d'un oeil étant très complexe, le choix de la méthode utilisée pour effectuer le suivi a été basé sur plusieurs caractéristiques devant à la fois être invariantes d'un individu à l'autre tout en étant très visibles sur les images. La méthode utilisée a grandement été inspirée de [62] mais plusieurs modifications y ont été apportées afin qu'elle soit mieux adaptée à notre contexte. Divers essais avaient d'ailleurs été effectués sur [46] et [72] mais des résultats peu concluants avaient été obtenus.

C'est en analysant beaucoup d'images traitées comme les arêtes, les vallées et les sommets (voir Annexe II) qu'une méthode a pu être proposée. Plusieurs algorithmes et essais de paramètres ont été essayés afin d'obtenir ces images traitées dans le but d'avoir

la meilleure visibilité possible sur des éléments importants à la localisation. Par exemple, sur l'image des arêtes, les contours des yeux n'étaient pas souvent visibles mais le bas de l'iris l'était cependant très souvent. Les arêtes sont donc utilisées pour localiser. Pour les arêtes, l'algorithme Canny avec des paramètres spécifiques a été utilisé car d'autres méthodes essayées ne permettaient pas d'obtenir une aussi bonne visibilité du bas de l'iris. Évidemment, moins une caractéristique recherchée sur une image est visible et isolée, plus il est difficile de la trouver et plus le taux de réussite risque d'être bas. Pour la recherche d'un élément complexe comme un œil, il est préférable de chercher d'abord les éléments simples, comme l'iris, et ensuite poursuivre avec les éléments de plus en plus difficiles. Cela permet de diminuer les recherches et de récolter le plus d'informations possibles aux recherches.

Le suivi des yeux a donc été décomposé selon les quatre étapes suivantes :

1. Localiser les deux iris potentiels ;
2. Estimer le sommet et la base de l'ouverture des yeux ;
3. Déterminer si l'oeil est ouvert ou fermé ;
4. Adapter le contour de l'œil.

5.1.3. Localiser les deux iris potentiels

Les caractéristiques 9 et 10 de l'ouverture de l'œil et les caractéristiques 2, 6, 7, 8 et 10 de l'iris (présentées au début de ce chapitre) sont utilisées pour effectuer la première étape du suivi. Bien que les yeux puissent être fermés, cette étape présume d'abord que les yeux soient ouverts et que les deux iris soient donc visibles. Les iris

sont les éléments recherchés en premier car ils sont habituellement très visibles sur les images, ceci permet de faciliter la détection. De plus, le ratio du rayon des iris sur la distance qui les sépare est pratiquement constant, ceci sera élaboré plus loin. Les iris sont donc recherchés et pour ce faire, l'image des arêtes est obtenue afin de mettre en évidence le contour des iris comme l'illustre la figure 5.1. Sur cette figure, le contour du bas des iris est très visible contrairement au reste des yeux. Le bas des iris sera donc recherché pour ensuite obtenir deux groupes d'iris potentiels G_{IG} et G_{ID} pour les iris de gauche et de droite respectivement. Pour trouver ces groupes d'iris potentiels, un modèle M_I représentant le bas de l'iris est parcouru dans une région de recherche R_I. Ce modèle est illustré à la figure 5.3.c) et 5.4. Ce dernier est constitué d'un demi-cercle d'une épaisseur $2 \cdot \Delta M_I$ afin de pouvoir contenir les arêtes du bas de l'iris. Le rayon R et la demi-épaisseur ΔM_I peuvent être estimés grâce à l'anthropométrie du visage et la distance D entre les deux yeux, comme illustré par l'équation 5.1.

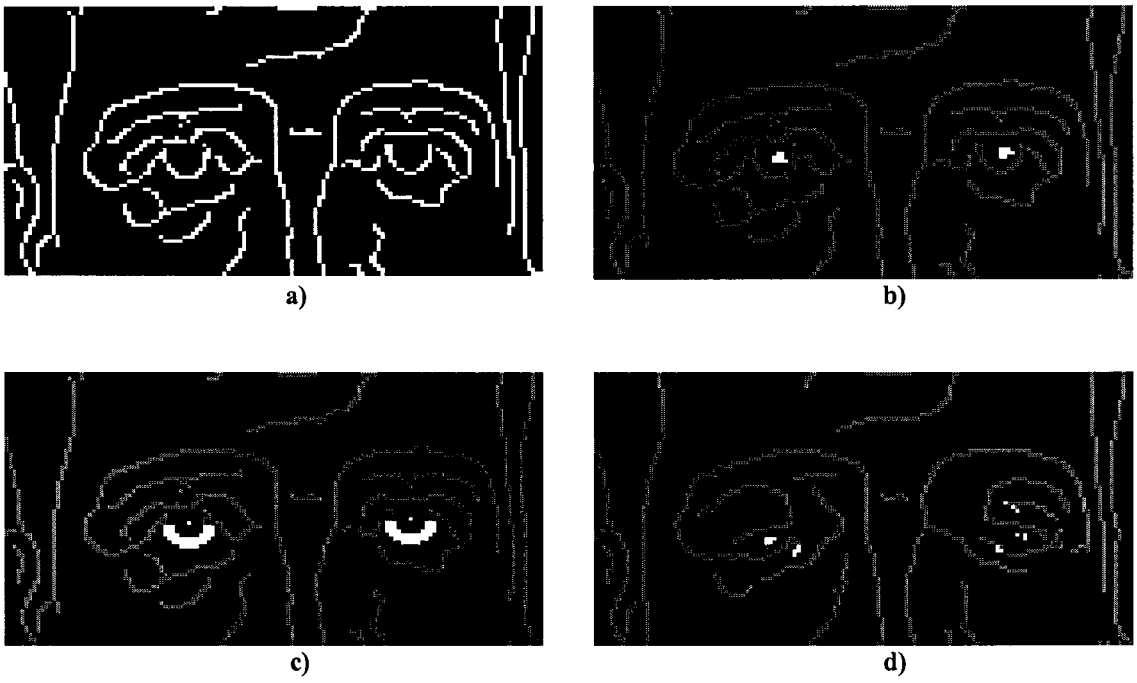


Figure 5.3. La détection des groupes d'iris potentiels G_{IG} et G_{ID}. En a), l'image des arêtes. En b), les iris potentiels détectés (en blanc). En c), les modèles M_I vis-à-vis du bas des iris (en blanc). En d), les iris potentiels détectés lorsque les yeux sont fermés (en blanc).

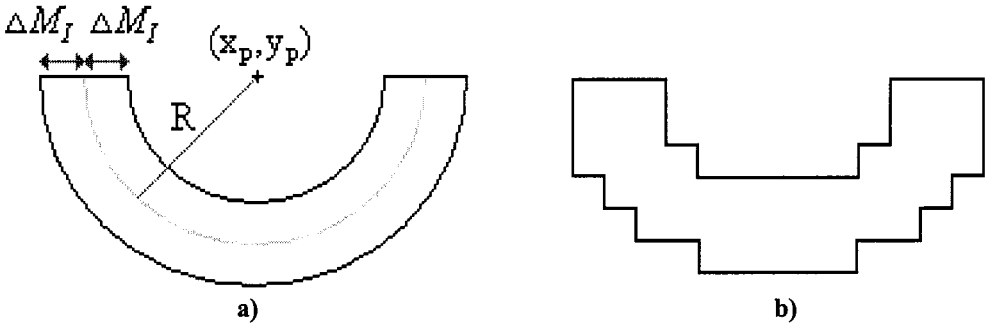


Figure 5.4. Création du modèle M_I pour détecter les iris potentiels. En a), le modèle théorique. En b), un exemple de modèle pratique utilisé.

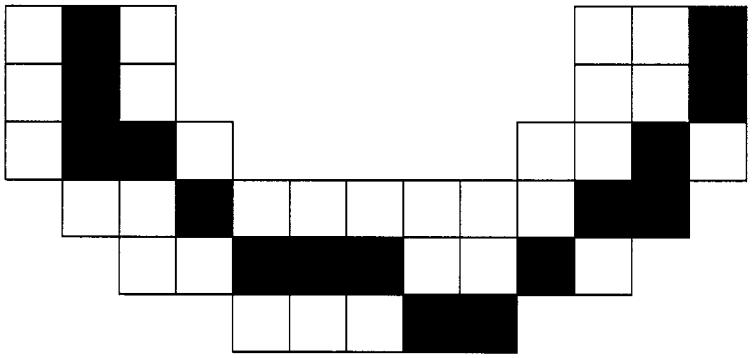
$$R \approx \frac{D}{9} \tag{5.1}$$

$$\Delta M_I \approx \frac{R}{4} \approx \frac{D}{36}$$

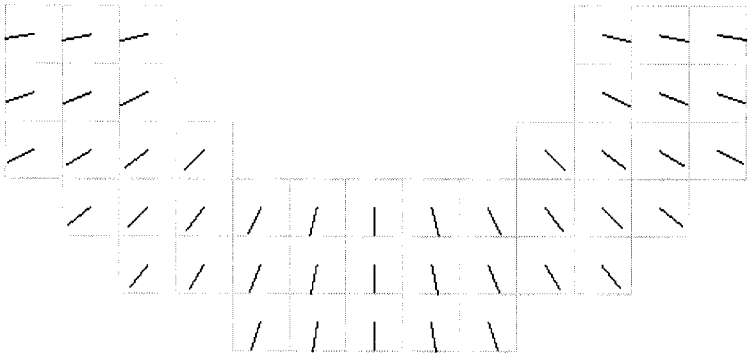
Lorsque M_I est créé, celui-ci doit ensuite parcourir tous les pixels de R_I afin d'identifer les N positions où les plus grands nombres d'arêtes ont été rencontrés à l'intérieur de M_I. Ces N positions, pour la gauche et la droite, constitueront les groupes d'iris potentiels G_{IG} et G_{ID} respectivement. L'équation 5.2 permet de calculer le nombre de pixels d'arêtes pour un certain M_I. La figure 5.5.a) illustre un exemple où 16 pixels d'arêtes ont été rencontrés.

$$Nombre\_arête = \sum_{M_I} I_{arête}(\vec{p}) \tag{5.2}$$

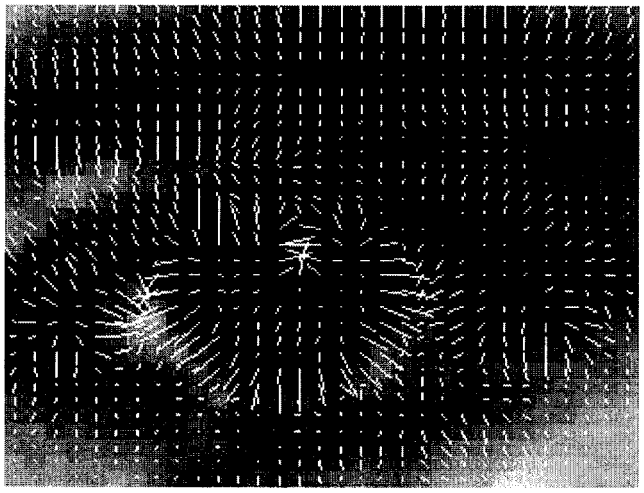
où $I_{arête}$ représente une image binaire des pixels d'arêtes avec la valeur 1 s'il s'agit d'un pixel d'arête et 0 sinon. Et $\vec{p}$ est un vecteur de position des pixels de M_I sur $I_{arête}$.



a)



b)



c)

Figure 5.5. Exemple d'utilisation des modèles M_I et M_{grad}. En a), un chevauchement de M_I sur 16 pixels d'arêtes de l'iris. En b), un modèle M_{grad} pour vérifier les gradients. En c), les gradients réels obtenus sur l'image d'un oeil, une distribution particulière est obtenue pour l'iris.

Lorsque les groupes d'iris potentiels G_{IG} et G_{ID} sont obtenus avec M_I, un autre modèle M_{grad} de même dimension que M_I est utilisé pour identifer le meilleur iris à l'aide de la distribution des gradients. La plupart du temps, l'iris d'un oeil est plus foncé que le reste de la région de l'oeil. Sur le contour du bas de l'iris sur l'image des gradients I_{grad}, il y a donc de forts gradients pointant vers l'extérieur. Le modèle M_{grad} est donc constitué de gradients normalisés, comme l'illustre la figure 5.5.b), afin d'imiter le bas de l'iris. Un gradient ($g_p.x$, $g_p.y$) vis-à-vis une position $\vec{p}$ de M_{grad} est obtenu grâce à l'équation 5.3.

$$g_p.x = \frac{\overrightarrow{d.x}}{|\vec{d}|} \qquad (5.3)$$

$$g_p.y = \frac{\overrightarrow{d.y}}{|\vec{d}|}$$

où $\vec{d}$ est le vecteur formé entre la position $\vec{p}$ et le centre (x_p,y_p) de M_{grad}. Le meilleur iris de G_{IG} ou G_{ID} (les groupes d'iris potentiels) sera donc celui dont la direction des gradients maximisent la ressemblance entre M_{grad} à l'endroit correspondant dans I_{grad}. Le score de la ressemblance est obtenu grâce à l'équation 5.4. Il s'agit en fait d'une sommation de produits scalaires : des vecteurs avec une direction semblable contribuent donc à augmenter le score. Cette équation est utilisée pour exploiter une propriété intéressante du produit scalaire : la différence d'inclinaison entre 2 vecteurs peut être quantifiée très rapidement, sans même une analyse directe de l'angle entre les deux.

Ceci est très utile lorsque des essais avec M_{grad} sont effectués sur plusieurs positions et que beaucoup de vecteurs sont concernés.

$$score\_grad = \sum_{M_{grad}} \left(g_p.x \cdot G_p.x + g_p.y \cdot G_p.y \right) \tag{5.4}$$

où ($G_p.x$, $G_p.y$) représente le gradient à la position correspondante dans I_{grad} vis-à-vis du gradient à la position (g_p.x, g_p.y) dans M_{grad}.

En résumé, les iris potentiels sont trouvés grâce au nombre de pixels d'arêtes de $I_{\text{arête}}$ chevauchant M_{I}. Ensuite, seul l'iris ayant le meilleur score de l'équation 5.4 est retenu. Ceci permet donc de choisir un iris ayant beaucoup d'arêtes sur son contour du bas tout en ayant de bonnes directions pour les gradients.

Lorsque les iris sont détectés, des erreurs de localisation peuvent être engendrées car le rayon utilisé pour former le modèle M_{I} n'est qu'approximatif. Il est donc difficile de superposer un demi-cercle sur un autre dans l'image lorsque ces derniers ne sont pas de même rayon. Une correction est alors apportée sur la position d'un iris de la façon suivante : un filtre passe-bas est appliqué sur l'image des arêtes $I_{\text{arêtes}}$ afin d'étendre un peu les données et former ainsi un champ d'arêtes, ceci forme l'image $I_{\text{arêtes\_champ}}$. Ensuite, dans un voisinage de quelques pixels sur $I_{\text{arêtes\_champ}}$ (2 pixels de part et d'autre de l'ancienne position de l'iris), des essais sont effectués afin de maximiser le passage d'un demi-cercle sur les valeurs dans $I_{\text{arêtes\_champ}}$. Les nouvelles positions ainsi obtenues sont généralement plus précises.

5.1.4. Estimer le sommet et la base de l'ouverture des yeux

Bien que les iris des deux yeux aient été sélectionnés lors de la première étape, ces iris peuvent en fait être absents puisque l'usager peut très bien avoir les yeux fermés. Cependant, ces iris sont supposés être présents dans l'image pour l'instant afin de pouvoir calculer le sommet et la base de l'ouverture des yeux. Ce n'est qu'à la prochaine étape que les yeux pourront être déterminés comme ouverts ou fermés car certaines données doivent être obtenues préalablement.

Le sommet et la base de l'ouverture sont estimés grâce à la position du centre de l'iris obtenue à l'étape précédente. À partir de cette position, deux régions de recherche R_{haut} et R_{bas} seront parcourues afin de déterminer la position du haut et du bas de l'oeil respectivement. Puisqu'il y a un changement brusque d'intensité lumineuse sur l'image lors du passage de l'intérieur de l'oeil sur la paupière, l'image des gradients I_{grad} sera observée afin de localiser les frontières verticales de l'oeil. Sur ces dernières, de forts gradients sont rencontrés et ceux-ci sont étendus à l'horizontale, relativement à l'inclinaison des deux yeux. Pour le haut et le bas, deux segments de droite S_{haut} et S_{bas} sont positionnés afin de maximiser le passage sur de forts gradients de I_{grad}, c'est-à-dire maximiser les équations 5.5 et 5.6.

$$score\_haut = \int_{S_{haut}} \sqrt{I_{grad}(\vec{p}).dx^2 + I_{grad}(\vec{p}).dy^2} \cdot \vec{dp} \tag{5.5}$$

$$score\_bas = \int_{S_{bas}} \sqrt{I_{grad}(\vec{p}).dx^2 + I_{grad}(\vec{p}).dy^2} \cdot \vec{dp} \tag{5.6}$$

où $\vec{p}$ est le vecteur de position sur I_{grad}. Les segments S_{haut} et S_{bas} sont centrés, selon l'horizontale, vis-à-vis du centre de l'iris (x_p, y_p) et doivent être assez longs pour éviter de se positionner sur les frontières de l'iris ou de la pupille : de forts gradients peuvent aussi être rencontrés vis-à-vis de ces endoits. Le fait d'avoir des segments assez longs encourage le passage sur des gradients distribués horizontalement plutôt que ceux distribués sur des cercles. Mais ces segments doivent aussi être assez courts afin d'être positionnés sur la frontière du contour de l'oeil qui, bien que courbe, a une courbure beaucoup plus faible que celle de l'iris ou de la pupille. De plus, les intervalles de recherche R_{haut} et R_{bas} doivent être définis afin d'être assez grands pour toutes les ouvertures possibles de l'oeil et assez courts pour éviter les autres gradients à l'extérieur de la région de l'oeil : les frontières des sourcils, par exemple, constituent de très forts gradients étendus à l'horizontale et ceux-ci doivent être évités. Les valeurs de ces paramètres sont arbitraires et ont été sélectionnées pour de bons résultats. Celles-ci sont montrées sur l'équation 5.7.

$$L_{haut} = L_{bas} = R \tag{5.7}$$

$$R_{haut} = R_{bas} = 1.5 \cdot R$$

où R, L_{haut} et L_{bas} sont respectivement le rayon de l'iris, la longueur de S_{haut} et la longueur de S_{bas}. Les hauteurs des ouvertures, par rapport au centre (x_p, y_p), sont désignées par h_{haut} et h_{bas} pour le haut et le bas respectivement et sont obtenues lors du positionnement de S_{haut} et S_{bas}. Ces hauteurs se situent donc dans les intervalles définis par l'équation 5.8.

$$0 \leq h_{haut} \leq R_{haut}$$
$$0 \leq h_{bas} \leq R_{bas} \tag{5.8}$$

La figure 5.6 illustre de façon géométrique l'obtention des hauteurs h_{haut} et h_{bas}. Il est à noter que cette détection de l'ouverture est basée sur le fait que l'iris est très souvent situé près du centre de l'oeil. Dans le cas où un usager regarde à l'extrême gauche ou droite, des imprécisions peuvent s'ajouter puisque les frontières de l'oeil détectées par S_{haut} et S_{bas} ne seront plus vis-à-vis du centre de la concavité de l'ouverture. Cependant, l'étendue horizontale des gradients est plus forte vis-à-vis de ce centre, ceci contribue tout de même à positionner S_{haut} et S_{bas} vis-à-vis du centre de la concavité de façon verticale. Sur la plupart des images d'un oeil ouvert, la frontière du bas de l'ouverture de l'oeil se situe vis-à-vis du bas de l'iris. Et le haut de l'iris est partiellement caché par la paupière. C'est un avantage pour la méthode utilisée puisque les gradients situés entre la paupière et l'iris sont plus faibles que ceux entre le blanc de l'oeil et l'iris. Il a d'ailleurs été constaté que la précision obtenue est plus grande lorsque le blanc de l'oeil en haut ou en bas de l'iris n'est pas visible.

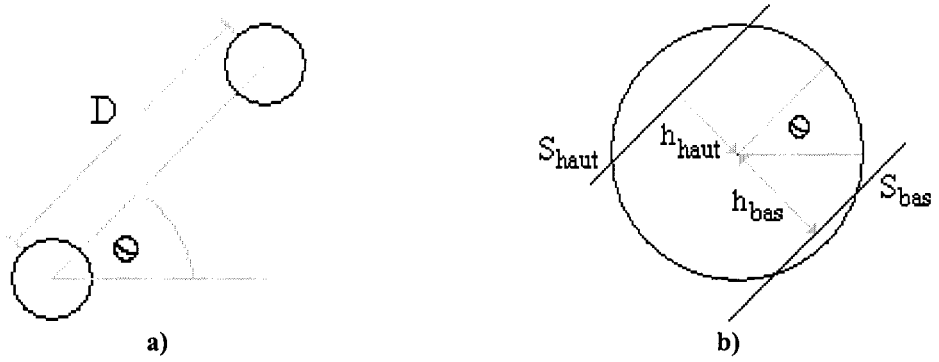


Figure 5.6. L'obtention des hauteurs h_{haut} et h_{bas} pour l'ouverture de l'oeil. En a), l'angle θ est obtenu grâce à l'inclinaison, par rapport à l'horizontale, des iris détectés. En b), l'emplacement des segments S_{haut} et S_{bas} permet d'obtenir h_{haut} et h_{bas} respectivement.

5.1.5. Déterminer si l'oeil est ouvert ou fermé

À partir de la position (x_p, y_p) de l'iris et des hauteurs h_{haut} et h_{bas} de l'ouverture de l'oeil, une méthode a été élaborée afin de déterminer si l'oeil est ouvert ou fermé. Cette étape est importante car selon l'état d'ouverture de l'oeil, le contour ne sera pas adapté de la même façon. Afin de déterminer l'état de l'ouverture, l'iris détecté ainsi que son voisinage sont analysés afin de vérifier s'il s'agit bel et bien d'un iris ou d'un autre objet indésiré. La méthode utilisée est principalement basée sur les caractérisques suivantes pour un oeil ouvert :

1. Sur l'image des vallées $I_{vallées}$, de fortes vallées sont obtenues vis-à-vis de l'iris ;
2. Sur l'image des sommets $I_{sommets}$, de forts sommets sont obtenus vis-à-vis du blanc de l'oeil.

Ces vallées et ces sommets sont donc recherchés pour valider l'ouverture de l'oeil. La figure 5.7 illustre un exemple mettant en évidence ces images pour un oeil ouvert et un oeil fermé. Sur cette figure, les vallées sont clairement visibles sur les iris, les sourcils et le contour des yeux lorsque ces derniers sont ouverts. Et les sommets sont très forts sur le blanc des yeux uniquement, surtout près de l'extérieur des iris. Lorsque les yeux sont fermés, les vallées sont très fortes sur les sourcils et les fermetures des yeux tandis que les sommets sont presque invisibles.

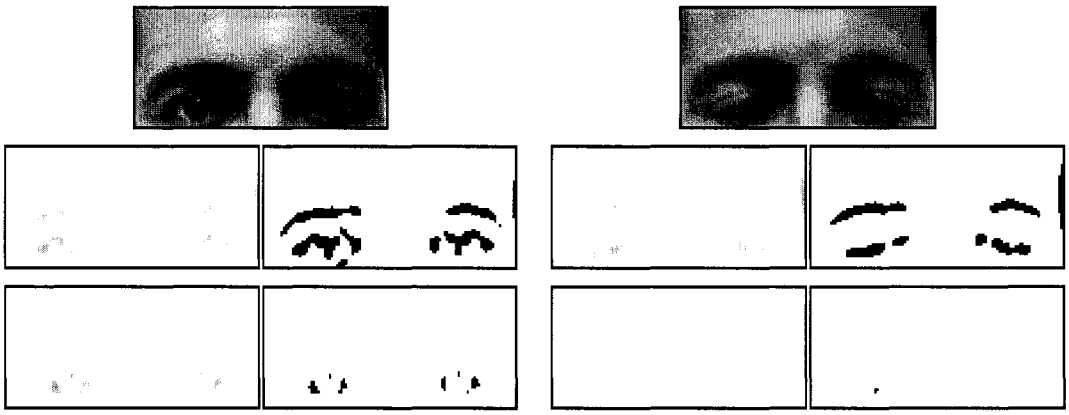


Figure 5.7. Les images des vallées et des sommets pour des yeux ouverts et fermés. En a) ou en b), les images du centre représentent $I_{vallées}$ à gauche et ces valeurs binarisées à droite. Les images du bas représentent $I_{sommets}$ à gauche et ces valeurs binarisées à droite.

La détermination de l'état d'ouverture des yeux est donc effectuée à l'aide des 3 tests suivants :

1. Vérifier s'il y a des vallées vis-à-vis de l'iris ;
2. Vérifier s'il y a des sommets vis-à-vis du blanc de l'œil ;
3. Vérifier si la région de l'iris est assez foncée.

Si ces 3 tests sont positifs, l'oeil est alors considéré ouvert. Et si au moins l'un des tests est négatif, l'oeil est alors considéré fermé. Un autre test pourrait aussi vérifier si l'iris est circulaire mais ceci est en fait effectué pour la détection de l'iris. Étant donné que ce sont soit les sommets forts ou les vallées fortes qui sont recherchés, un seuil est appliqué aux valeurs afin d'obtenir les images binaires $I_{sommets\_bin}$ et $I_{vallées\_bin}$ respectivement. En supposant que les valeurs de $I_{sommets}$ et $I_{vallées}$ varient de 0 à 255, les seuils ont été fixés à 40 et 20 respectivement pour de bons résultats.

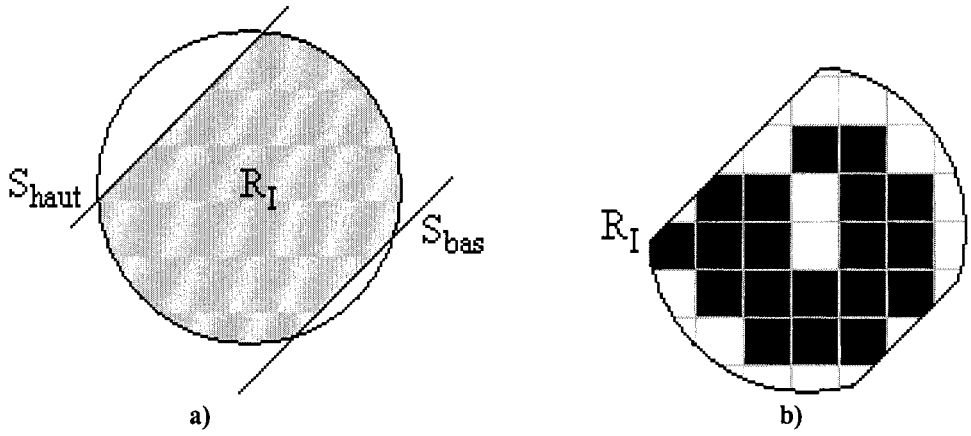


Figure 5.8. L'obtention de la concentration des vallées à l'intérieur de la région visible de l'iris. En a), R_I est la région à l'intérieur du cercle de l'iris entre les segments S_{haut} et S_{bas}. En b), la concentration est obtenue d'après le ratio du nombre de vallées binaires sur le nombre de pixels à l'intérieur de R_I.

5.1.5.1. Vérifier s'il y a des vallées vis-à-vis de l'iris

Ce test consiste à vérifier la présence des vallées binaires à l'intérieur de la région visible de l'iris R_I tel qu'illustré à la figure 5.8. La concentration des vallées binaires à l'intérieur de R_I est utilisée pour ce test comme le montre l'équation 5.9.

$$concentration\_vallées = \frac{\int_{R_I} I_{vallees\_bin}(\vec{p}) \cdot \vec{dp}}{\int_{R_I} \vec{dp}} \tag{5.9}$$

où $\vec{p}$ est un vecteur de position à l'intérieur de R_I. Le test suivant est effectué pour déterminer si l'iris est absent :

Si (*concentration_vallées* < α) alors l'iris absent.

Pour de bons résultats, le seuil α a été fixé à 0.3, ceci signifie que plus du tiers de la région de l'iris doit comporter des vallées. Ce seuil doit être assez élevé pour distinguer l'iris du voisinage tout en étant assez bas pour éviter que des yeux pâles ou

des reflets de lumière empêchent la détection. Il est à noter que plusieurs endroits sur l'image peuvent satisfaire ce test : il suffit d'être situé sur une région concentrée en vallées comme sur la fermeture d'un oeil. D'autres tests doivent donc être utilisés mais ceux-ci permettent d'éviter plusieurs endroits indésirés.

5.1.5.2. Vérifier s'il y a des sommets vis-à-vis du blanc de l'oeil

Ce test consiste à vérifier s'il y a de fortes concentrations de sommets dans le blanc de l'oeil à gauche et à droite de l'iris. Pour ce faire, deux régions R_G et R_D sont définies pour la gauche et la droite respectivement. Il est à noter que l'iris n'est pas nécessairement positionné au centre de l'oeil et c'est pourquoi il est assez difficile de déterminer avec précision les régions R_G et R_D. De plus, lorsqu'un usager regarde à l'extrême droite ou gauche, l'une de ces régions peut être absente même si l'oeil est ouvert. Étant donné que les vallées semblent davantage concentrées près du contour de l'iris, R_G et R_D sont définis selon la figure 5.9. Sur cette figure, ces régions sont constituées de deux cercles de même rayon que l'iris croisant le centre de l'iris (x_p, y_p). Les centres des trois cercles sont colinéaires et l'inclinaison obtenue est celle engendrée par les iris des deux yeux. Le choix de ces régions provient de la simplicité géométrique utilisée pour bien englober les fortes concentrations de sommets. D'après les résultats obtenus, des distributions assez variées sont rencontrées près du blanc des yeux. Il est donc difficile de justifier la géométrie des régions de recherche utilisées.

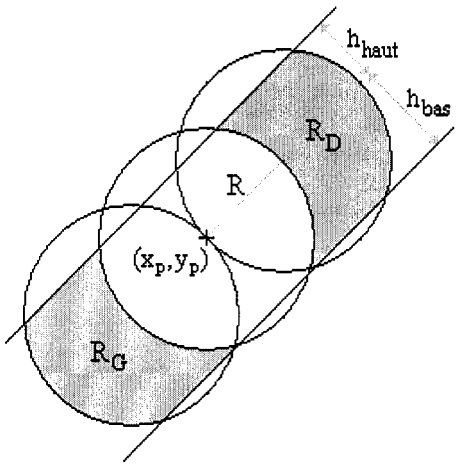


Figure 5.9. Modèle géométrique pour obtenir les régions de recherche R_G et R_D. Ces régions sont utilisées pour récupérer les sommets dans le blanc de l'oeil.

Les équations 5.10 et 5.11 illustrent comment les concentrations des sommets à l'intérieur de R_G et R_D respectivement sont obtenues.

$$concentration\_sommets\_G = \frac{\int_{R_G} I_{sommets\_bin}(\overrightarrow{pG}) \cdot \overrightarrow{dpG}}{\int_{R_G} \overrightarrow{dpG}} \tag{5.10}$$

$$concentration\_sommets\_D = \frac{\int_{R_D} I_{sommets\_bin}(\overrightarrow{pD}) \cdot \overrightarrow{dpD}}{\int_{R_D} \overrightarrow{dpD}} \tag{5.11}$$

où $\overrightarrow{pG}$ et $\overrightarrow{pD}$ sont des vecteurs de positions à l'intérieur de R_G et de R_D respectivement. Puisque R_G ou R_D peut être absent mais jamais les deux en même temps lorsque l'oeil est ouvert, le test suivant est effectué pour déterminer si l'iris est absent :

Si [(*concentration_sommets_G* + *concentration_sommets_D*) < β]
alors l'iris absent.

Pour de bons résultats, le seuil β a été fixé au ratio de la grandeur de l'ouverture de l'oeil sur l'aire engendré par les régions R_G et R_D, ceci revient à avoir au moins assez de pixels de sommet pour la grandeur de l'ouverture. Le blanc des yeux semble le seul endroit dans le voisinage des yeux où de fortes concentrations de sommets peuvent être rencontrées. Cependant, à cause des conditions d'éclairage non contrôlées, des sommets peuvent être perçus à cause des reflets sur la peau. Ces sommets indésirables sont surtout rencontrés sur les coins intérieurs des yeux et sur les paupières comme l'illustre la figure 5.10. Afin d'éviter ce problème, la région de recherche des iris potentiels ainsi que R_G et R_D doivent être assez restreintes pour éviter ces endroits.

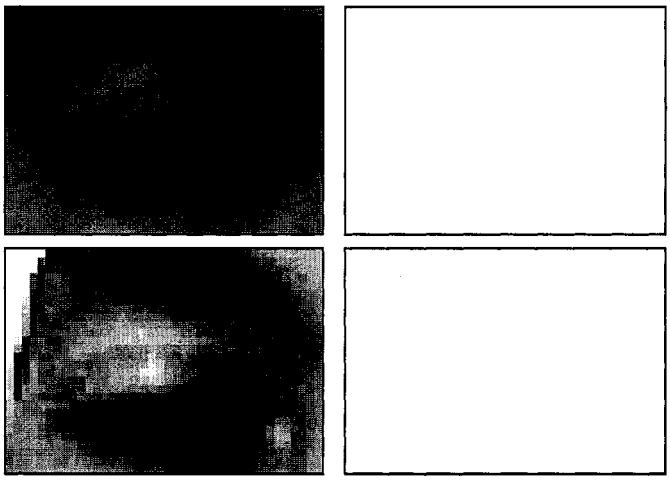


Figure 5.10. Quelques exemples où des sommets indésirés sont rencontrés. Ces sommets sont causés par les reflets de la lumière sur la peau. Les images de gauche sont les sommets obtenus des images de droite.

5.1.5.3. Vérifier si la région de l'iris est assez foncée

Ce test peut paraître semblable au premier mais voici ce qui le distingue : un sommet n'est pas nécessairement sur une région très foncée, il suffit que cette région soit plus foncée que son entourage. Donc pour éviter les sommets situés sur des régions trop pâles, l'intensité lumineuse est vérifiée à l'aide du calcul de la concentration des pixels foncés effectuée par l'équation 5.12.

$$concentration\_foncé = \frac{\int_{R_I} I_n(\vec{p}) \cdot \overrightarrow{dp}}{\int_{R_I} \overrightarrow{dp}} \tag{5.12}$$

où I_n est une image binaire où seulement les pixels foncés sont retenus, R_I est la région de l'iris et $\vec{p}$ est un vecteur de position dans R_I. En utilisant un ton de gris sur 256 valeurs où le noir est obtenu à 0 et le blanc à 255, un pixel est considéré foncé s'il a une valeur inférieure à 100. L'équation 5.12 est identique à 5.9 sauf que I_n est utilisé au lieu de $I_{vallées\_bin}$. Il faut donc obtenir une valeur très faible de l'équation 5.12. Le test suivant est effectué pour déterminer si l'iris est absent :

Si (*concentration_foncé* < γ) alors l'iris est absent.

Pour de bons résultats, le seuil γ a été fixé à 0.5, ceci signifie qu'au moins la moitié des pixels à l'intérieur de l'iris doivent être assez foncés. Un seuil plus élevé serait moins robuste aux reflets de lumière souvent rencontrés. A part l'iris, les endroits très foncés sont surtout rencontrés sur le contour des yeux et les sourcils. Bien que la pupille, qui est un point d'un noir très foncé, soit censée contribuer grandement à ce test, il arrive

souvent que des reflets indésirés viennent perturber l'image. Ces reflets sont souvent près du centre de l'iris et habituellement d'un blanc très vif comme le montre la figure 5.11. Le seuil γ ne peut donc pas être très élevé pour éviter que les yeux ouverts soient considéré fermés. Heureusement, ces reflets couvrent généralement de petites régions et le reste de l'iris permet de compenser leur effet.

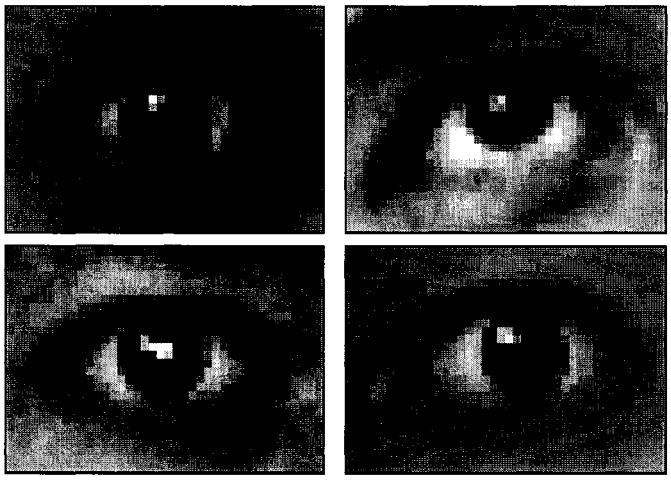


Figure 5.11. Quelques exemples où des reflets d'un blanc très vif sont rencontrés. Ces reflets sont souvent rencontrés près du centre des iris.

Exemples de localisation :

La figure 5.12 illustre des exemples de localisation de l'iris. En a), les iris n'ont pas été détectés car les yeux sont presques fermés, ceci nuit à la détection du cercle de l'iris. En b) et c), les iris sont correctement localisés (les cercles blancs) malgré la présence de reflet de lumière dans l'iris. L'ouverture de l'oeil est aussi détectée (les droites blanches traversant les iris). En d), à cause de la présence de reflet de lumière, un seul iris a été détecté.

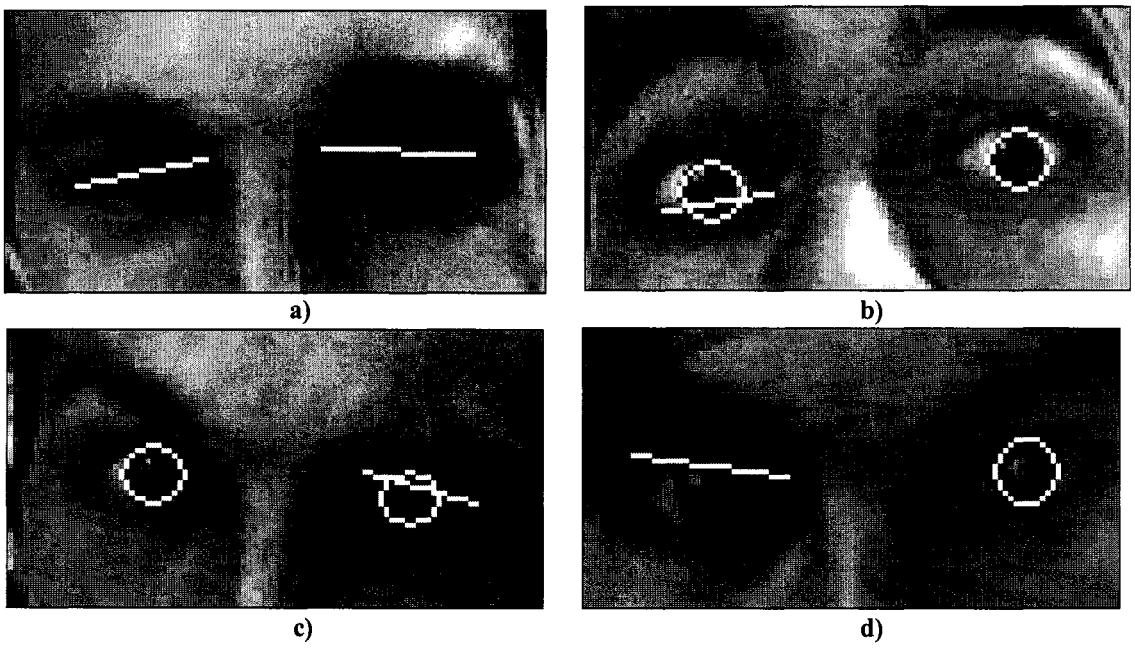


Figure 5.12. Exemples de localisation de l'iris.

5.1.6. Adapter le contour de l'oeil

La dernière étape pour le suivi des yeux consiste à adapter le contour des yeux en fonction des données obtenues aux étapes précédentes : il est donc essentiel que ces données soient obtenues avec précision. Puisque que le contour des yeux peut varier beaucoup d'une image à l'autre et qu'il est parfois difficile à distinguer sur certaines images, la méthode utilisée est principalement basée sur l'anthropométrie du visage afin de définir les modèles des contours devant être adaptés. Ainsi, certaines propriétés géométriques concernant les yeux demeurent presque constantes d'un individu à l'autre et c'est pourquoi certaines d'entre elles seront utilisées.

La méthode proposée est basée sur les caractéristiques suivantes étant d'ailleurs parmi celles mentionnées au début du chapitre :

- Les sommets des concavités des courbes des paupières sont souvent au centre de ces courbes ;
- La largeur des yeux est constante peu importe l'état ;
- Les quatre coins engendrés par les deux yeux sont presque colinéaires ;
- L'espace libre entre les deux yeux est presque de la même largeur qu'un œil ;
- Un oeil peut être ouvert ou fermé ;
- L'ouverture d'un oeil est indépendante de l'ouverture de l'autre ;
- Les deux yeux regardent dans la même direction ;
- Pour un oeil fermé, la courbe du contour est très souvent concave vers le bas.

Selon l'état des yeux, ouvert ou fermé, les trois cas suivants doivent être considérés :

1. Les deux yeux sont ouverts ;
2. Les deux yeux sont fermés ;
3. Un oeil est ouvert et l'autre est fermé.

5.1.6.1. Premier cas : les deux yeux sont ouverts

Pour traiter ce cas, le modèle géométrique représenté sur la figure 5.13 est utilisé. Selon cette figure, la distance entre le centre des deux yeux est estimée à partir de la distance entre les deux iris obtenue auparavant. Une telle estimation peut être effectuée avec beaucoup de précision puisque les deux yeux regardent dans la même direction : même si le centre des yeux n'est pas vis-à-vis de celui des iris à cause du regard, la distance D demeure presque constante. Il y a cependant quelques rares configurations où cela ne s'applique pas (le deux yeux qui regardent vers le centre par exemple).

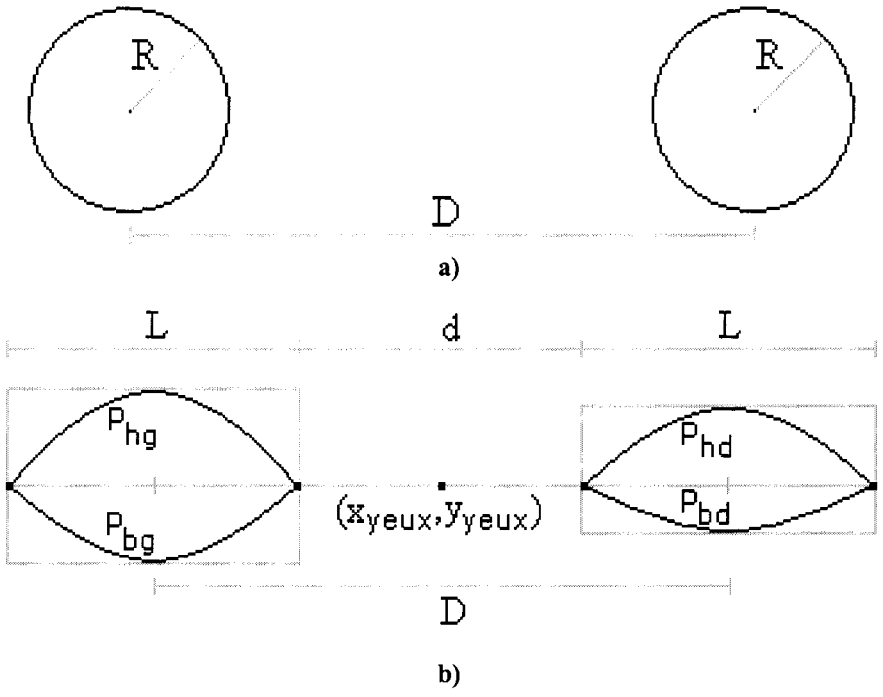


Figure 5.13. Le modèle géométrique pour les deux yeux ouverts. En a), la distance D est obtenue à partir des deux iris. En b), D est utilisée pour estimer la distance entre le centre de deux yeux. La largeur de l'oeil L est estimée en fonction de D.

Le contour d'un oeil est exprimé par deux paraboles dont l'une a sa courbure vers le bas et l'autre, vers le haut. Ces paraboles sont présentées sur la figure 5.13.b). Bien que plusieurs méthodes [50][72] utilisent l'image des arêtes pour placer les contours, une approche différente a été choisie pour ce projet étant donné les difficultés rencontrées pour obtenir les arêtes des yeux : les images utilisées ne sont pas d'une très bonne qualité et les arêtes recherchées sont parfois difficiles à distinguer. L'image des vallées $I_{\text{vallées}}$ sera donc utilisée car les données semblent plus fiables. Ainsi, puisque le contour des yeux contient beaucoup de vallées, les paraboles du modèle devront maximiser le passage sur les fortes vallées de $I_{\text{vallées}}$. Cependant, la région d'un oeil contient beaucoup de vallées sur l'iris, le sourcil, etc. Il faut donc éviter que les paraboles se localisent en ces endroits. Pour ce faire, très peu de liberté sera attribuée aux déplacements de ces paraboles. La figure 5.14 illuste un exemple d'une parabole avec les paramètres nécessaires pour la définir.

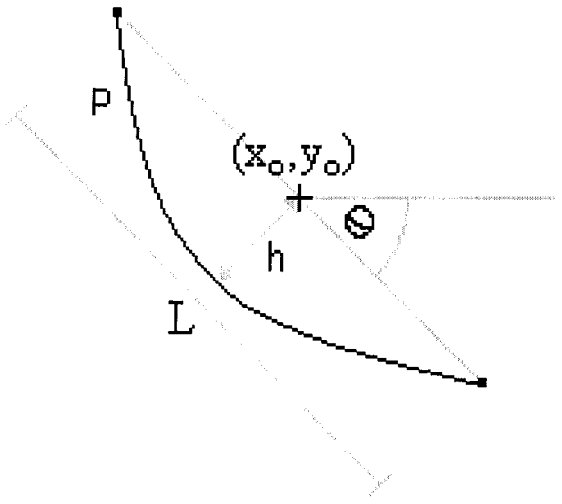


Figure 5.14. Paramètres nécessaires pour définir une parabole P.

Voici les contraintes qui seront imposées pour diminuer la liberté des paramètres :

- La largeur des paraboles des deux yeux demeure constante à L ;
- La position du point de courbure d'une parabole est située sur le segment S_{haut} ou S_{bas} (pour la parabole du haut ou du bas respectivement), ceci permet de définir la valeur h ;
- Les deux paraboles d'un oeil se croisent vis-à-vis des coins de l'œil, donc la position (x_0, y_0) est partagée ;
- La distance d séparant les paraboles de gauche et de droite demeure constante ;
- Les paraboles des deux yeux ont toutes la même orientation constante θ définie selon l'orientation obtenue par les deux iris ;

D'après ces contraintes, les paramètres des paraboles seront obtenus seulement en déplaçant le centre (x_{yeux},y_{yeux}) situé entre les deux yeux. Ceci positionne les coins des yeux uniquement. Puisque L, d et θ sont constants, les quatre coins des yeux se positionneront en même temps que (x_{yeux},y_{yeux}) et la recherche des fortes vallées se fera sur les deux yeux en même temps. La maximisation de l'équation 5.13 est donc recherchée en tenant compte du fait que les vallées du bas de l'oeil sont moins visibles. Pour ce faire, des essais de (x_{yeux},y_{yeux}) sont effectués pour chaque pixel dans un interval RxR. Les paramètres des paraboles sont obtenus ensuite grâce à l'équation 5.14.

$$concentration\_vallées = conc\_hg + \frac{1}{2}conc\_bg + conc\_hd + \frac{1}{2}conc\_bd \qquad (5.13)$$

$$\text{Où} \qquad conc\_hg = \int_{P_{hg}} I_{vallées}(\vec{p})d\vec{p}$$

$$conc\_bg = \int_{P_{bg}} I_{vallées}(\vec{p})d\vec{p}$$

$$conc\_hd = \int_{P_{hd}} I_{vallées}(\vec{p})d\vec{p}$$

$$conc\_bd = \int_{P_{bd}} I_{vallées}(\vec{p})d\vec{p}$$

$\vec{p}$: vecteur de position sur P_{hg}, P_{bg}, P_{hd} ou P_b (5.14)

Il y a des avantages et des inconvénients d'avoir si peu de liberté concernant les paramètres des paraboles :

Inconvénients :

1. Un oeil qui serait bien localisé seul peut être mal localisé si l'image de l'autre oeil est mauvaise ;
2. Les largeurs L et d ne sont qu'approximatives et peuvent en fait varier d'un individu à l'autre ;
3. Des imprécisions sont rencontrées si les quatre coins des yeux ne sont pas colinéaires
4. Si l'usager a les deux pupilles regardant au centre, les yeux seront estimés comme étant plus petits à cause de D ;
5. Si la tête de l'usager a subi une grande rotation autour de l'axe vertical, l'anthropométrie s'applique mal sur l'image obtenue ;
6. Si seulement l'un des deux iris est mal positionné, les contours seront imprécis pour les deux yeux.

Avantages :

1. Le contour d'un oeil ayant une image mauvaise peut tout de même être bien localisé si l'image de l'autre oeil est de bonne qualité ;
2. L'anthropométrie des yeux est respectée ;
3. Le temps de calcul est considérablement diminué ;
4. Les probabilités sont plus faibles de se positionner sur des vallées indésirables.

Les avantages obtenus sont plus importants que les inconvénients et c'est pourquoi cette méthode a été retenue. Le fait que le contour d'un oeil soit aussi positionné en fonction de l'autre est à la fois un inconvénient et un avantage : quelques erreurs peuvent parfois être obtenues mais la plupart du temps, cela rend le système plus robuste en empêchant qu'une configuration étrange soit obtenue par l'ensemble des deux yeux. L'équation 5.13 peut en fait être exprimée à l'aide de deux variables seulement, soit x_{yeux} et y_{yeux}. Il en résulte donc un gain énorme en temps de calcul comparativement à une autre méthode utilisant tous les paramètres nécessaires pour traiter les yeux séparément. Pour chaque oeil, ces paramètres supplémentaires seraient : la largeur de l'oeil, l'orientation et la position du centre de l'oeil par rapport au centre de l'iris. Puisque l'intégration d'une parabole sur $I_{\text{vallées}}$ est un traitement assez lent pour un ordinateur, il est préférable qu'un nombre minimal de paramètres soit utilisé afin de n'avoir que très peu d'essais à effectuer. Certaines méthodes, comme celles utilisant les masques déformables [72], utilisent des algorithmes de minimisation numérique telle la descente du gradient. Il a cependant été constaté que ces méthodes s'appliquent mal à ce projet

car des convergences sur de mauvais minimums (ou maximums) étaient trop souvent rencontrées. Ce type de méthode numérique nécessite certaines conditions initiales pour que les paramètres à ajuster soient situés assez près des bonnes valeurs pour que la convergence soit bien atteinte. Dans ce projet, rien ne peut garantir cette convergence et plusieurs essais expérimentaux ont démontré qu'elle était rarement atteinte. Avec la méthode utilisée dans ce projet, le maximum recherché est en échange toujours atteint puisque toutes les valeurs dans les intervalles alloués pour les paramètres sont essayées. Ceci serait pratiquement impossible si le nombre de paramètres était moindrement élevé.

Lorsque le modèle géométrique est positionné, le contour des yeux est obtenu et grâce aux informatons obtenues aux étapes précédentes, les informations suivantes sont recueillies pour définir entièrement un oeil :

- La position (x_o,y_o) du centre de l'œil ;
- La hauteur de l'ouverture du haut : $H_{haut} = h_c + h_{haut}$;
- La hauteur de l'ouverture du bas : $H_{bas} = h_{bas} - h_c$;
- La largeur de l'oeil L ;
- L'inclinaison de l'oeil θ ;
- Le centre de l'iris (x_p,y_p) ;
- Le rayon de l'iris R.

où h_c représente la distance verticale entre (x_o,y_o) et (x_p,y_p), ceci est défini par l'équation 5.15. La figure 5.15 illustre la représentation géométrique de ces informations

(sauf pour θ). La figure 5.16 est un exemple où le modèle géométrique du contour seulement a été adapté sur l'image I_t grâce aux données de $I_{vallées}$. La figure 5.17 illustre des cas où la localisation a mal été effectuée. Ces cas se produisent souvent lorsque l'image de l'œil est floue ou que l'œil est presque fermé.

$$h_c = \vec{u}.x \cdot \overrightarrow{op}.y - \vec{u}.y \cdot \overrightarrow{op}.x \tag{5.15}$$

où $\quad \vec{u} = (\cos\theta, \sin\theta)$

$$\overrightarrow{op} = (x_p - x_o, y_p - y_o)$$

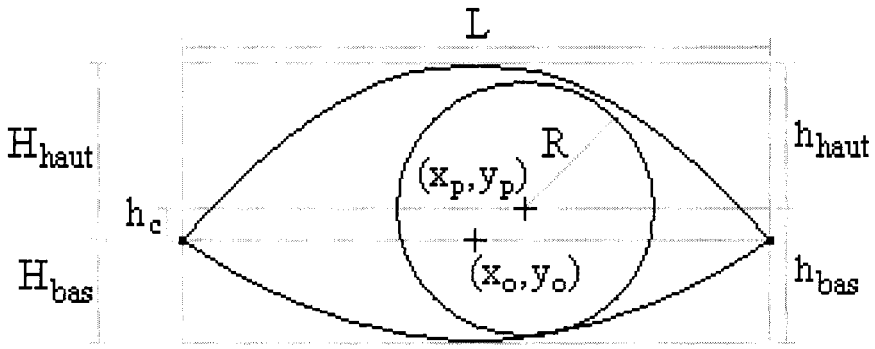


Figure 5.15. La représentation géométrique des informations pour un oeil. Seule l'inclinaison θ n'est pas illustrée.

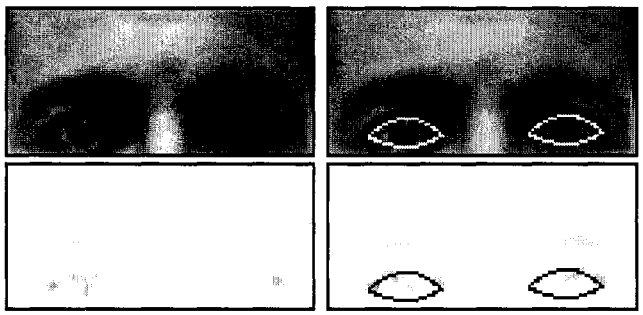
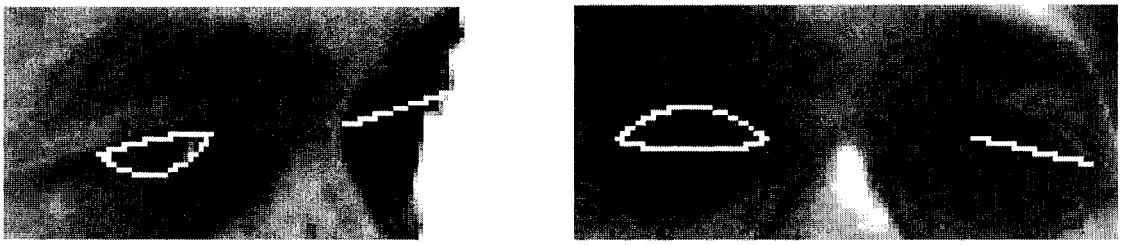


Figure 5.16. Exemple d'adaptation du modèle géométrique du contour de l'oeil. À gauche, les images I_t et $I_{vallées}$. À droite, le modèle obtenu grâce aux données sur $I_{vallées}$.



a) b)

Figure 5.17. Des cas rencontrés où la localisation a mal été effectuée.

5.1.6.2. Deuxième cas : les deux yeux sont fermés

Lorsque les yeux sont fermés, un autre modèle géométrique est utilisé. Dans ce modèle, chaque oeil n'est représenté que par une parabole puisque le haut et le bas de l'ouverture se chevauchent. Tout comme pour le cas des yeux ouverts, l'image des vallées $I_{\text{vallées}}$ est utilisée danc ce projet car les valeurs obtenues sont très fiables comparativement aux autres observées (celles des arêtes par exemple). Cependant, les deux yeux seront traités séparément puisqu'il est difficile de pouvoir construire un modèle aussi simple que celui du cas des deux yeux ouverts. Cette difficulté provient du fait que les iris aient été définis absents et que la position de ces derniers soit donc invalide. Sans la position des iris, les données θ et D ne peuvent donc pas être obtenues aussi facilement. Pour les obtenir, les fermetures des deux yeux devront d'abord être obtenues. Voici maintenant quelques caractéristiques qui permettront de faciliter l'adaptation des modèles pour les fermetures :

1. Entre deux images consécutives I_{t-1} et I_t, les valeurs D, L, θ et les positions des yeux varient peu ;

2. De fortes vallées sont rencontrées dans $I_{\text{vallées}}$ vis-à-vis une fermeture et très peu de vallées indésirables sont présentes dans le voisinage (à part le sourcil qui est assez éloigné).

Grâce à la première caractéristique, les valeurs nécessaires dans l'image courante au temps t seront estimées grâce à celles obtenues au temps t-1. Ainsi, les valeurs suivantes seront définies initialement:

$L_0 = L$

$\theta_0 = \theta$

$D_0 = D$

$(x_{0o}, y_{0o}) = (x_o, y_o)$

R_f : région de recherche pour la position (x_o, y_o) de la fermeture

Puisque les deux fermetures sont traitées séparément, l'adaptation d'une parabole sera expliquée pour un oeil seulement. À partir de la position (x_o, y_o) de l'oeil (ouvert ou fermé) au temps t-1, une région de recherche R_f centrée en (x_o, y_o) est définie afin de trouver la nouvelle position de l'oeil fermé au temps t. R_f doit être assez grande pour couvrir les mouvements rigides de la tête et les imprécisions obtenues. Toutefois, R_f doit être assez petite pour éviter les fortes vallées indésirables. Pour de bons résultats, les dimensions de R_f ont été fixées à $(2 \cdot R) \times (2 \cdot R)$. Ensuite, à l'intérieur de R_f, plusieurs paraboles sont utilisées afin de trouver celle qui maximisera le passage sur de fortes vallées de $I_{\text{vallées}}$. Il s'agit donc de maximiser l'équation 5.16 en essayant un intervalle de possibilités pour le centre (x_o, y_o), l'inclinaison θ_f et la hauteur h_{bas}.

$$concentration\_vallées = \int_{P_f} I_{vallées}(\vec{p})d\vec{p} \tag{5.16}$$

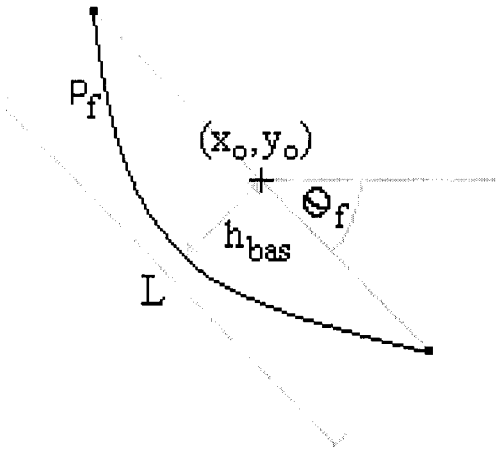


Figure 5.18. Les paramètres nécessaires pour construire une parabole P_f.

où $\vec{p}$ est un vecteur de position dans $I_{\text{vallées}}$ et P_f est la parabole de la fermeture. La figure 5.18 illustre un exemple de la parabole utilisée avec les paramètres nécessaires. La largeur L est fixée à L_0 puisque la largeur de l'oeil varie très peu d'une image à l'autre.

Voici maintenant les intervalles définis pour faire varier les paramètres :

$$x_{0o} - R \le x_o \le x_{0o} + R$$

$$y_{0o} - R \le y_o \le y_{0o} + R$$

$$\theta_0 - \Delta\theta \le \theta_f \le \theta_0 + \Delta\theta$$

$$h_{min} \le h_{bas} \le 0$$

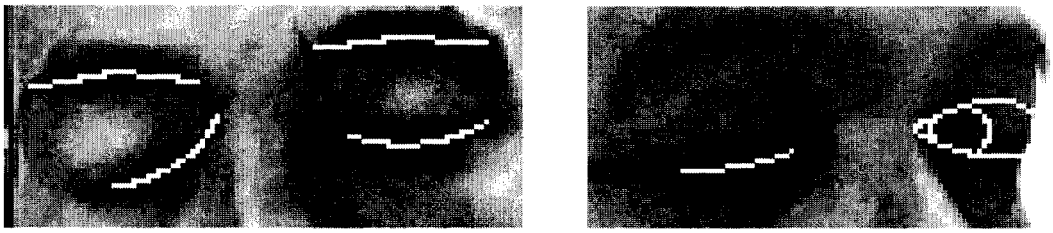
Et pour de bons résultats, les valeurs de $\Delta\theta$ et h_{min} ont été fixées à $0.03 \cdot \pi$ et $-0.1 \cdot L$ respectivement. Le choix de $\Delta\theta$ permet de compenser les rotations de la tête autour de l'axe de la profondeur entre les temps t et t-1. Le fait que la borne supérieure

de h_{bas} soit zéro empêche la parabole d'être courbée vers le haut et cela pour les deux raisons suivantes :

1. La courbure de la fermeture de l'oeil est presque toujours vers le bas ;
2. Le sourcil, qui est une forte concentration de vallées, est très souvent courbé vers le haut.

Grâce à l'intervalle admis pour h_{bas}, les paraboles inutiles sont évitées et les probabilités d'une mauvaise localisation sur le sourcil sont faibles. Lorsque les deux paraboles des fermetures sont localisées, la distance D et l'orientation θ entre les deux yeux peuvent être recalculées pour ensuite continuer le suivi à l'image suivante. La figure 5.19 illustre un exemple où les paraboles des fermetures ont été adaptées sur l'image I_t grâce aux données de $I_{vallées}$. Bien que les yeux ne soient pas complètement fermés sur cette figure, les iris sont presque invisibles et ont donc été considéré absents. La figure 5.20 illustre des cas où la détection a été mal effectuée. Ces cas se produisent souvent lorsque le voisinage de la fermeture de l'œil est foncé sur une grande surface (en a) ou lorsqu'un iris est détecté même s'il est absent (en b).

Figure 5.19. Exemple de l'adaptation des paraboles pour la fermeture des yeux. À gauche, les images I_t et $I_{vallées}$. À droite, les fermetures obtenues grâce aux données sur $I_{vallées}$.



a) **b)**

Figure 5.20. Des cas où la localisation a été mal effectuée

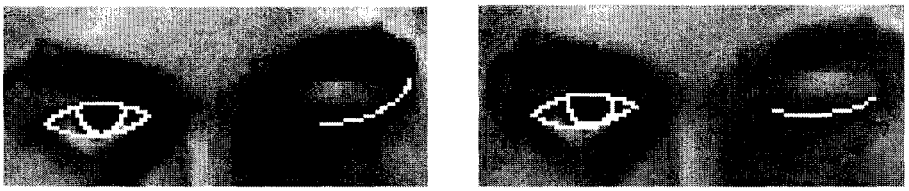
5.1.6.3. Troisième cas : un oeil est ouvert et l'autre est fermé

Pour ce troisième cas, un mélange des deux premiers cas sera utilisé. Pour l'oeil fermé, l'adaptation de la fermeture s'effectue de la même manière puisqu'au deuxième cas, les yeux sont traités séparément. Cependant, une modification doit être apportée pour l'oeil ouvert puisqu'il ne peut plus être traité en même temps que l'autre oeil. Pour ce cas, l'oeil ouvert est adapté en ignorant l'autre oeil. L'équation 5.17 doit donc être

maximisée pour positionner l'oeil ouvert. Cette maximisation est effectuée avec des essais sur un intervalle de possibilités comme expliqué pour le cas des deux yeux ouverts.

$$concentration\_vallées = \int_{P_h} I_{vallées}(\overrightarrow{p_h})d\overrightarrow{p_h} + \frac{1}{2}\int_{P_b} I_{vallées}(\overrightarrow{p_b})d\overrightarrow{p_b} \qquad (5.17)$$

où P_h, P_b, $\overrightarrow{p_h}$ et $\overrightarrow{p_b}$ sont respectivement la parabole du haut, la parabole du bas, le vecteur de position sur P_h et le vecteur de position sur P_b. Pour positionner un oeil seul, les paramètres θ et L sont estimés aux valeurs obtenues à l'image précédente. Lorsque les deux yeux sont positionnés, ces paramètres peuvent être recalculés convenablement pour respecter les proportions anthropométriques expliquées précédemment. Puisque l'oeil ouvert est traité seul, les avantages de l'anthropométrie sont peu utilisés et les mauvaises détections sont donc plus probables. Mais pour l'oeil fermé, il n'y a aucune différence. Bien qu'il soit assez rare qu'un usager puisse faire un clin d'oeil durant la séquence d'images, ce troisième cas est tout de même fréquent lorsqu'un des deux yeux est mal positionné : une fermeture est donc recherchée car l'iris est jugé absent. Il est à noter que le système de correction fonctionne aussi bien pour le cas d'un seul oeil fermé. La figure 5.21 en illustre un exemple : en a), la fermeture de l'œil droit est mal positionnée mais est rétablie à l'image suivante en b).



a) b)

Figure 5.21. Correction de la localisation dans le cas où il y a un œil ouvert et l'autre fermé.

5.1.7. Système de correction d'erreurs

Puisque ce projet doit traiter des séquences d'images, il est très important qu'une mauvaise détection sur une image n'affecte pas automatiquement tout le reste de la séquence. Heureusement, la méthode utilisée dans ce projet pour localiser la fermeture des deux yeux est aussi un système de correction d'erreurs. Ainsi, si la position des yeux (ouverts ou fermés) s'avère imprécise sur une image, le système essaiera souvent de diminuer cette imprécision d'une image à l'autre. Les lignes qui suivent en décrivent les raisons.

Lorsque la position des yeux est très imprécise, les iris sont rarement détectés et les deux fermetures sont donc recherchées même si les yeux sont ouverts. Les paraboles des fermetures essaieront donc, d'image en image, de se déplacer aux endroits qui maximiseront le passage sur de fortes vallées dans $I_{\text{vallées}}$. Mais sur chaque image, ces déplacements ne pourront pas dépasser les régions de recherche R_f. Puisque de fortes vallées sont rencontrées à la fois sur un oeil ouvert ou fermé, les fermetures sont donc attirées sur les yeux. Et si une fermeture se positionne vis-à-vis un oeil ouvert, alors la position sera assez souvent précise pour que la détection de l'iris soit effectuée convenablement sur l'image suivante. Cependant, ce système de correction a les faiblesses suivantes :

- L'imprécision à corriger doit être assez petite pour que la parabole ne soit pas attirée sur d'autres concentrations de vallées (les cheveux ou les sourcils par exemple) car cela pourrait même augmenter l'imprécision ;

- Lorsqu'un oeil est ouvert, la distribution des vallées fait en sorte que la parabole puisse rester sur le coin extérieur sans essayer de mieux se positionner ;
- Une suite de plusieurs images est parfois nécessaire avant que la correction soit effectuée.

La meilleure situation où le système de correction fonctionne est donc lorsque l'imprécision est petite et que l'usager a les yeux fermés. La pire situation est probablement le cas où les paraboles sont localisées sur les sourcils : il est alors pratiquement impossible de s'en dégager. La figure 5.22 illustre quelques exemples de la correction d'erreurs. En a), la correction est effectuée entre les images 90 et 168. La mauvaise localisation de l'oeil gauche est causée par la présence de fortes vallées sur les cheveux. En b), la correction est effectuée entre les images 1 et 5. L'oeil droit était mal positionné au départ mais a tout de même été corrigé après quelques images. En c), la correction est effectuée entre les images 79 et 120. Les deux yeux étaient très mal positionnés, la correction a été effectuée mais pas complètement car les iris sont imprécis. En d), la correction est effectuée entre les images 103 et 114. Comme le démontre cette figure, la mauvaise localisation d'un oeil affecte souvent l'autre.

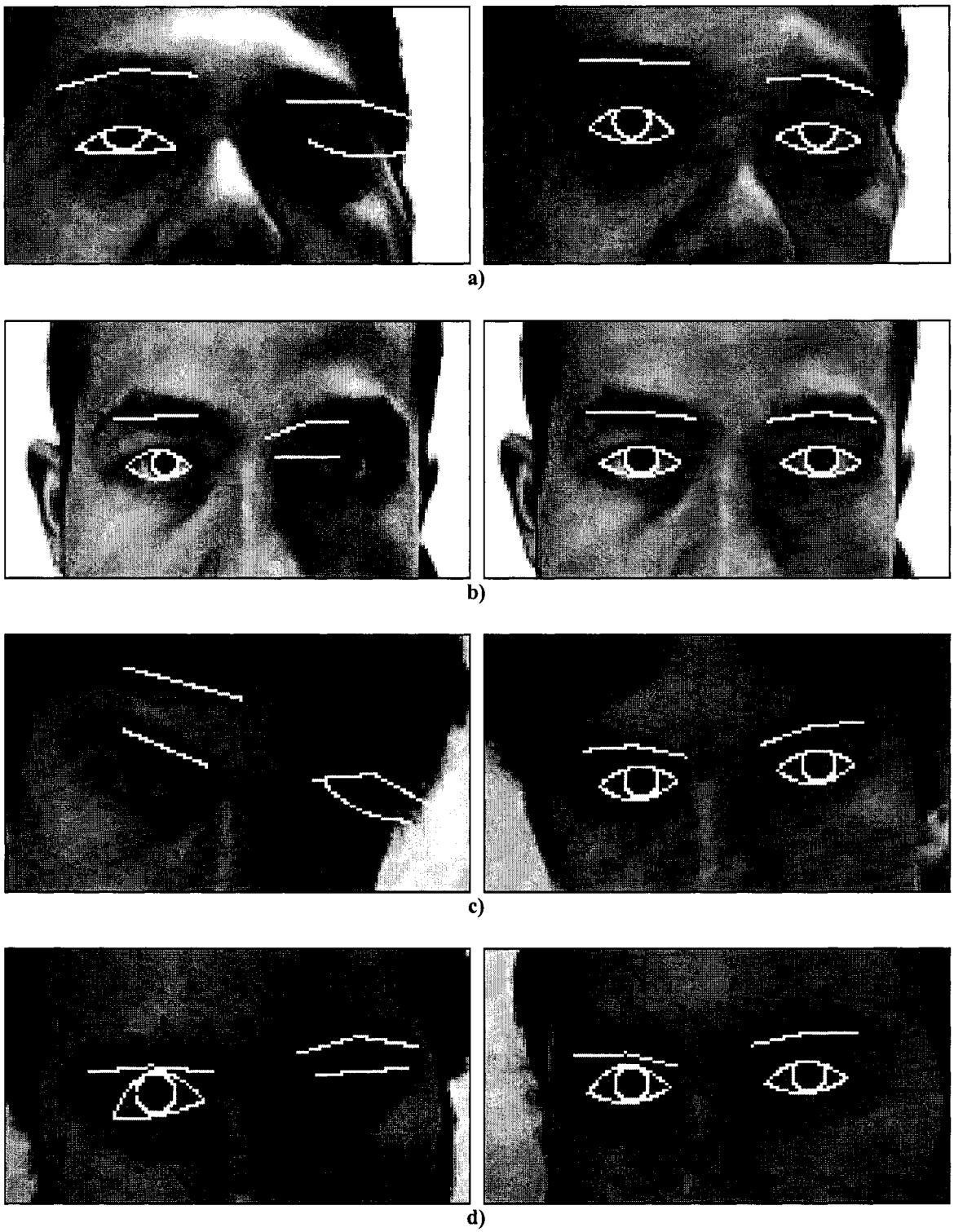


Figure 5.22. Exemples de la correction d'erreurs.

5.1.8. Analyse des résultats obtenus

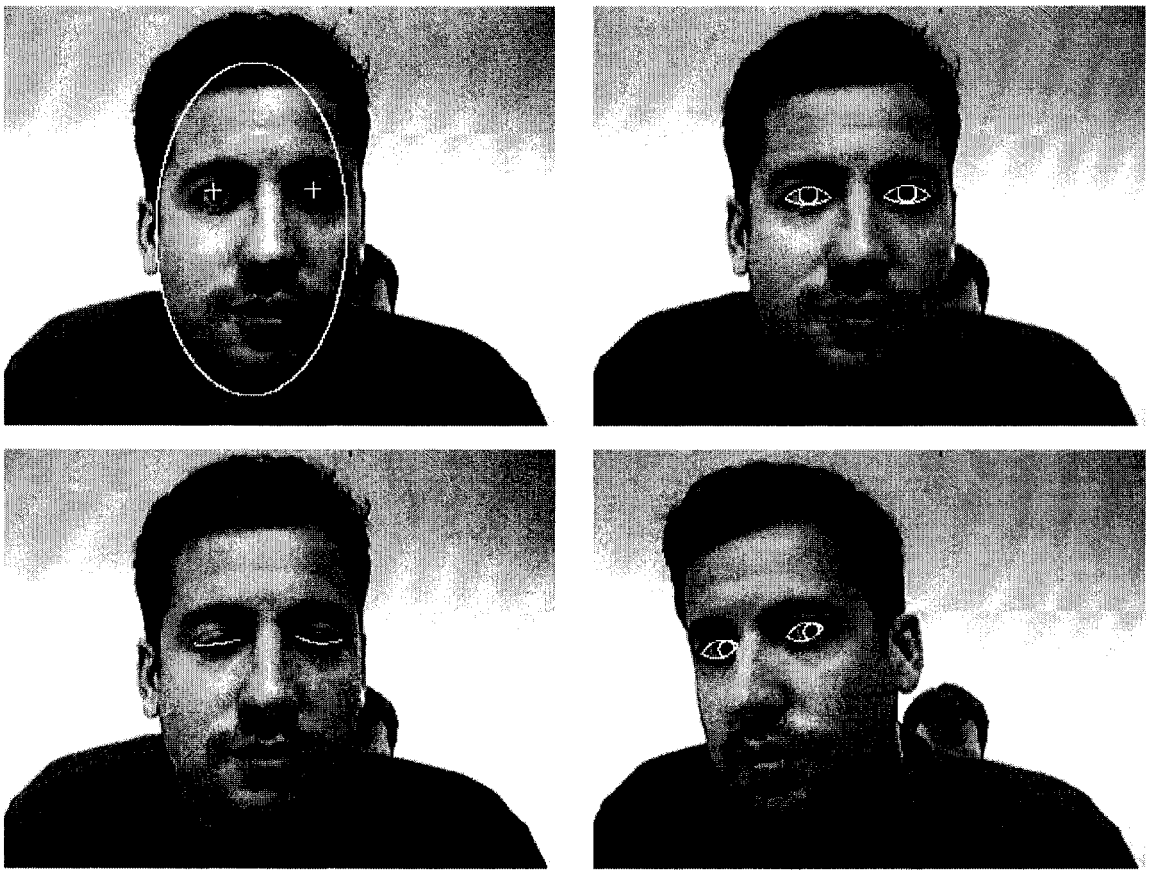


Figure 5.23. Exemple du suivi des yeux à partir de la détection initiale.

La figure 5.23 illustre un exemple où le suivi des yeux a été effectué avec la position initiale obtenue selon les travaux de [84]. Même si les positions sont initialement peu précises, le suivi s'effectue bien. Il serait important d'apporter des améliorations à [84] afin d'obtenir des positions initiales plus précises car pour certaines séquences, le suivi doit être effectué avec des positions initiales sélectionnées manuellement. De façon générale, le suivi fonctionne bien lorsque la détection de l'iris et celle de l'état de l'ouverture sont bien effectuées. Parfois un oeil ouvert est considéré

fermé à cause d'une mauvaise qualité de l'image vis-à-vis de l'iris. Ce dernier est donc détecté absent. Un oeil fermé est cependant rarement considéré ouvert. Quelques déformations sont quelques fois rencontrées lorsque la tête de l'usager est trop tournée selon l'axe vertical. Une telle configuration respecte moins l'anthropométrie du visage, les modèles géométriques peuvent donc difficilement s'y adapter. Cependant, les détections sont généralement bien effectuées. Lorsque l'usager a la tête penchée vers l'avant et qu'un sourcil se situe trop près de l'oeil, il arrive que le haut de l'ouverture se situe sur ce sourcil car de forts gradients sont présents. Il en résulte que l'oeil est trop ouvert mais le reste du modèle n'est généralement pas affecté. De plus, l'ouverture d'un oeil est parfois mal estimée lorsqu'un oeil est trop ouvert et que le blanc de l'oeil est visible en haut ou en bas de l'iris. Il s'agit cependant d'une configuration assez rare. Des résultats plus détaillés sont présentés en Annexe III.

5.2. Le suivi des sourcils

La méthode utilisée pour le suivi des sourcils est inspirée de [10] où un modèle de sourcil doit s'adapter en maximisant un score. Ce score augmente lorsque le modèle est situé vis-à-vis des pixels de l'image sur de fortes vallées. Cependant, plusieurs modifications ont été apportées, le modèle géométrique utilisé est très semblable mais l'adaptation s'effectue autrement. Cette adaptation est effectuée en fonction des données du centre des yeux obtenu auparavant, ceci permet d'effectuer très peu de recherches. De plus, la dimension des modèles est calculée en fonction de la distance entre le centre des deux yeux. Dans ce projet, les sourcils sont recherchés afin de pouvoir estimer les mouvements non rigides du visage. Étant donné que les mouvements des sourcils sont très limités, seulement quelques données peuvent être recueillies et il n'est donc pas nécessaire de prélever le contour complet des sourcils. Voici les mesures jugées suffisantes pour ce projet :

1. La hauteur du centre du sourcil ;
2. L'inclinaison de la moitié gauche du sourcil ;
3. L'inclinaison de la moitié droite du sourcil.

Les tableaux 5.4, 5.5 et 5.6 indiquent quelques caractéristiques concernant les images des sourcils par rapport à l'orientation du visage.

Caractéristiques sur l'intensité des pixels	
1	La couleur des sourcils est généralement plus foncée que celle de la peau
2	La couleur est assez uniforme sur tout le sourcil

Tableau 5.4. Caractéristiques sur l'intensité des pixels.

Caractéristiques sur le mouvement	
1	Le mouvement des sourcils n'est que vertical
2	Le mouvement des sourcils est assez lent

Tableau 5.5. Caractéristiques sur le mouvement.

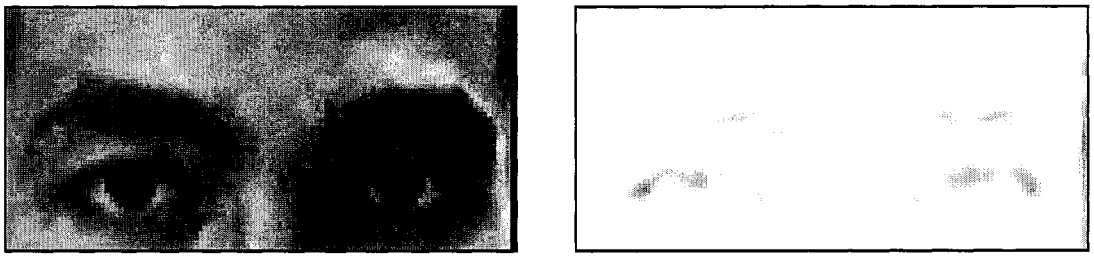
Caractéristiques sur la géométrie	
1	L'étendue du sourcil est surtout horizontale
2	L'étendue du sourcil forme une courbe simple
3	Les sourcils sont toujours en haut des yeux
4	La courbe du sourcil est très souvent convexe vers le haut
5	La courbe du sourcil est très rarement convexe vers le bas
6	La courbure du sourcil est assez faible
7	Le sourcil est généralement plus large que l'oeil
8	Selon l'horizontale, le sourcil inclut l'oeil au complet généralement
9	Le sourcil n'est pas d'une épaisseur uniforme
10	Selon l'horizontale, le centre du sourcil est près du centre de l'oeil

Tableau 5.6. Caractéristiques sur la géométrie.

5.2.1. Description générale de la méthode utilisée

Les premières caractéristiques de chaque tableau cité précédemment ont permis d'élaborer rapidement une méthode simple et efficace pour la détection des sourcils. En obtenant l'image des vallées $I_{\text{vallées}}$ (voir Annexe V) à partir de l'image originale des sourcils, de grandes étendues sont obtenues vis-à-vis de ces derniers comme le démontre la figure 5.24. Ce sont donc ces étendues qui seront recherchées grâce aux données obtenues à partir des yeux et des autres caractéristiques élaborées à partir des sourcils. La méthode organisée est effectuée selon les 4 étapes suivantes :

1. Récupérer les données déjà acquises sur les yeux
2. Obtenir l'image des vallées $I_{\text{vallées}}$
3. Positionner un segment de droite S_c sur le centre du sourcil
4. Positionner deux segments de droite S_g et S_d sur le sourcil



a) b)

Figure 5.24. Exemple d'images des sourcils. En a), l'image originale. En b), l'image des vallées $I_{\text{vallées}}$.

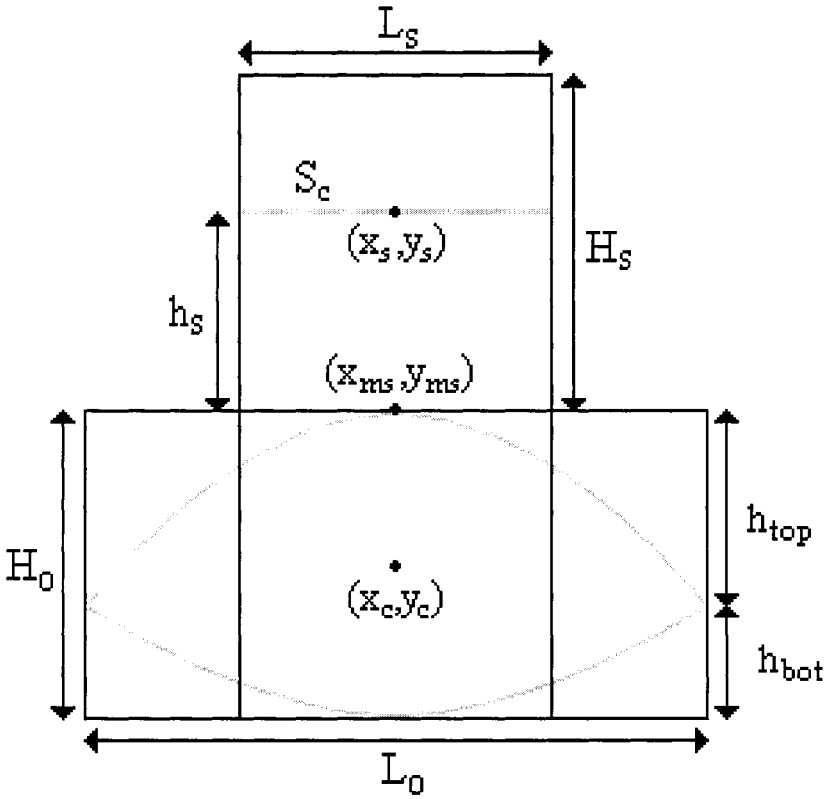


Figure 5.25. Mesures utilisées pour trouver la position du centre du sourcil.

5.2.2. Obtention des données sur les yeux

Les données essentielles sur les yeux sont : la hauteur H_o, la largeur L_o, le centre (x_c,y_c) et l'inclinaison des yeux θ_{yeux}. Selon la figure 5.25, la recherche du centre du sourcil S_c doit débuter à l'extérieur de l'oeil au point (x_{ms},y_{ms}) défini par l'équation 5.18.

$$x_{ms} = x_c - \frac{H_o}{2}\cdot\sin(\theta_{yeux}) \qquad (5.18)$$

$$y_{ms} = y_c + \frac{H_o}{2}\cdot\cos(\theta_{yeux})$$

Il est très important de définir ce point de recherche correctement car si ce dernier est trop éloigné, le sourcil ne sera pas inclus et s'il est trop près, une partie de l'oeil sera incluse et la détection de S_c sera alors instable. En effet, la détection est basée sur l'image des vallées et l'oeil en contient beaucoup; c'est pourquoi celui-ci doit être exclu de la région de recherche.

La largeur de S_c, nommée L_s ainsi que la hauteur de la région de recherche H_s sont ensuite définies selon l'équation 5.19.

$$\begin{aligned} L_s &= \alpha_1 \cdot L_o \\ H_s &= \alpha_2 \cdot L_o \end{aligned} \qquad (5.19)$$

Et α_1 et α_2 représentent respectivement le diamètre de l'iris et la distance séparant le centre des yeux. Bien que ces choix soient très arbitraires, les résultats observés semblent amplement suffisants.

5.2.3. Positionnement d'un segment de droite sur le centre du sourcil

D'après la figure 5.25, localiser un segment de droite S_c revient à trouver h_s tel que $0 \leq h_s \leq H_s$. En utilisant l'image des vallées $I_{vallées}$, h_s est choisi en maximisant le score de l'équation 5.20, c'est-à-dire en maximisant le passage de S_c sur les fortes vallées de $I_{vallées}$ tout en étant situé à une distance probable du centre de l'oeil. Cette maximisation est effectuée avec des essais sur un intervalle de positions possibles pour S_c.

$$Score\_S_c = P(d) \int_{Sc} I_v(\vec{p}) \cdot d\vec{p} \tag{5.20}$$

où $d = \sqrt{(x_c - x_s)^2 + (y_c - y_s)^2}$

avec $\vec{p}$: un vecteur de position parcourant S_c

$P(\cdot)$: fonction de probabilité

La fonction de probabilité $P(\cdot)$ est utilisée afin d'éviter que S_c se localise vis-à-vis de fortes vallées indésirables. Cela peut arriver par exemple lorsque la tête tourne selon l'axe vertical et que l'un des sourcils devienne de moins en moins visible. Sur l'image courante I_t, $P(\cdot)$ tend à favoriser les distances étant le plus près possible de celle obtenue sur l'image précédente I_{t-1}, comme le montre l'équation 5.21.

$$P(d) = \frac{e^{\frac{-(d-d_{-1})^2}{2\sigma^2}}}{\sqrt{2\pi \cdot \sigma^2}} \tag{5.21}$$

où d : distance sur I_t

d_{-1} : distance obtenue sur I_{t-1}

σ : déviation standard

Dans ce projet, σ a été fixé au tiers de la distance séparant le centre des yeux pour de bons résultats, ceci permet d'assurer que les sourcils soient assez près des yeux. Une valeur trop élevée de σ rend $P(\cdot)$ inefficace tandis qu'une valeur trop faible rend la méthode très peu robuste aux mouvements rapides des sourcils.

5.2.4. Positionner deux segments de droite sur le sourcil

Étant donné que le sourcil est une courbe faiblement courbée, deux segments de droite seulement seront utilisés pour le définir complètement. Ces segments sont nommés S_g et S_d pour la gauche et la droite respectivement. La longueur W_s d'un segment (S_g ou S_d) est définie par l'équation 5.22.

$$W_s = \alpha_3 \cdot D \tag{5.22}$$

où D est la distance séparant le centre des yeux et α_3 est fixé à 0.3 pour respecter la proportion autant que possible. Il est à noter que la largeur réelle des 2 moitiés du sourcil n'est pas importante puisque que ce n'est que le centre et les inclinaisons de S_g et S_d qui sont recherchées. Il suffit donc d'avoir un α_3 permettant de couvrir suffisamment le sourcil. S_g et S_d sont d'ailleurs connectés sur le point (x_s,y_s), c'est-à-dire sur le centre de S_c tel qu'illustré sur la figure 5.26. Dans ce projet, le centre de l'oeil (x_c,y_c) et le centre du sourcil (x_s,y_s) sont considérés comme étant vis-à-vis l'un de l'autre selon la verticale. Cette supposition n'est pas exacte mais cela permet de simplifier la recherche tout en ayant de très bons résultats. Avec la position (x_s,y_s) obtenue, il suffit ensuite de

trouver les bonnes inclinaisons θ_g et θ_d pour S_g et S_d respectivement. Puisqu'un sourcil a une courbure assez faible et très rarement convexe vers le bas, les valeurs de θ_g et θ_d doivent respecter les contraintes suivantes :

$$\theta_{yeux} + \pi \leq \theta_g \leq \theta_{yeux} + \Delta\theta + \pi \tag{5.23}$$

$$\theta_{yeux} - \Delta\theta \leq \theta_d \leq \theta_{yeux}$$

où $\Delta\theta$ a été fixé à $0.8 \cdot \pi$ pour de bons résultats, ceci laisse peu de liberté d'inclinaison aux sourcils. Tout comme pour S_c, la maximisation du passage de S_g et S_d sur de fortes vallées de $I_{vallées}$ est utilisée pour le positionnement de ces segments. La maximisation des équations 5.24 et 5.25 est effectuée en parcourant plusieurs inclinaisons possibles dans un intervalle pour chaque segment. Les inclinaisons recherchées sont θ_g et θ_d.

$$Score\_S_g = \int_{Sg} I_v(\overrightarrow{p_g}) \cdot \overrightarrow{dp}_g \tag{5.24}$$

$$Score\_S_d = \int_{Sd} I_v(\overrightarrow{p_d}) \cdot \overrightarrow{dp}_d \tag{5.25}$$

Où $\vec{p}_g$ et $\vec{p}_d$ sont les vecteurs bidimensionnels parcourant S_g et S_d respectivement. La figure 5.27 illustre brièvement les 4 étapes mentionnées pour la localisation des sourcil

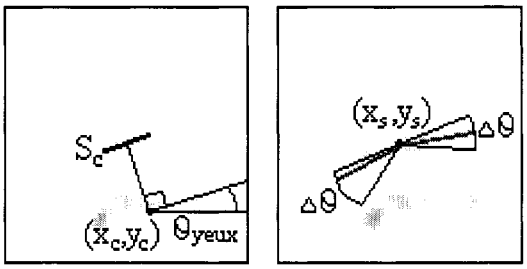


Figure 5.26. Positionnement des segments de droite S_g et S_d. En a), S_c est localisé selon son centre et son inclinaison. En b), à partir du centre de S_c, S_g et S_d sont positionnés.

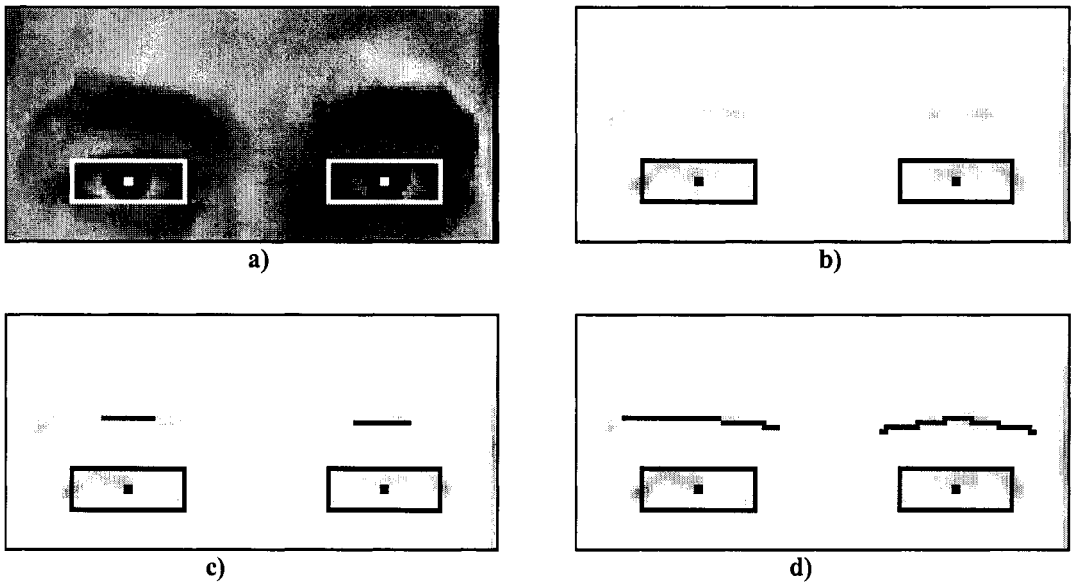


Figure 5.27. Les étapes pour trouver les sourcils à l'aide de l'image des vallées. En a), contour des yeux déjà acquis. En b), la hauteur, la largeur et le centre des yeux. En c), les petits segments S_c positionnés au centre des sourcils. En d), la détection obtenue pour les sourcils.

5.2.5. Analyse des résultats obtenus

La figure 5.28 illustre des résultats obtenus sur une séquence à partir de la première image jusqu'à la 555^e. Bien que quelques imprécisions soient apparues durant la séquence (tels les sourcils mal inclinés en b), il n'y a pas d'accumulation d'erreurs et de bons résultats sont obtenus même après plusieurs centaines d'images. La

figure 5.29 illustre quelques exemples où les sourcils sont mal localisés pour diverses raisons. En a), l'oeil droit est mal localisé selon l'horizontale, le segment Sc est donc vis-à-vis des vallées rencontrées sur les cheveux. En b), la tête a subi une grande rotation selon l'axe vertical et le sourcil droit n'est donc plus vis-à-vis de l'oeil. En c), à la suite d'une grande rotation de la tête selon l'axe vertical sur les images précédentes, le sourcil s'est mal positionné et est maintenant sur une position jugé probable, et S_c s'est placé sur de fortes vallées rencontrées sur les cheveux. En d), l'oeil droit est mal positionné et S_c s'est positionné sur les plus fortes vallées rencontrées sur la fermeture de l'oeil. Ces images démontrent donc l'importance d'avoir une bonne estimation des yeux avant de positionner les sourcils. Plus de détails sont présentés en Annexe III.

Figure 5.28. Exemple du suivi des sourcils sur une séquence d'images. En a), la 1ère image. En b), la 70e image. En c), le 303e image. En d), la 555e image.

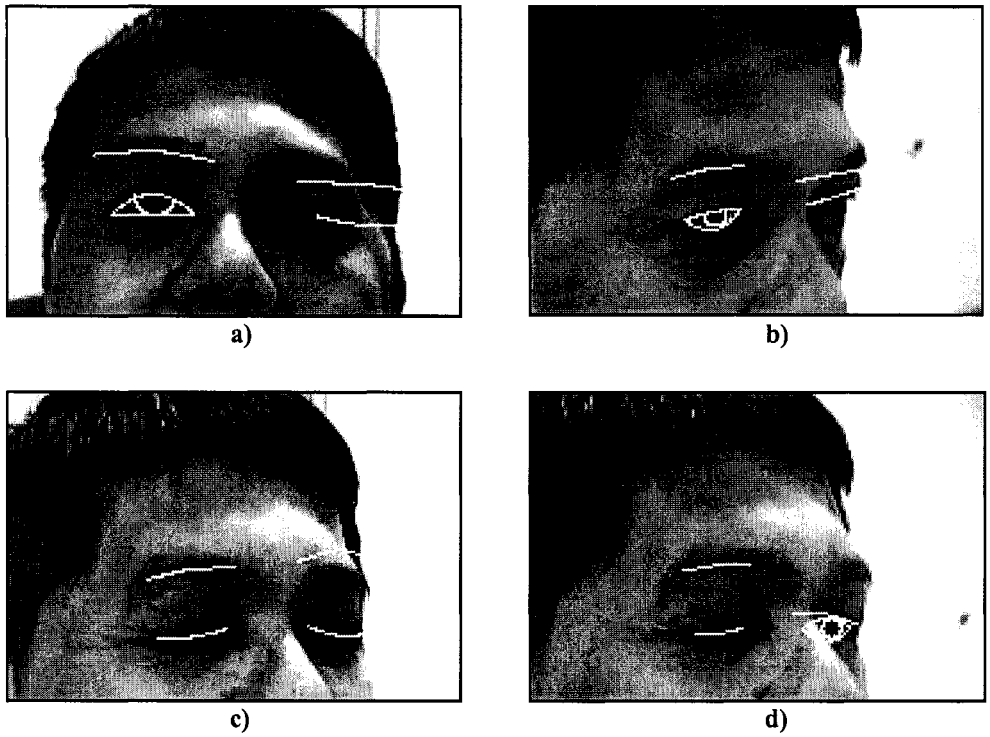


Figure 5.29. Quelques exemples où les sourcils sont mal localisés.

5.3. Le suivi de la bouche

Le suivi de la bouche consiste à pouvoir déterminer le contour de l'ouverture de la bouche d'une image à l'autre. Pour localiser la bouche, voici les informations recherchées jugées suffisantes :

- La position des deux coins extérieurs de la bouche ;
- La position du centre des contours extérieurs des lèvres ;
- La position du centre des contours intérieurs des lèvres.

Pour mieux justifier la méthode utilisée, les tableaux 5.7, 5.8 et 5.9 indiquent quelques caractéristiques de la bouche.

Caractéristiques sur la géométrie	
1	La bouche est plus large que le nez sauf dans quelques rares positions
2	Les coins de la bouche se situent aux extrémités horizontales
3	La géométrie des dents du haut est très différente de celle du bas
4	Les dents du haut sont souvent plus visibles que celles du bas
5	Si les dents du haut sont visibles, elles sont immédiatement en bas de la lèvre supérieure
6	Si les dents du bas sont visibles, elles sont immédiatement en haut de la lèvre inférieure
7	Les dents du haut sont souvent perçues presque de la même hauteur. Il en est de même pour les dents du bas
8	La courbe du contour de la lèvre supérieure est souvent plus complexe que celle de la lèvre inférieure
9	La bouche est presque toujours symétrique selon l'axe vertical au centre du visage

10	La bouche au repos peut rarement être plus large que la distance entre le centre des 2 yeux
11	La hauteur de l'ouverture de la bouche est rarement plus grande que la largeur
12	Pour chaque lèvre (l'inférieure et la supérieure), les terminaisons aux coins de la bouche forment des angles aigus

Tableau 5.7. Caractéristiques selon la géométrie.

Caractéristiques sur l'intensité des pixels	
1	La bouche peut être peu visible lorsqu'elle est entourée d'une moustache ou d'une barbe
2	Les lèvres supérieure et inférieure sont d'une couleur semblable
3	Quelquefois, un changement brusque d'intensité est obtenu entre les lèvres et la peau
4	Un changement brusque d'intensité est obtenu entre les lèvres et l'intérieur de la bouche
5	L'image de l'intérieur de la bouche est souvent très foncée sauf pour les dents qui ont une image très pâle
6	Les écarts entre les dents, selon l'orientation de la tête, forment des contours foncés presque verticaux
7	La couleur est uniforme sur les lèvres mais l'éclairage peut facilement augmenter l'intensité à certains endroits
8	Le relief de la peau est plus complexe en haut de la lèvre supérieure qu'en bas de la lèvre inférieure
9	Lorsque la bouche est fermée, la séparation des 2 lèvres est souvent très visible, très foncée et elle relie les 2 coins de la bouche avec une courbe assez simple
10	Pour chaque lèvre (l'inférieure et la supérieure), le contour situé près de l'intérieur de la bouche est souvent plus visible que celui situé près de l'extérieur

Tableau 5.8. Caractéristiques sur l'intensité des pixels.

Caractéristiques sur le mouvement	
1	Lorsque la bouche est en mouvement, plusieurs parties du visage (les joues par exemple) sont habituellement aussi en mouvement
2	Les deux lèvres peuvent parfois se chevaucher, il y en a alors une des deux qui est invisible entièrement ou en partie

Tableau 5.9. Caractéristiques sur le mouvement.

En connaissant les coins et le centre des contours, les courbes des lèvres peuvent ainsi être estimées à l'aide de paraboles.

5.3.1. Description générale de la méthode utilisée

Pour effectuer le suivi de la bouche, le contour devra d'abord être déterminé sur la première image I_0 afin de diminuer les recherches sur les images suivantes. Pour ce faire, on suppose que la bouche est initialement fermée sur I_0 et en position de repos afin de faciliter les traitements. La méthode utilisée pour le suivi sera donc divisée selon les deux étapes suivantes :

1. Initialisation d'un modèle de la bouche sur l'image initiale ;
2. Réajustement du modèle sur l'image courante.

Pour localiser la bouche, deux modèles géométriques sont utilisés pour les cas d'une bouche ouverte ou fermée. Les figures 5.30 et 5.31 illustrent ces modèles géométriques ainsi que les mesures qui seront recherchées.

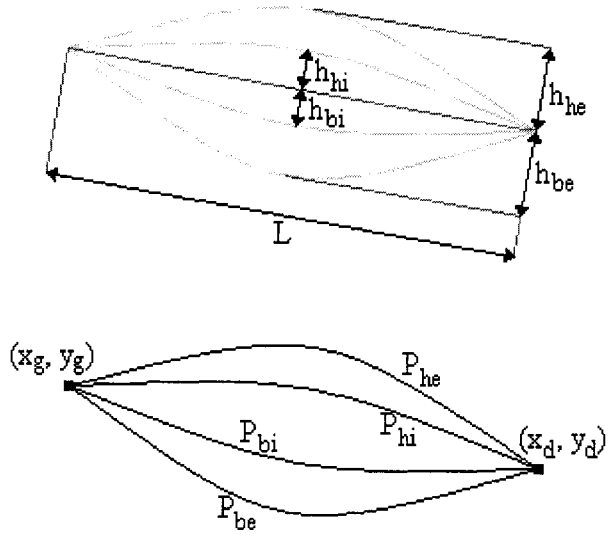


Figure 5.30. Modèle géométrique pour une bouche ouverte.

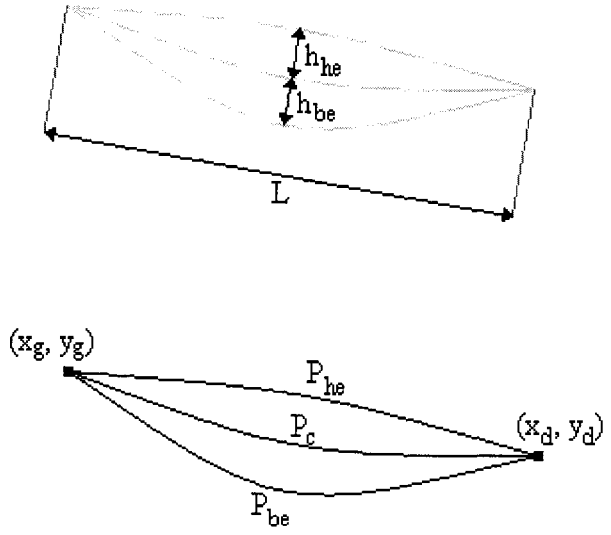


Figure 5.31. Modèle géométrique pour une bouche fermée.

5.3.2. Initialisation d'un modèle de la bouche sur l'image initiale

Sur la première image de la séquence, l'usager doit être d'une position frontale par rapport à la caméra et avec la bouche fermée au repos. La fermeture des deux lèvres forme ainsi une étendue foncée selon une direction horizontale. Certaines méthodes [63][73][78] utilisent l'extraction de la couleur rouge pour localiser la bouche. Pour isoler le rouge, la base HSV est parfois utilisée [78] pour les couleurs et un intervalle est défini sur H pour délimiter le rouge. Il est cependant difficile de pouvoir définir le bon intervalle de couleur pour traiter tous les usagers possibles. Il faudrait un intervalle permettant de ne retenir que la bouche et ignorer la peau autour. Ceci est très délicat car, chez plusieurs individus, les lèvres sont pratiquement de la même couleur que la peau. De plus, les lèvres ne sont pas toujours d'une couleur uniforme étant donné les reflets de l'éclairage. La figure 5.32 illustre quelques exemples où une tentative d'extraction des lèvres a été effectuée avec la base HSV et l'équation 5.26. Selon cette équation, une couleur a un score se situant entre 0 ou 1 : plus une couleur est près du rouge idéal, plus le score est élevé. Il est à noter que sur la figure 5.32, les images présentées sont en tons de gris mais les résultats ont en fait été obtenus à l'aide d'images couleurs dans le format RGB.

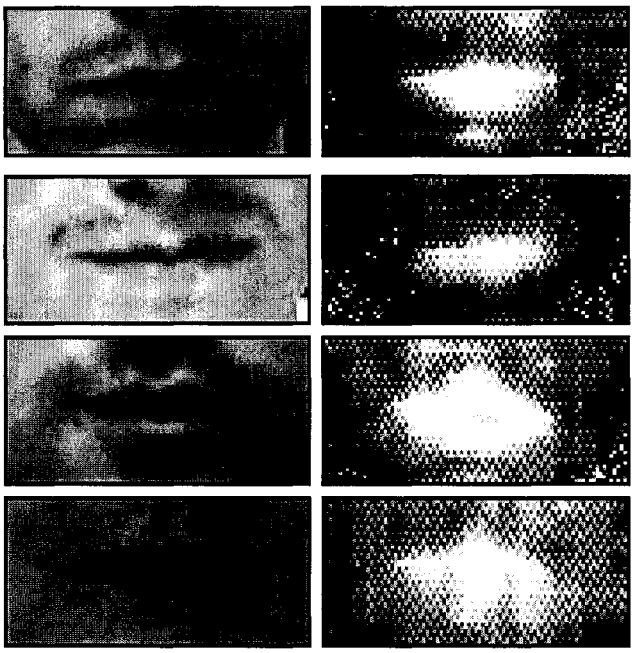


Figure 5.32. Quelques tentatives d'extration des lèvres en utilisant la couleur. La base de couleur HSV est utilisée avec l'équation 5.26.

$$\begin{aligned} &\text{Si } (H > 180°) \\ &\qquad H = -(360-H) \\ &\\ &\text{Si } (|H-H_0| > w_o) \\ &\qquad \text{Score\_rouge} = 0 \\ &\text{Sinon} \\ &\qquad \text{Score\_rouge} = 1 - (H-H_0)^2/w_0^2 \end{aligned} \tag{5.26}$$

Les paramètres H_0 et w_0 ont été fixés à 0.333 et 60 respectivement pour obtenir les meilleurs résultats possibles. D'après les résultats expérimentaux obtenus, l'approche décrite semble inappropriée pour ce projet. L'information de la couleur ne sera donc pas utilisée et une autre approche basée sur les caractéristiques géométriques de la bouche a été élaborée. D'autres techniques [72][77] utilisent des masques déformables mais des essais sur de telles techniques ont souvent conduit à de mauvais résultats. Tout comme les essais similaires effectués pour les yeux, des convergences sur des minimums locaux

étaient obtenues alors que les minimums globaux étaient recherchés. Une nouvelle approche a donc été élaborée pour détecter la bouche en utilisant les mesures déjà obtenues pour les yeux. Tout d'abord, les coins extérieurs de la bouche sont recherchés car ces derniers semblent les éléments de la bouche les plus faciles à localiser puisqu'ils sont très riches en textures. Ces coins seront recherchés afin de connaître les endroits où se croiseront les courbes du contour de la bouche. Il est à noter que le contour extérieur de la bouche est parfois peu visible étant donné que le changement d'intensité entre la peau et les lèvres s'effectue parfois de façon graduelle. Voici donc les étapes nécessaires pour initialiser le modèle :

1. Définir deux fenêtres de recherche pour les coins de la bouche ;
2. Localiser les coins extérieurs de la bouche ;
3. Localiser la fermeture de la bouche ;
4. Localiser les contours extérieurs de la bouche ;
5. Recueillir les informations nécessaires au suivi.

5.3.2.1. Fenêtres de recherche pour les coins de la bouche

Afin de faciliter la détection des coins extérieurs de la bouche, deux fenêtres R_g et R_d seront définies pour limiter les recherches. Pour définir ces fenêtres, le centre de la bouche doit d'abord être estimé. L'équation 5.27 décrit comment le centre de la bouche C_b est estimé grâce à l'anthropométrie du visage. Afin d'augmenter la précision de ce centre, une fenêtre de recherche R_{med} sera définie autour de C_b afin de récupérer la position médiane C_{med} des vallées binaires dans $I_{vallées\_bin}$.

$$C_b.x = \frac{oeil\_G.x + oeil\_D.x}{2} + Ly \cdot \sin(ang\_yeux)$$
$$C_b.y = \frac{oeil\_G.y + oeil\_D.y}{2} - Ly \cdot \cos(ang\_yeux) \quad (5.27)$$

Les symboles *oeil_G* et *oeil_D* sont les positions des centres de l'oeil gauche et droit respectivement et *ang_yeux* est l'angle formé entre les deux yeux et l'horizontale. Dist_yeux est la distance entre le centre des deux yeux et l'équation 5.28 décrit comment le coefficient *Ly* a été obtenu en considérant l'anthropométrie du visage.

$$Ly = \frac{2}{9} \cdot Dist\_yeux \quad (5.28)$$

À partir de l'estimation C_b, les limites de la fenêtre R_{med} sont obtenues grâce aux équations 5.29 et 5.30. La fenêtre R_{med} est rectangulaire et délimitée par le point inférieur gauche ($R_{med}.x1$, $R_{med}.y1$) et le point supérieur droit ($R_{med}.x2$, $R_{med}.y2$). La position médiane C_{med} sera ensuite obtenue en calculant la valeur médiane sur les vallées binaires dans $I_{vallées\_bin}$ à l'intérieur de R_{med}. Pour obtenir les deux coordonnées de la position, le traitement est effectué selon l'horizontale et la verticale séparément. La figure 5.33 illustre comment C_b et R_{med} sont obtenus géométriquement, la figure 5.34 explique à l'aide d'histogrammes comment obtenir C_{med} et la figure 5.35 illustre comment obtenir les fenêtres de recherche R_g et R_d pour rechercher ultérieurement les coins de la bouche.

$$
\begin{aligned}
R_{med}.x1 &= C_b.x - \frac{Sx}{2} \\
R_{med}.y1 &= C_b.y - \frac{Sy}{2} \\
R_{med}.x2 &= C_b.x + \frac{Sx}{2} \\
R_{med}.y2 &= C_b.y + \frac{Sy}{2}
\end{aligned}
\tag{5.29}
$$

$$
\begin{aligned}
Sx &= \frac{4}{5} \cdot Dist\_yeux \\
Sy &= \frac{1}{2} \cdot Dist\_yeux
\end{aligned}
\tag{5.30}
$$

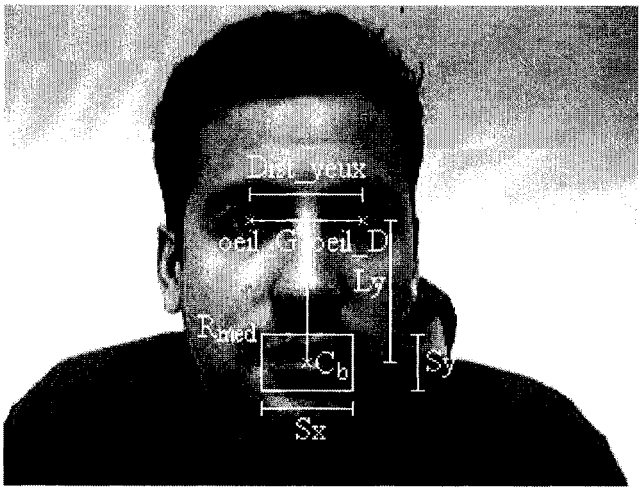


Figure 5.33. Géométrie utilisée à l'aide de l'anthropométrie du visage pour obtenir C_b et R_{med}.

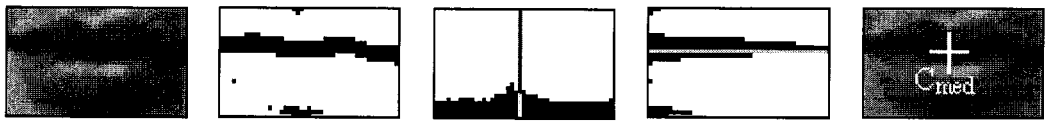


Figure 5.34. Démonstration de l'obtention de C_{med} à l'aide d'histogrammes. En a), l'image intiale I_0 à l'intérieur de la fenêtre R_{med}. En b), l'image $I_{vallées\_bin}$ à l'intérieur de R_{med}. En c), la médiane effectuée sur l'histogramme de la projection horizontale. En d), la médiane effectuée sur l'histogramme de la projection verticale. En e), la position C_{med} obtenue à l'aide des deux médianes.

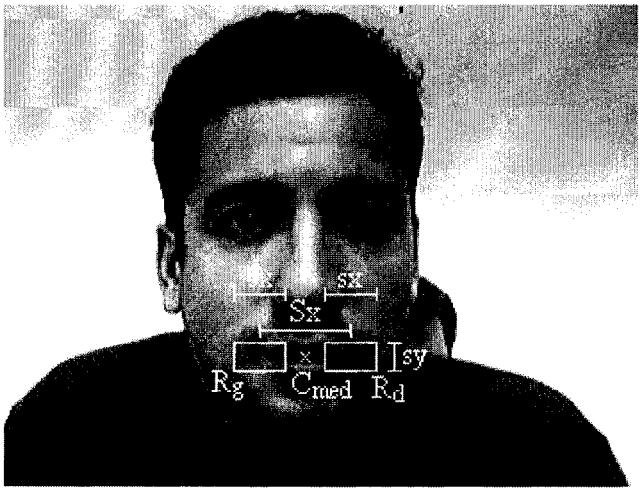


Figure 5.35. Démonstration pour obtenir les fenêtres de recherche R_g et R_d

5.3.2.2. Localisation des coins extérieurs de la bouche

Les coins extérieurs de la bouche sont recherchés à l'intérieur des fenêtres de recherche R_g et R_d pour les coins gauche et droit respectivement. Ces coins seront nommés C_g et C_d. Voici quelques caractéristiques qui définissent les coins extérieurs de la bouche :

- C_g et C_d sont situés respectivement à l'extrême gauche et droite de la bouche ;
- C_g et C_d sont situés sur des vallées ;
- Des distributions particulières des gradients de l'image I_t sont obtenues dans les régions de C_g et C_d.

L'image des vallées est donc utilisée pour localiser les coins. Puisque les coins sont situés sur des vallées mais que l'amplitude de ces dernières n'a pas beaucoup d'importance, l'image des vallées binaires $I_{vallées\_bin}$ sera utilisée où le seuil de binarisation a été fixé à 20 pour de bons résultats en sachant que l'intervalle de $I_{vallées}$ va

de 0 à 255. Ce seuil permet d'isoler uniquement la fermeture de la bouche. À partir de $I_{vallées\_bin}$, il faut ensuite éliminer certaines positions des coins afin de ne garder qu'un groupe de coins potentiels. La première élimination consiste à retirer toutes les positions où les vallées binaires sont absentes. Ensuite, il faut éliminer celles qui ne sont pas succeptibles d'être à l'extrême gauche ou droite pour C_g ou C_d respectivement. Pour ce faire, le voisinage immédiat de chaque pixel de vallée binaire est analysé afin de vérifier si d'autres vallées n'empêchent pas le pixel concerné d'être à l'une des extrêmes. De plus, les petites vallées binaires causées par le bruit sont éliminées. Voici donc comment un coin potentiel est recherché :

- Pour une vallée binaire dans R_g, s'il y a un voisin à droite et aucun voisin à gauche, alors le pixel concerné est un coin potentiel ;
- Pour une vallée binaire dans R_d, s'il y a un voisin à gauche et aucun voisin à droite, alors le pixel concerné est un coin potentiel.

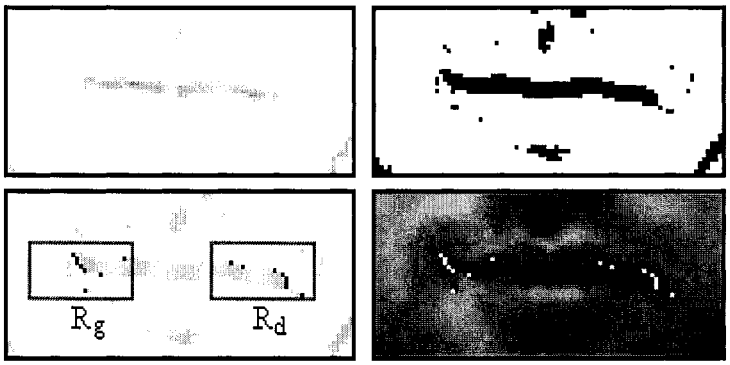


Figure 5.36. Un exemple du test effectué pour éliminer des pixels. Ce test est effectué sur le voisinage des pixels de vallée afin d'éliminer des possibilités de coins. L'image des vallées $I_{vallées}$ est obtenue puis binarisée pour obtenir $I_{vallées\_bin}$. Ensuite, les coins potentiels sont recherchés dans R_g et R_d.

La figure 5.36 démontre un exemple du test effectué dans le voisinage de chaque pixel de vallée binaire dans les fenêtres R_g et R_d afin de déterminer si un pixel est un coin potentiel. Deux groupes de coins potentiels, G_g et G_d, sont ainsi définis pour la gauche et la droite respectivement. À partir des coins potentiels dans G_g et G_d, un dernier test utilisant l'image des gradients I_{grad} est effectué afin de ne retenir qu'un seul coin pour la gauche et la droite. Ce test est basé sur la distribution des gradients habituellement rencontrés sur un des coins de la bouche. Sur le coin gauche par exemple, les gradients ont tendance à pointer vers l'extérieur du coin et sont orientés d'une façon circulaire, sauf vis-à-vis du début de la fermeture à droite du coin. Sur ce début de fermeture, les gradients ont plutôt tendance à être pointés vers l'extérieur tout en étant orientés d'une façon verticale. Ce phénomène s'explique bien par le fait que le coin gauche est une région foncée entourée d'une région plus pâle sauf vis-à-vis de la fermeture de la bouche. Le coin droit a les caractéristiques opposées selon l'horizontale. La figure 5.37 illustre deux exemples où la distribution des gradients est affichée dans la région du coin extérieur gauche de deux bouches. Ces distributions des gradients ne sont que très rarement rencontrées ailleurs que dans la région des coins et c'est pourquoi elles seront utilisées pour choisir les coins uniques.

a) b)

Figure 5.37. Deux exemples de la distribution des gradients. Ces gradients sont situés dans la région du coin extérieur gauche d'une bouche.

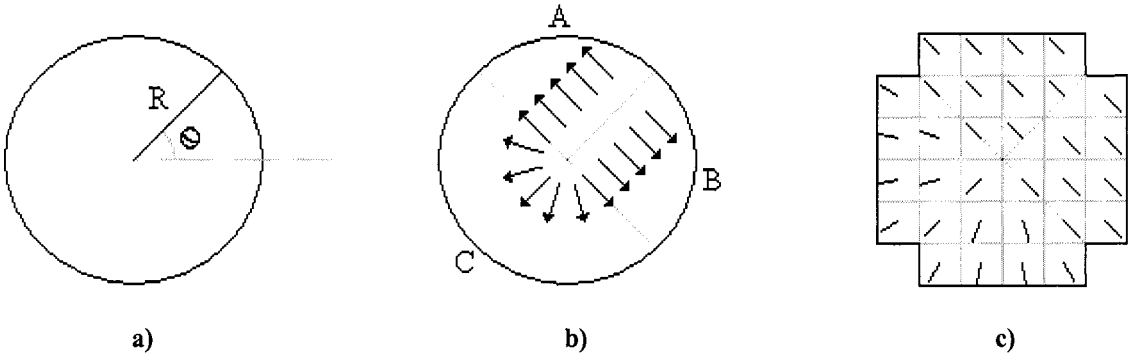


Figure 5.38. Le modèle M_c de distribution des gradients. Ce modèle est utilisé pour imiter la distribution dans le voisinage d'un coin de la bouche.

Afin de reconnaître les distributions recherchées dans I_{grad}, un modèle M_c de distribution des gradients semblables est créé. Ce modèle est illustré à la figure 5.38. L'équation 5.31 permet d'attribuer un score de ressemblance entre la position dans I_{grad} et le modèle M_c.

$$score\_grad = \sum_{M_c} \left(g_p.x \cdot G_p.x + g_p.y \cdot G_p.y \right) \tag{5.31}$$

où $(G_p.x, G_p.y)$ représente le gradient à la position correspondante dans I_{grad} vis-à-vis du gradient à la position $(g_p.x, g_p.y)$ dans M_c. La maximisation de l'équation 5.31 est donc recherchée afin d'avantager les gradients de même orientation entre I_{grad} et M_c. Les paramètres du modèle sont ajustés en fonction du coin (gauche ou droit) et de l'inclinaison possible que ce dernier peut avoir en fonction de l'orientation de la bouche. L'angle θ du modèle M_c a initialement la valeur de l'inclinaison obtenue entre les deux yeux. Un intervalle de $0.5 \cdot \pi$ centré en θ est ensuite défini afin de récupérer la meilleure inclinaison de M_c qui maximise l'équation 5.31 pour un coin potentiel. L'unique coin retenu (pour la gauche et la droite) est celui qui obtient un score maximal. Ainsi, C_g et

C_d ont de fortes probabilités d'avoir été obtenus sur des vallées aux extrêmes gauche ou droite (selon le coin) et respectant en autant que possible la distribution des gradients d'un coin de la bouche. La figure 5.39 illustre des exemples où les coins de la bouche ont été détectés avec la méthode proposée.

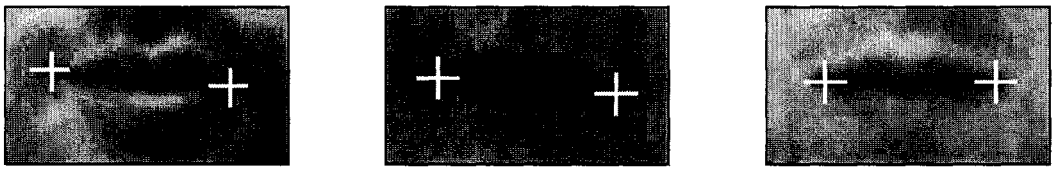


Figure 5.39. Des exemples où les coins de la bouche ont été détectés.

La détection des coins de la bouche avait déjà été effectuée dans les travaux de [84]. La méthode consistait à récupérer plusieurs coins potentiels dans la région inférieure du visage à l'aide d'un algorithme de détection de coins. Grâce à l'axe vertical du visage estimé auparavant, les coins potentiels obtenus étaient séparés en deux groupes pour la gauche et la droite. Ensuite, des couples de coins, formés d'un coin gauche et d'un coin droit, étaient formés et une probabilité était associée à chaque couple en fonction de leur configuration. Le couple qui maximisait la probabilité d'avoir une bonne configuration, selon l'anthropométrie du visage, était retenu et les deux coins de la bouche étaient trouvés. La détection des coins de la bouche a cependant été refaite dans ce travail car beaucoup d'imprécision caractérisait la méthode de [84]. Peu de caractéristiques de la bouche étaient en fait exploitées. Par exemple, il n'y avait pas de contrainte concernant l'orientation des coins de la bouche, alors qu'il

s'agit d'une contrainte très importante car un coin gauche est toujours orienté vers la droite et vice-versa.

5.3.2.3. Fermeture de la bouche

Lorsque les coins extérieurs de la bouche sont disponibles, ceux-ci sont utilisés pour localiser la courbe de la fermeture. Bien que plusieurs techniques consistent à rechercher les arêtes pour localiser la fermeture, le type d'images rencontrées dans ce projet ne permettait généralement pas d'obtenir des arêtes bien définies sur la fermeture. D'autres techniques vont rechercher la couleur rouge des lèvres afin de localiser la fermeture là où la couleur est plus sombre. Avec l'analyse de la couleur des lèvres dans ce projet, les résultats étaient peu convaincants et une approche plus fiable semblait nécessaire. En analysant les images des vallées, une très forte démarcation est obtenue sur la fermeture de la bouche comme le montre la figure 5.40. Puisque la fermeture forme une courbe peu complexe pouvant être estimée par une parabole, la méthode utilisée consiste donc à utiliser une parabole et l'image des vallées.

Entre les coins C_g et C_d, de fortes vallées sont rencontrées le long de la fermeture des deux lèvres. Une parabole P_c sera donc positionnée en maximisant le passage sur de fortes vallées de $I_{\text{vallées}}$, ceci est décrit par l'équation 5.32.

$$score\_vallées = \int_{P_c} I_{vallées}(\vec{p}) \cdot \overrightarrow{dp} \tag{5.32}$$

où $\vec{p}$ est un vecteur de position sur la parabole P_c. La parabole permettant de maximiser l'équation 5.32 est donc celle recherchée. Il est à noter qu'un seul paramètre de

l'équation est variable, soit la hauteur. Et pour trouver le maximum, plusieurs essais dans un intervalle sont effectués en variant la hauteur. Puisque les coins définissent les limites de la parabole et que l'on suppose que la bouche est presque symétrique, l'orientation et la largeur restent constantes. Un intervalle de hauteur est défini grâce aux positions trouvées pour C_g et C_d. La figure 5.40 illustre l'intervalle H_p admissible pour h afin de positionner la parabole P_c et l'équation 5.33 permet de déterminer cet intervalle pour de bons résultats en se référant à l'anthropométrie du visage. Il est à noter qu'un grand intervalle H_p peut être admissible puisque la fermeture de la bouche contient beaucoup plus de fortes vallées qu'ailleurs dans le voisinage de la bouche. La parabole P_c permettra ainsi de trouver le contour extérieur des lèvres avec plus de facilité.

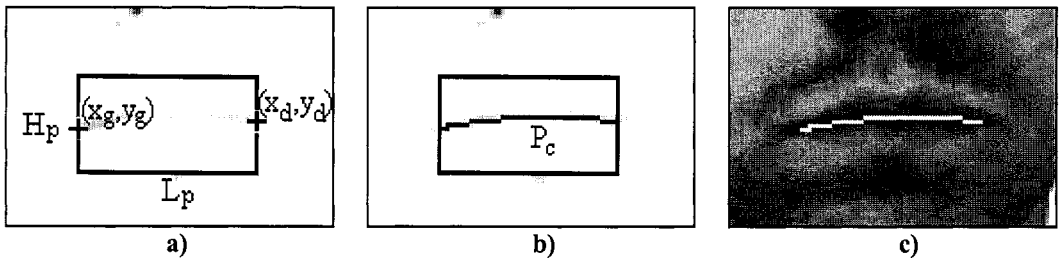


Figure 5.40. Localisation de la parabole P_c de la fermeture à l'intérieur de l'intervalle défini.

$$\frac{H_p}{2} \leq h \leq \frac{H_p}{2} \tag{5.33}$$

où $$H_p = \frac{L_p}{2} = \frac{x_d - x_g}{2}$$

5.3.2.4. Contours extérieurs de la bouche

Comme il a été mentionné plus tôt, les contours extérieurs de la bouche peuvent être très difficiles à détecter puisque le changement de couleur entre la peau et les lèvres ne se fait pas toujours de façon brusque. Pour améliorer la détection, il faut donc réduire l'intervalle de recherche de ces contours. Puisque les coins extérieurs et la courbe de fermeture de la bouche sont maintenant disponibles, il est plus facile de trouver les contours extérieurs car il suffit de relier deux paraboles aux coins C_g et C_d. Ces paraboles seront nommées P_{he} et P_{be} pour celles du haut et du bas respectivement et auront une hauteur assez près de la fermeture P_c. P_{he} sera en haut de P_c et P_{be} sera en bas. Pour ces deux paraboles, un intervalle de hauteur est donc fixé à $0.25 \cdot L_p$ pour de bons résultats, ceci permet d'inclure diverses épaisseurs de lèvre. Plusieurs techniques consistent à rechercher les arêtes pour localiser le contour extérieur. Cependant, d'après les images rencontrées dans ce projet, beaucoup de lèvres ne permettaient pas d'avoir des arêtes bien définies car les frontières entre les lèvres et la peau sont parfois peu visibles. D'autres techniques vont analyser la couleur rouge des lèvres mais là encore de mauvais résultats étaient obtenus après des essais effectués. En analysant plusieurs images traitées (les arêtes, les gradients, les vallées, la couleur rouge, etc) il a semblé préférable d'analyser la magnitude des gradients tout en utilisant une estimation de l'épaisseur des lèvres à l'aide des proportions anthropométriques du visage. L'image $I_{\text{grad\_mod}}$ des modules des gradients sera donc utilisée pour localiser P_{he} et P_{be}. Ces paraboles devront maximiser le passage par des gradients de magnitudes élevées ainsi qu'une probabilité de hauteur de la lèvre comme le décrivent les équations 5.34, 5.35 et

5.36. La figure 5.41 illustre deux exemples de la distribution des gradients sur le contour de la bouche. D'après cette figure, des gradients de magnitudes élevées sont souvent rencontrés sur le contour extérieur de la bouche mais ils sont surtout concentrés près de la fermeture. De plus, ces gradients sont parfois très faibles sur le contour extérieur lorsqu'il y a une faible démarcation entre les lèvres et la peau comme pour le cas de la lèvre inférieure sur la figure 5.41.b). La probabilité sur la hauteur de la lèvre est donc utilisée pour contourner ce problème.

$$score\_grad\_h = P(h_{he}) \cdot \int_{P_{he}} I_{grad\_\mathrm{mod}}(\overrightarrow{p_h}) \cdot \overrightarrow{dp_h} \tag{5.34}$$

$$score\_grad\_b = P(h_{be}) \cdot \int_{P_{be}} I_{grad\_\mathrm{mod}}(\overrightarrow{p_b}) \cdot \overrightarrow{dp_b} \tag{5.35}$$

$$P(h) = \frac{e^{\frac{-(h-h_0)^2}{2\sigma^2}}}{\sqrt{2\pi \cdot \sigma^2}} \tag{5.36}$$

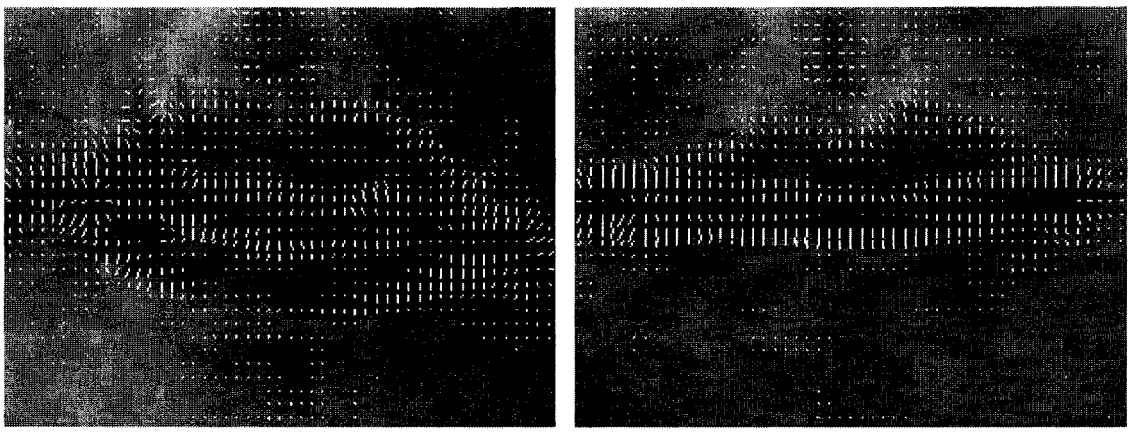
où h : hauteur de la parabole

h_0 : hauteur espérée de la parabole

σ : déviation standard

Pour de bons résultats, la valeur de h_0 a été fixée à $0.15 \cdot L_p$ et $0.2 \cdot L_p$ pour les paraboles P_{he} et P_{be} respectivement. Ces valeurs ont été obtenues en analysant plusieurs visages et en calculant les épaisseurs moyennes des lèvres. Et pour les deux cas, σ a été fixé à 3. Cette valeur de σ a été choisie en fonction des visages analysés mais a été réduite pour que le contour extérieur des lèvres ait plutôt tendance à se positionner à un endroit approximatif qu'à un autre endroit dans l'image forte en gradients. En effet, les

gradients des contours ne sont pas toujours présents et la localisation précise du contour extérieur n'a pas beaucoup d'importance pour estimer les mouvements de la bouche : ce sera plutôt le contour intérieur qui permettra d'estimer les mouvements.



a) b)

Figure 5.41. Deux exemples de la distribution des gradients sur le contour de la bouche.

5.3.2.5. Informations nécessaires au suivi

Le suivi de la bouche ne peut pas être effectué de la même façon que durant l'intitialisation du modèle puisque trop de liberté sur la géométrie de la bouche devra être considérée. À partir de l'adaptation initialement effectuée, certaines informations devront être prélevées afin de faciliter le suivi tout en laissant assez de liberté à la bouche. Bien que certaines techniques [72][74][76][77] utilisent des masques déformables pour s'adapter aux arêtes de la bouche, les résultats expérimentaux ont démontré qu'avec les images utilisées dans ce projet il est préférable d'utiliser une autre approche. Une méthode a été développée d'après les caractéristiques suivantes qui ont été observées lors de mouvements normaux de la bouche, lorsque l'usager parle ou sourit par exemple:

- Entre l'image courante I_t et l'image initiale I_0, l'image des coins extérieurs de la bouche ainsi que le centre des contours extérieurs ne varient pas beaucoup même lorsque la bouche est en mouvement ;
- La bouche est très souvent symétrique par rapport à l'axe vertical ;
- Sur toute la lèvre du haut ou du bas, une région d'images sur une certaine position horizontale peut être très semblable à celle sur une autre position (peu de variations sur les lèvres selon l'horizontale).

Puisque certaines régions d'images sur la bouche peuvent bien se comparer entre I_t et I_0 et que la recherche de la bouche d'image en image semble peu robuste avec la méthode utilisée initialement, une autre méthode est nécessaire pour le suivi. Il est à noter que tout le contour extérieur de la bouche peut être exprimé à l'aide de quatre positions, soit les deux coins et le centre des contours extérieurs du haut et du bas. La méthode consiste donc à prélever, sur l'image initiale I_0, quatre petites régions d'images qui serviront à effectuer des comparaisons de régions sur les autres images de la séquence. Ces quatre régions seront nommées T_g, T_d, T_h et T_b pour le coin gauche, le coin droit, le centre du contour extérieur du haut et le centre du contour du bas respectivement.

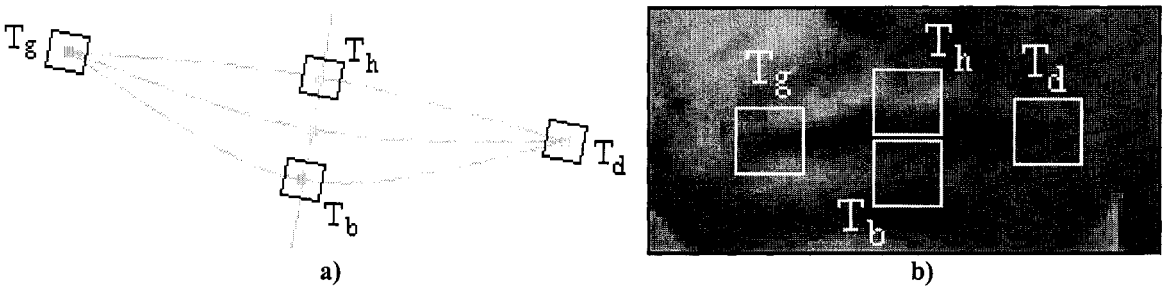


Figure 5.42. Quatre régions d'images prélevées sur I_0 à l'aide d'un modèle géométrique. En a), le modèle. En b), les régions prélevées sur I_0.

La figure 5.42 illustre comment sont prélevées ces quatre régions à l'aide d'un modèle géométrique. Les informations prélevées sur les quatre régions T_g, T_d, T_h et T_b ainsi que les traitements effectués sont les mêmes que ceux décrits pour le suivi des points 2D (chapitre 3). En effet, les quatres regions prélevées initialement sur I_0 seront recherchées sur les autres images de la séquence. De plus, les épaisseurs des lèvres en leur centre, H_h et H_b pour la lèvre du haut et du bas respectivement, sont retenues pour faciliter le suivi comme il sera expliqué plus loin.

5.3.3. Réajustement du modèle sur l'image courante

Le réajustement du modèle de la bouche se fait selon les étapes suivantes :

1. Adapter les coins extérieurs de la bouche ;
2. Adapter le centre des courbes extérieures de la bouche ;
3. Adapter le centre des courbes intérieures de la bouche.

5.3.3.1. Adaptation des coins extérieurs de la bouche

Lorsque les régions T_g et T_d sont recueillies sur I_0 vis-à-vis des coins extérieurs de la bouche, elles sont ensuite recherchées avec la méthode décrite pour le suivi des points 2D (chapitre 3). Bien que les coins de la bouche subissent plusieurs transformations tout au long de la séquence, peu de régions dans un petit voisinage ressemblent aux coins recherchés. Les positions obtenues par ce suivi ont donc tendance à se localiser aux bons endroits même si le score de la corrélation est quelquefois peu élevé. En effet, le coin d'une bouche est une région forte en texture et en changement d'intensité lumineuse. Mais dans le voisinage d'un coin, comme sur la peau ou les lèvres, la texture est plus uniforme et le changement d'intensité est donc plus faible. Pour ces raisons, un point du suivi reste assez bien accroché à un coin de la bouche. De plus, d'après les configurations probables de la bouche (lorsqu'une personne ne fait pas de grimace par exemple), l'angle d'ouverture des coins ne varie pas beaucoup puisqu'il se situe aux extrêmités de la bouche, il n'y a donc pas beaucoup de variations sur l'image. Et lors d'un sourire par exemple, les variations dans l'image sont surtout causées par des rotations et la méthode utilisée est conçue pour affronter ce genre de situation. La figure 5.43 illustre quelques exemples de variations rencontrées pour les coins d'une même bouche, les régions ont été sélectionnées manuellement sur cette figure. En a), les régions des coins sont recueillies sur l'image initiale I_0 où la bouche est fermée, au repos et étendue à l'horizontale. En b), une légère rotation de la bouche est obtenue et la région de gauche risque d'être difficile à retrouver étant donné la présence de l'image de l'arrière-plan.

En c), une rotation est obtenue et le coin de gauche est plus difficile à voir. En d), les angles d'ouverture des coins sont plus grands et les régions sont donc quelque peu différentes de celles sur I_0. En e), il y a à la fois une rotation et des angles d'ouverture plus grands pour les coins. Et en f), il s'agit d'une situation très difficile pour la méthode proposée puisque les angles d'ouverture des coins sont tellement grands qu'il y a maintenant peu de ressemblance avec les régions recueillies sur I_0. Sur ce dernier cas, il y aura sans doute des imprécisions mais les coins détectés devraient tout de même se situer assez près des vraies positions puisqu'aucun autre endroit dans le voisinage n'est autant semblable à ceci a été recueilli sur I_0.

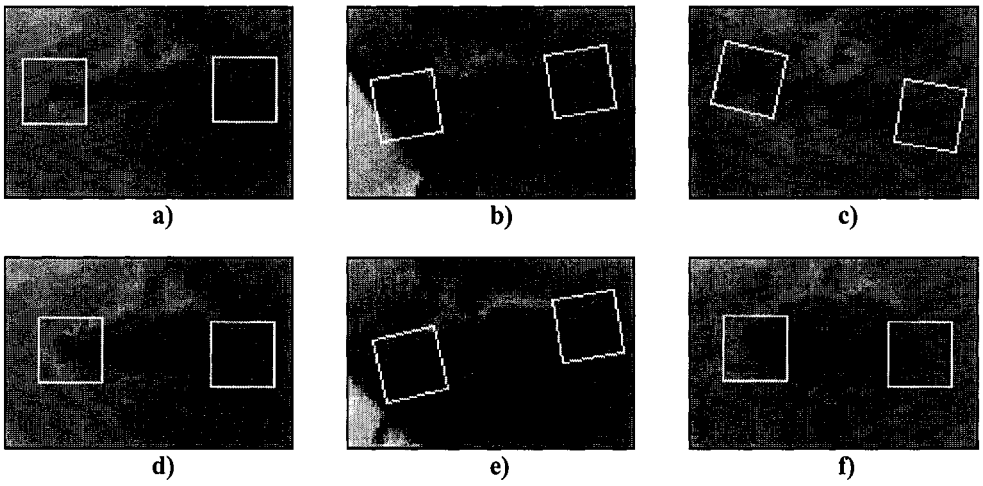


Figure 5.43. Quelques exemples de variations rencontrées pour les coins d'une même bouche. Les régions ont été sélectionnées manuellement.

5.3.3.2. Adaptation du centre des courbes extérieures de la bouche

Tout comme pour T_g et T_d, les régions T_h et T_b sont recherchées grâce à la méthode utilisée pour le suivi des points 2D (chapitre 3). Cependant, une légère modification a été appliquée concernant les positions potentielles. Plutôt que la recherche de la position soit effectuée sur une région rectangulaire, celle-ci est plutôt effectuée sur une région linéaire : puisque la position recherchée doit être située à mi-chemin, selon l'horizontale, entre les deux coins de la bouche, une recherche verticale est effectuée relativement à l'angle formé par les deux coins de la bouche. Et pour laisser un peu de liberté au cas où la bouche ne serait pas symétrique, un petit intervalle horizontal de quelques pixels a d'ailleurs été admis pour la recherche. Pour de bons résultats, cet intervalle a été fixé à deux pixels à droite et à gauche. Cette méthode a été appliquée afin de résoudre un problème d'instabilité rencontré dû au fait que le haut du centre de la lèvre est semblable à d'autres endroits sur le haut de la lèvre. Ainsi, lors d'un mauvais suivi, le haut de la lèvre était atteint mais pas vis-à-vis du centre de la bouche. Et il en est de même avec le bas de la lèvre inférieure. En diminuant l'intervalle de recherche, la position obtenue est donc toujours presque à mi-chemin entre les deux coins. La figure 5.44 illustre des exemples d'instabilité de positions qui a été corrigée grâce à la méthode proposée. Sur cette figure, la même image est utilisée pour chaque rangée. Sur la colonne de gauche, la correction n'a pas été effectuée tandis qu'elle l'a été sur celle de droite.

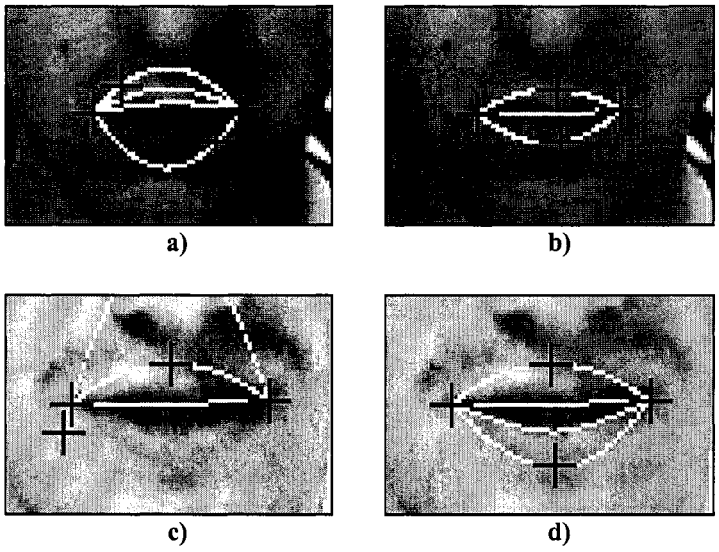


Figure 5.44. Quelques exemples d'instabilités de positions qui ont été corrigées.

5.3.3.3. Adaptation du centre des courbes intérieures de la bouche

Les courbes intérieures de la bouche sont difficiles à retrouver car l'image sur ces régions peut avoir beaucoup de variations : bouche ouverte ou fermée, dents visibles ou non, etc. Pour cette raison, cette étape n'est traitée qu'une fois que toutes les autres informations sur la bouche aient été obtenues. Plusieurs techniques consistent à rechercher les arêtes du contour, à analyser la couleur rouge des lèvres, à analyser l'image des sommets engendrés par les dents, etc. D'après des essais effectués avec de telles techniques, les résultats n'étaient pas convaincants principalement à cause des diverses variations à l'intérieur de la bouche. Il a donc été jugé préférable de se fier aux informations déjà recueillies et d'exploiter le fait que l'épaisseur des lèvres varie peu en fonction de la configuration de la bouche. Bien que de forts gradients puissent être rencontrés sur les dents ou la langue, ils seront tout de même utilisés pour localiser le contour intérieur de la bouche car ils sont souvent là aussi présents. Puisque les

contours se confondent lorsque la bouche est fermée, cette caractéristique sera exploitée afin de déterminer si la bouche est ouverte ou fermée. Au départ, la bouche sera donc traitée comme si elle était ouverte et un test permettra de déterminer si elle est en fait fermée.

D'après l'intialisation de la bouche sur I_0, les épaisseurs des lèvres H_h et H_b ont été retenues et seront maintenant utilisées pour faciliter le suivi. La bouche a principalement deux états qui sont "ouvert" et "fermé". Ces états sont détectés au tout début selon la distance qui sépare le centre des paraboles P_{he} et P_{be} et les épaisseurs H_h et H_b obtenues. L'algorithme suivant permet de déterminer l'état de l'ouverture:

- Positionner les paraboles P_{hi} et P_{bi} à l'aide de $I_{\text{grad\_mod}}$ et la fonction de probabilité (on suppose la bouche ouverte) ;
- Si P_{bi} se retrouve au dessus de P_{hi} alors
  - La bouche est fermée ;
  - Positionner la parabole P_c à l'aide de $I_{\text{vallées}}$.
- Sinon la bouche est ouverte.

5.3.3.3.1. Premier cas : la bouche est ouverte

Pour positionner les paraboles P_{hi} et P_{bi} des courbes intérieures d'une bouche ouverte, l'image des magnitudes des gradients $I_{\text{grad\_mod}}$ est utilisée. Ces paraboles doivent maximiser le passage sur de forts gradients tout en maximisant la probabilité que l'épaisseur des lèvres soit semblable à celle obtenue lors de l'initialisation de la bouche sur I_0. Les équations 5.37 et 5.38 doivent donc être maximisées pour obtenir les bonnes

paraboles et l'équation 5.39 décrit la fonction de probabilité utilisée. La maximisation est effectuée avec plusieurs essais de Δh_h et Δh_b sur des intervalles entre 0 et $2 \cdot \Delta h_0$, ceci permet une variation respectable de l'épaisseur de la lèvre.

$$score\_grad\_h = P(\Delta h_h) \cdot \int_{P_{hi}} I_{grad\_\mathrm{mod}}(\overrightarrow{p_h}) \cdot \overrightarrow{dp_h} \tag{5.37}$$

$$score\_grad\_b = P(\Delta h_b) \cdot \int_{P_{bi}} I_{grad\_\mathrm{mod}}(\overrightarrow{p_b}) \cdot \overrightarrow{dp_b} \tag{5.38}$$

$$P(\Delta h) = \frac{e^{\frac{-(\Delta h - \Delta h_0)^2}{2\sigma^2}}}{\sqrt{2\pi \cdot \sigma^2}} \tag{5.39}$$

où $\Delta h_h = h_{he} - h_{hi}$: épaisseur du centre de la lèvre du haut

$\Delta h_b = h_{be} - h_{bi}$: épaisseur du centre de la lèvre du bas

Δh_0 : épaisseur espérée de la lèvre

σ : déviation standard

Pour de bons résultats, la valeur de Δh_0 a été fixée à l'épaisseur au centre des lèvres obtenue intialement sur I_0. Et pour les deux contours intérieurs, σ a été fixé à 3 d'après des essais expérimentaux. Cette valeur de σ a d'ailleurs été choisie assez faible afin d'augmenter les probabilités que l'épaisseur des lèvres soit semblable à ce qui a été obtenu initialement. Ceci permet d'éviter que les contours intérieurs soient détectés sur d'autres endroits forts en gradients à l'intérieur de la bouche comme sur la pointe des dents par exemple. Il est à noter qu'encore une fois, seules la hauteur des paraboles est

variable puisque les coins sont fournis et que l'on suppose que la bouche est symétrique par rapport à l'axe vertical.

5.3.3.3.2 Deuxième cas : la bouche est fermée

Lorsque la bouche est fermée, seulement une parabole P_c doit être positionnée. Puisque les vallées sont très fortes sur la fermeture, l'image des vallées $I_{vallées}$ sera utilisée et l'équation 5.40 devra être maximisée. Pour ce cas, les résultats sont tellement stables qu'il a été jugé préférable de ne pas imposer que l'épaisseurs des lèvres soit semblable à celle obtenue initialement.

$$score\_vallées = \int_{P_c} I_{vallées}(\vec{p}) \cdot \overrightarrow{dp} \tag{5.40}$$

où $\vec{p}$ est un vecteur de position sur la parabole P_c.

5.3.4. Analyse des résultats obtenus

La figure 5.45 illustre quelques exemples d'adaptation de la bouche obtenus dans diverses situations. Sur cette figure, les images de la colonne de gauche ont été obtenues sur l'image initiale I_0. Afin d'isoler le problème du suivi de la bouche, la position initiale des yeux a été effectuée manuellement. Les images des colonnes du centre et de droite ont été obtenues pendant le suivi et pour chaque rangée, des séquences différentes d'individus différents ont été utilisées.

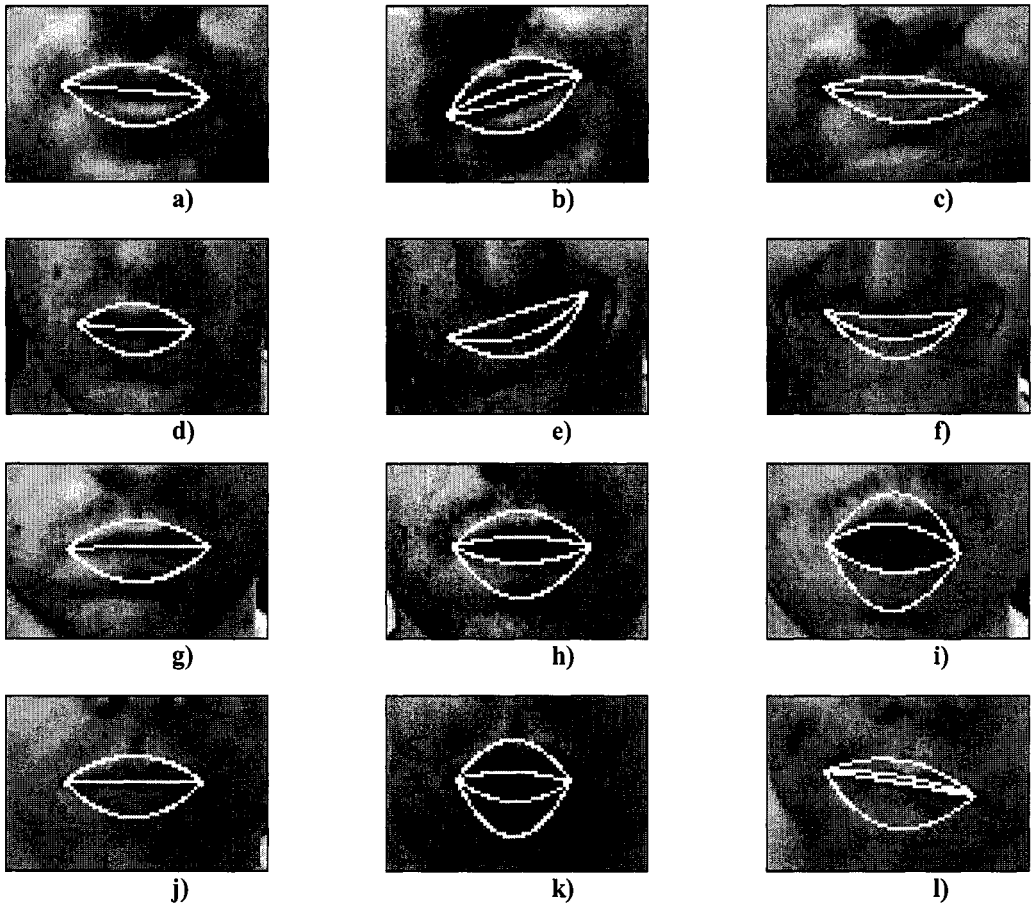


Figure 5.45. Quelques exemples d'adaptation de la bouche pour diverses situations.

En a), d), g) et j), les modèles de la bouche ont été adaptés sur I_0 avec une précision acceptable bien que le contour extérieur soit quelquefois peu précis étant donné le faible changement d'intensité entre les lèvres et la peau. En k), une grande imprécision est rencontrée sur le coin droit, principalement causée par la grande différence de la région du coin par rapport à celle obtenue initialement. En i) et k), le contour extérieur du bas de la bouche est assez imprécis vu la faible démarcation entre la lèvre inférieure et la peau. La figure 5.46 illustre deux exemples où l'adaptation de la

bouche lors du suivi est tout de même obtenue avec précision même si les images sont floues étant donné le mouvement rapide des usagers.

a) b)

Figure 5.46. Des exemples de l'adaptation sur des images floues. La bouche a été adaptée même sur des images floues dues à un mouvement rapide de l'usager.

a) b)

Figure 5.47. Des exemples où la bouche est mal localisée.

La figure 5.47 illustre des exemples où la bouche a été mal localisée. En a), la lèvre inférieure est mal localisée à cause de la langue et le peu de changement d'intensité entre la peau et la lèvre. En b), le coin gauche de la bouche est mal localisé à cause du manque de ressemblance avec le même coin sur l'image initiale. De plus, la présence des dents nuit à la localisation de la lèvre inférieure. Plus de détails sur les résultats sont présentés en Annexe III.

5.4. Intégration et analyse des résultats du suivi

Dans ce travail, le suivi des mouvements non rigides inclut celui des yeux, des sourcils et de la bouche. Bien que ces derniers aient été décrits séparément dans ce chapitre, ils sont tous intégrés au système.

Il existe diverses façons d'analyser les résultats obtenus mais il est difficile d'en quantifier la précision étant donné la quantité de paramètres utilisés et leur interdépendance. Il existe une très grande quantité d'exemples, en voici quelques-uns :

1. Si l'œil détecté est trop large, l'ouverture peut sembler réduite ;
2. Si un coin de l'œil est mal positionné, la direction du regard peut sembler imprécise ;
3. Une localisation trop éloignée des yeux va augmenter la largeur des yeux ;
4. Une mauvaise localisation des yeux peut facilement engendrer une mauvaise localisation des sourcils ;
5. Des iris bien positionnés, mais dont le rayon est imprécis, n'a aucun impact pour le mouvement ;
6. Un mauvais angle d'ouverture des sourcils pourrait tout de même donner une bonne représentation sur un modèle virtuel de l'usager ;
7. Une mauvaise localisation des contours extérieurs de la bouche pourrait tout de même donner une bonne représentation sur un modèle virtuel de l'usager ;
8. Des localisations totalement fausses pourraient tout de même donner une bonne représentation sur un modèle virtuel de l'usager.

Il est donc difficile de quantifier correctement la précision des résultats. De plus, il est difficile de classifier les résultats en fonction des séquences utilisées car là aussi il existe une très grande variété. En voici quelques exemples :

1. L'âge, la race et le sexe de l'usager ;
2. La qualité des images et la visibilité de l'usager ;
3. Les mouvements de l'usager et les diverses combinaisons possibles.

Dans ce travail, l'analyse des résultats est donc plutôt qualifiée que quantifiée et ces résultats seront présentés à l'aide des modèles géométriques directement sur les images. C'est d'ailleurs ceci a été effectué d'après les articles étudiés sur ce sujet. Il est d'ailleurs difficile de comparer les résultats avec ceux des articles étudiés pour deux principales raisons :

1. Les algorithmes implantés sont difficilement accessibles, de même que les séquences d'images utilisées. Pour bien comparer, les séquences devraient être les mêmes, bien représentatives au problème et assez vastes pour concerner une grande population d'usagers ;
2. Le problème n'est souvent pas posé de la même façon, surtout en ceci concerne les connaissances initiales, les types de mouvements et la résolution des images. Par exemple, certaines techniques utilisent des connaissances à priori des usagers comme la configuration initiale des yeux. D'autres fois, les mouvements non rigides sont analysés en supposant qu'il

n'y a pas de mouvement rigide. Les usagers semblent d'ailleurs souvent sélectionnés afin de permettre de bons résultats et pour ce faire, les images utilisées sont de haute résolution.

Dans ce projet, les méthodes ont été proposées en se basant sur des algorithmes existants mais ont par la suite été adaptées au problème rencontré. En général, de très mauvais résultats auraient été obtenus si de tels algorithmes avaient été implantés intégralement. Des résultats très décevants avaient d'ailleurs été obtenus après plusieurs essais. Les résultats obtenus dans ce projet sont donc très appréciés comparativement aux essais obtenus avec d'autres techniques. Finalement, dans la plupart des articles rencontrés, la réelle performance n'est pas bien expliquée, même d'une façon qualificative : seulement de bons résultats sont présentés et, bien que les forces des algorithmes soient expliquées, les faiblesses ne le sont pas et semblent quelquefois très évidentes pour le lecteur. L'analyse des résultats obtenus est détaillée en Annexe III.

5.5. Conclusion

Dans ce chapitre, le suivi des yeux, des sourcils et de la bouche ont été implantés. D'après les résultats obtenus, on peut conclure que le système fonctionne bien dans plusieurs situations mais manque souvent de précision lorsque des situations imprévues sont rencontrées (des images floues ou des reflets lumineux par exemple). De plus, il a été constaté qu'une mauvaise localisation initiale sur la première image a souvent de graves conséquences sur les autres images de la séquence. Cependant, le

système peut tout de même corriger quelquefois les imprécisions engendrées intialement, surtout concernant les yeux et les sourcils.

5.6. Améliorations suggérées

Afin de rendre le système plus robuste, il serait d'abord nécessaire d'augmenter la précision des localisations sur l'image initiale. Ensuite, exploiter davantage les caractéristiques des éléments non rigides et leurs interdépendances. Par exemple, la couleur des lèvres diffère de celle de la peau, les yeux sont habituellement grand ouverts lorsqu'ils regardent vers le haut, les deux sourcils sont plus susceptibles d'avoir la même hauteur, etc. De plus, il serait intéressant d'étudier d'autres images intermédiaires que celles des vallées, des sommets et des arêtes (voir Annexe V). Le système a d'ailleurs quelques problèmes de lenteur étant donné le nombre de traitements effectués, il serait donc intéressant d'utiliser des traitements optimaux pour les intégrations de paraboles, de cercles et de segments de droites (voir Annexe IV) car ces opérations sont très fréquentes. Finalement, l'utilisation d'un modèle virtuel précis de l'usager permettrait sans doute d'augmenter la robustesse en imposant des contraintes aux mouvements non rigides.

CHAPITRE 6

CONCLUSION

Dans ce mémoire, les mouvements rigides de la tête d'un usager ainsi que la localisation des éléments non rigides ont été étudiés. Puisque les mouvements rigides sont évalués en fonction d'un modèle virtuel de l'usager, il est important que ce modèle soit très fidèle. Malheureusement, ce modèle est estimé d'une façon assez grossière car peu d'informations sont disponibles concernant l'usager. En effet, seule l'image frontale de l'usager est fournie, ceci est insuffisant pour recueillir toutes les mesures nécessaires. Il en résulte donc que les mouvements rigides ne peuvent être estimés avec beaucoup de précision avec la méthode utilisée, même si le suivi des points 2D est effectué avec une grande précision. Étant donné l'imprécision des mouvements rigides obtenus, ceux-ci n'ont pas été utilisés pour la paramétrisation des mouvements non rigides du visage. Si les mouvements rigides avaient pu être obtenus avec une grande précision, les mouvements non rigides auraient pu être paramétrisés plus facilement car les positions peu probables auraient pu être éliminées grâce à la projection 2D du modèle 3D. Heureusement, les contraintes anthropométriques du visage ont pu être utilisées pour paramétriser ces mouvements non rigides, ceci a permis d'obtenir des résultats acceptables. Des essais avaient déjà été effectués sans tenir compte de ces contraintes et les résultats étaient très décevants. Les grandes difficultés rencontrées dans ce projet proviennent principalement du fait que les images à traiter sont très difficiles à décrire. Pour l'image d'un oeil, par exemple, plusieurs caractéristiques varient en fonction des

mouvements rigides et non rigides, du type d'usager, de l'éclairage, du bruit, etc. Parmi les images intermédiaires utilisées, celles des vallées et des sommets [53] ont permis de recueillir plusieurs invariances qui ont facilité la paramétrisation tandis que l'image des arêtes et l'information de la couleur ont été très peu utilisées car peu d'invariances utiles étaient retrouvées. Les gradients ont d'ailleurs été souvent utilisés. Les algorithmes de détection de coins semblaient peu efficaces pour détecter les coins des yeux ou de la bouche. Plusieurs essais avaient été effectués avec des méthodes de minimisation par descente du gradient pour le suivi des points 2D [79] et l'adaptation des modèles géométriques pour les yeux et la bouche [72][76][77]. Ces méthodes ont été abandonnées puisque des minimums locaux étaient trop souvent rencontrés.

6.1. Améliorations suggérées au projet

Tout d'abord, il serait important d'améliorer la précision du modèle virtuel de l'usager puisque ce modèle est à la fois utilisé pour représenter l'usager et évaluer les mouvements rigides. Il pourrait d'ailleurs être utilisé pour paramétriser les mouvements non rigides du visage. Plusieurs techniques [3][9] consistent à prélever des images de l'usager selon certaines configurations contrôlées. Plus de mesures pourraient ainsi être obtenues avec plus de précisions et ainsi reconstituer plus fidèlement le modèle virtuel. La paramétrisation des mouvements non rigides pourrait entre autre profiter des nouvelles informations concernant l'usager. Le projet pourrait d'ailleurs comporter un algorithme de reconnaissance de visages, ceci éviterait d'avoir à reconstruire le modèle virtuel à chaque fois qu'un même usager utilise le système. Certaines informations du

visage pourraient aussi être analysées afin d'évaluer les expressions faciales de l'usager. Le son de la voix émise par l'usager pourrait d'ailleurs être analysé afin d'attribuer de bonnes configurations à la bouche. Finalement, de meilleurs résultats pourraient sans doute être obtenus si une caméra de meilleure qualité était utilisée et que l'environnement était mieux contrôlé.

RÉFÉRENCES

[1] D.Maio, D.Maltoni. « **Real-time face location on gray-scale static images** ». Pattern Recognition, September 2000, vol.33, no. 9, pp. 1525-1539

[2] Tzyy-Yuang Shiang, « **A statistical approach to data analysis and 3-D geometric description of the human head and face** ». Proc. Natl. Sci. Counc. ROC(B), 1999, vol. 23, no. 1, pp. 19-26

[3] T. Akimoto, Y. Suenaga, R. S. Wallace « **Automatic Creation of 3D Facial Models** ». IEEE Computer Graphics and Applications, 1993, vol. 13, no. 5, pp. 16-22

[4] T. Horprasert, Y. Yacoob, L. S. Davis. « **An anthropometric shape model for estimating head orientation** ». 3rd International Workshop on Visual Form, Capri, Italy, May 1997.

[5] D. DeCarlo, D. Metaxas, M. Stone. « **An anthropometric face model using variational techniques** ». In Proceeding of the 25th annual conference on Computer graphics and interactive techniques, 1998, pp. 67-74

[6] A. Yilmaz, K. Shafique, M. Shah. « **Estimation of rigid and non-rigid facial motion using anatomical face model** ». 16th International Conference on Pattern Recognition, 2002, vol. 1, pp. 10377-10380

[7] J. W. Davis, S. Vaks. « **A perceptual user interface for recognizing head gesture acknowledgements** ». In Proceedings of the 2001 workshop on Perceptive user interfaces, 2001, pp. 1-7

[8] D. DeCarlo, D. Metaxas. « **Deformable model-based shape motion analysis from images using motion residual error** ». In Proceedings ICCV '98, 1998, pp. 113-119

[9] M. Malciu, F. Preteux, « **A robust model-based approach for 3D head tracking in video sequences** ». 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'2000), Grenoble, France, 2000, pp. 169-174

[10] A. Bottino. « **Real time head and facial features tracking from uncalibred monocular views** ». The 5th Asian Conference on Computer Vision, Melbourne, Australia, 23-25 January 2002,

[11] A. Schödl, A. Haro, I. A. Essa. « **Head tracking using a textured polygonal model** ». In Proceedings of Perceptual User Interfaces Workshop, (held in Conjunction with ACM UIST 1998), San Francisco, CA., November 1998

[12] A. H. Geed, R. Cipolla. « **Fast visual tracking by temporal consensus** ». Image and Vision Computing, vol. 14, no. 2, pp.105-114, 1996

[13] G. D. Hager, P. N. Belhumeur. « **Efficient region tracking with parametric models of geometry and illumination** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, October 1998, vol. 20, no. 10,

[14] S. Basu, I. Essa, A. Pentland. « **Motion regularization for model-based head tracking** ». In Proceeding of the IEEE Int'I Conf. On Pattern Recognition (ICPR 1996) (Vienna, Australia), vol. 3, pp. 611-616

[15] M. La Cascia, J. Isodora, S. Sclaroff. « **Head tracking via robust registration in texture map images** ». Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1998, pp. 508-514

[16] A. Schödl, K. Schwan, I. A. Essa. « **Adaptive parallelization of model-based head tracking** ». In Proceedings of 1999 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'99), Monte Carlo Resort, Las Vegas, Nevada, USA, June 1999.

[17] Y. Zhang, C. Kambhamettu. « **Robust 3D head tracking under partial occlusion** ». Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition 2000, pp. 176-182

[18] S. Birchfield. « **Elliptical head tracking using intensity gradients and color histograms** ». Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Santa Barbara, California), June 1998, pp. 232-237

[19] T. Horprasert, Y. Yacoob, L. S. Davis. « **Computing 3-D head orientation from a monocular image sequence** ». Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (FG '96), 1996, pp. 242-248

[20] H. A. Rowley, S. Baluja, T. Kanade. « **Neural network-based face detection** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, January 1998, vol. 20, no. 1,

[21] H. Wu, Q. Chen, M. Yachida. « **Face detection from color images using a fuzzy pattern matching method** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, June 1999, vol. 21, no. 6

[22] A. Smolic, B. Makai, T. Sikora. « **Real-time estimation of long-term 3-D motion parameters for SNHC face animation and model-based coding applications** ». IEEE Transactions on Circuits and Systems for Video Technology, March 1999, vol. 9, no. 2

[23] C. S. Wiles, A. Maki, N. Matsuda. « **Hyperpatches for 3D model acquisition and tracking** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, December 2001, vol. 23, no. 12

[24] K. Okada, S. Akamatsu, C. Von Der Malsburg. « **Analysis and synthesis of pose variations of human faces by a linear PCMAP model and its application for pose-invariant face recognition system** ». In Proceedings 4th IEEE International Conference on Automatic Face and Gesture Recognition, March 2000, pp. 142-149

[25] C. Zhang, F. S. Cohen. « **Face shape extraction and recognition using 3D morphing and distance mapping** ». Proceedings 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2000), March 2000, pp. 28-33

[26] K. Sengupta, W. Shiqin, C. C. Ko, P. Burman. « **Automatic face modeling from monocular image sequences using modified non parametric regression and an affine camera model** ». Proceedings 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2000), March 2000, pp. 524-529

[27] M. La Casia, S. Sclaroff. « **Fast, reliable head tracking under varying illumination : An approach based on registration of Texture-Mapped 3D models** ». IEEE Transactions On Pattern Analysis and Machine Intelligence, April 2000, vol. 22, no. 4, pp. 322- 336

[28] J. Ahlberg. « **Real-time facial feature tracking using an active model with fast image warping** ». Proceedings of the International Workshop on Very Low Bitrate Video (VLBV) (Athens, Greece), October 2001, pp. 39-43

[29] S. Shan, W. Gao, J. Yan, H. Zhang, X. Chen. « **Individual 3D face synthesis based on orthogonal photos and speech-driven facial animation** ». International Conference on Image Processing, 2000, vol. 3, pp. 238-241

[30] J. Ström, T. Jebara, S. Basu, A. Pentland. « **Real time tracking and modeling of faces : an EKF-based analysis by synthesis approach** ». IEEE International Workshop on Modelling People, September 20, 1999, pp. 55-61

[31] A. Pentland, B. Moghaddam, T. Starner, O. Oliyide, M. Turk. «**View-based and modular eigenspaces for face recognition** ». Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'94), 1994, pp. 84-91

[32] C-H. Lin, J-L. Wu. « **Automatic facial feature extraction by genetic algorithms** ». IEEE Transactions on Image Processing, June 1999, vol. 8, no. 6

[33] C. Zhang, F. S. Cohen. « **3-D face structure extraction and recognition from images using 3-D morphing and distance mapping** ». IEEE Transactions on Image Processing, November 2002, vol. 11, no. 11

[34] D. Nandy, J. Ben-Arie. « **Shape from recognition : A novel approach for 3-D face shape recovery** ». IEEE Transactions on Image Processing, February 2001, vol. 11, no. 2

[35] R. Brunelli, T. Poggio. « **Face recognition : features versus templates** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, October 1993, vol. 15, no. 10

[36] A. Azarbayejani, T. Starner, B. Horowitz, A. Pentland. « **Visually controlled graphics** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, June 1993, vol. 15, no. 6

[37] D. Terzopoulos, K. Waters. « **Analysis and synthesis of facial image sequences using physical and anatomical models** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, June 1993, vol. 15, no. 6

[38] H. Li, P. Roivainen, R. Forchheimer. « **3-D motion estimation in model-based facial image coding** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, June 1993, vol. 15, no. 6

[39] R. Koch. « **Dynamic 3-D scene analysis through synthesis feedback control** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, June 1993, vol. 15, no. 6

[40] N. Kruger. « **An algorithm for the learning of weights in discrimination functions using a priori constraints** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, July 1997, vol. 19, no. 7

[41] I. A. Essa, A. P. Pentland. « **Coding, analysis, interpretations, and recognition of facial expressions** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, July 1997, vol. 19, no. 7

[42] A. Lanitis, C. J. Taylor, T. F. Cootes. « **Automatic interpretation and coding of face images using flexible models** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, July 1997, vol. 19, no. 7

[43] Y. Akini, Y. Moses, S. Ullman. « **Face recognition : The problem of compensating for changes in illumination direction** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, July 1997, vol. 19, no. 7

[44] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman. « **Eigenfaces vs. fisherfaces : Recognition using class specific linear projection** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, July 1997, vol. 19, no. 7

[45] L. Wiskott, J-M. Fellous, N. Kruger, C. Von der Malsburg. « **Face recognition by elastic bunch graph matching** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, July 1997, vol. 19, no. 7

[46] M. Pardas. « **Extraction and tracking of the eyelids** ». International Conference on Acoustics, Speech and Signal Processing ICASSP 2000 (Istambul, Turkey), June 2000, vol. 4, pp. 2357-2360

[47] G. Donato, M. Stewart Bartlett, J. C. Hager, P. Ekman, T. J. Sejnowski. « **Classifying facial actions** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, October 1999, vol. 21, no. 10

[48] M. Pantic, J. M. Rothkrantz. « **Automatic analysis of facial expressions : The state of the art** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, December 2000, vol. 22, no. 12

[49] G. C. Feng, P. C. Yuen. « **Multi-cues eye detection on gray intensity image** ». Pattern Recognition, May 2001, vol. 34, no. 5, pp. 1033-1046

[50] S. Sirohey, A. Rosenfeld, Z. Duric. « **A method of detecting and tracking irises and eyelids in video** ». Pattern Recognition, June 2002, vol. 35, no. 6, pp. 1389-1401

[51] Y-L. Tian, T. Kanade, J. F. Cohn. « **Recognizing action units for facial expression analysis** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, February 2001, vol. 23, no. 2

[52] Y-J. Ryoo, N-C. Kim. « **Valley operator for extracting sketch features : DIP** ». Electronics Letters, 14th April 1988, vol. 24, no. 8, pp. 461-463

[53] R-S. Wang, Y. Wang. « **Facial feature extraction and tracking in video sequences** ». IEEE First Workshop on Multimedia Signal Processing (MMSP97) (Princeton, NJ), June 1997, pp. 233-238

[54] V. Blanz, T. Vetter. « **A morphable model for the synthesis of 3D faces** ». Proceedings of the 26th annual conference on Computer graphics and interactive techniques, 1999, pp. 187-194

[55] J-G. Ko, K-N. Kim, R. S. Ramakrishna. « **Facial feature tracking for eye-head controlled human computer interface** ». IEEE, TENCON'99 (Cheju, Korea), September 1999

[56] L. Shihong, Y. Sumi, M. Kawade, F. Tomita. « **Building 3D facial models and detecting face pose in 3D space** ». 2nd International Conference on 3-D Imaging and Modeling (3DIM '99) (Ottawa, Canada), October 04-08, 1999

[57] R. Liao, S. Z. Li. « **Face recognition based on multiple facial features** ». Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 239-245

[58] L. Zhang, P. Lenders. « **Knowledge-based eye detection for human face recognition** ». Fourth International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies, 2000, vol. 1, pp. 117-120

[59] A. A. Amini, T. E. Weymouth, R. C. Jain. « **Using dynamic programming for solving variational problems in vision** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, September 1990, vol. 12, no. 9

[60] W. Huang, B. Yin, C. Jiang, J. Miao. « **A new approach for eye feature extraction using 3D eye template** ». Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing (Hong Kong). May 2-4 2001, pp. 340-343

[61] A. Kapoor, R. W. Picard. « **Real-time, fully automatic upper facial feature tracking** ». Proceedings from 5th International Conference on Automatic Face and Gesture Recognition, May 2002, pp. 0010-0015

[62] Y-L. Tian, T. Kanade, J. F. Cohn. « **Dual-state parametric eye tracking** ». Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), March 2000, pp. 110-115

[63] M. U. Ramos Sanchez, J. Mantas, J. Kittler. « **Statistical chromaticity models for lip tracking with B-splines** ». In International Conference on Acoustics, Speech and Signal Processing, (Munich, Germany, April 21-24), 1997, vol. 4, pp. 2973-2976

[64] T. Darrell, M. Covell. « **Correspondence with cumulative similarity transforms** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, February 2001, vol. 23, no. 2

[65] J. Hou, R. H. Bamberger. « **Orientation selective operators for ridge, valley, edge, and line detection in imagery** ». IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1994

[66] Y. Moses, D. Reynard, A. Blake. « **Determining facial expressions in real time** ». Proceedings of the 5th International Conference on Computer Vision, 1995, pp. 296-301

[67] T. Kawaguchi, D. Hidaka, M. Rizon. « **Robust extraction of eyes from face** ». International Conference on Pattern Recognition (ICPR'00) (Barcelona, Spain), 2000, vol. 1, pp. 5109-5114

[68] J. Serrat, A. Lopez, D. Floret. « **On ridges and valleys** ». International Conference on Pattern Recognition (ICPR'00) (Barcelona, Spain), 2000, vol. 4, pp. 4059-4066

[69] T. J. Chown, P. H. Lewis. « **EFRM-based detection and extraction of ridge and valley features in grey level images** ». In Proceedings of Proc. IAPR Int. Conf. on Pattern Recognition C, 1992, pages 339-342

[70] J-I. Choi, C-W. La, P-K. Rhee, Y-L. Bae. « **Face and eye location algorithms for visual user interface** ». IEEE First Workshop on Multimedia Signal Processing, 1997, pp. 239-244

[71] Y. Zhong, A. K. Jain, M-P. Dubuisson-Jolly. « **Object tracking using deformable templates** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, May 2000, vol. 22, no. 5

[72] A. L. Yuille, P. W. Hallinan, D. S. Cohen. « **Feature extraction from faces using deformable templates** ». International Journal of Computer Vision, August 1992, vol. 8, no. 2, pp. 99-111

[73] L. Yin, A. Basu. « **Color-based mouth shape tracking for synthesizing realistic facial expressions** ». In Proceedings of International Conference on Image Processing, 2002, vol. 1, pp. 161-164

[74] A. R. Mirhosseini, C. Chen, D. M. Lam, H. Yan. « **A hierarchical and adaptive deformable model for mouth boundary detection** ». ICIP 1997, vol. 2, pp. 756-759

[75] V. Pahor, S. Carrato. « **A fuzzy approach to mouth corner detection** ». In Proceedings of International Conference on Image Processing, 1999, vol. 1, pp. 667-671

[76] L. Zhang. « **Estimation of the mouth features using deformable templates** ». International Conference on Image Processing (ICIP '97), 1997, vol. 3, pp. 328-331

[77] G. Rabi, S. W. Lu. « **Energy minimization for extracting mouth curves in a facial image** ». IASTED International Conference on Intelligent Information Systems (IIS '97), 1997, pp. 381-387

[78] M. Pantic, M. Tomc, L. J. M. Rothkrantz. « **A hybrid approach to mouth features detection** ». Proc. of the IEEE Int. Conf. on System, Man and Cybernetics (SMC) (Tucson, Arizona, USA), October 2001, pp. 1188-1193

[79] H. Jin, P. Favaro, S. Soatto. « **Real-time feature tracking and outlier rejection with changes in illumination** ». International Conference on Computer Vision (ICCV'01), 2001, vol. 1, pp. 684-688

[80] C. Xu, J. L. Prince. « **Snakes, shapes, and gradient vector flow** ». IEEE Transactions on Image Processing, March 1998, vol. 7, no. 3

[81] T. Tommasini, A. Fusiello, E. Trucco, V. Roberto. « **Making good features track better** ». Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (Santa Barbara, USA), June 1998, pp. 178-183

[82] A. Azarbayejani, A. P. Pentland. « **Recursive estimation of motion, structure, and focal length** ». IEEE Transactions on Pattern Analysis and Machine Intelligence, June 1995, vol. 17, no. 6

[83] J. Shi, C. Tomasi. « **Good features to track** ». IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94), 1994, pp. 593-600

[84] Niu Kegong. « **Estimation de posture et d'expressions faciales pour la vidéoconférence** ». Thèse présentée en vue de l'obtention du diplôme de maîtrise en sciences appliquées (M. Sc. A), École Polytechnique de Montréal, Département de Génie Électrique, Juin 2002

Annexe I

Analyse des résultats du suivi de points 2D

Sur la figure A.1, le suivi utilise seulement l'étape de la SSD. Cela constitue d'ailleurs l'algorithme utilisé par [84] dont les résultats devaient être améliorés dans ce présent ouvrage afin d'obtenir un suivi plus stable qu'auparavant.

La séquence utilisée va de la 1ère image à la 90e pour ensuite revenir à la première, ceci fait un total de 179 images parcourues. Les imprécisions causées par les erreurs cumulatives sont d'ailleurs très visibles sur la 179e image, soit la figure A.1.d). Sur cet exemple, les points 1 et 3 sont demeurés assez précis tandis que les autres le sont très peu. Le tableau A.1 illustre les résultats obtenus.

La figure A.2 utilise les mêmes points initiaux et la même séquence d'images que la figure A.1 mais le suivi a été effectué en utilisant l'algorithme proposé basé sur la SSD et la NCC. D'après la figure A.2.d), une bien meilleure précision est obtenue grâce à l'élimination de l'erreur cumulative. Le tableau A.2 illustre les résultats obtenus.

Numéro du point	21e image		66e image		179e image	
	Erreur en X	Erreur en Y	Erreur en X	Erreur en Y	Erreur en X	Erreur en Y
1	1	3	4	-2	2	0
2	-2	-1	-2	5	7	-1
3	-4	2	-2	6	1	-4
4	4	4	-2	7	-4	4
5	-1	4	-1	5	-5	4

Tableau A.1. Résultats de la première séquence avec la SSD simple.

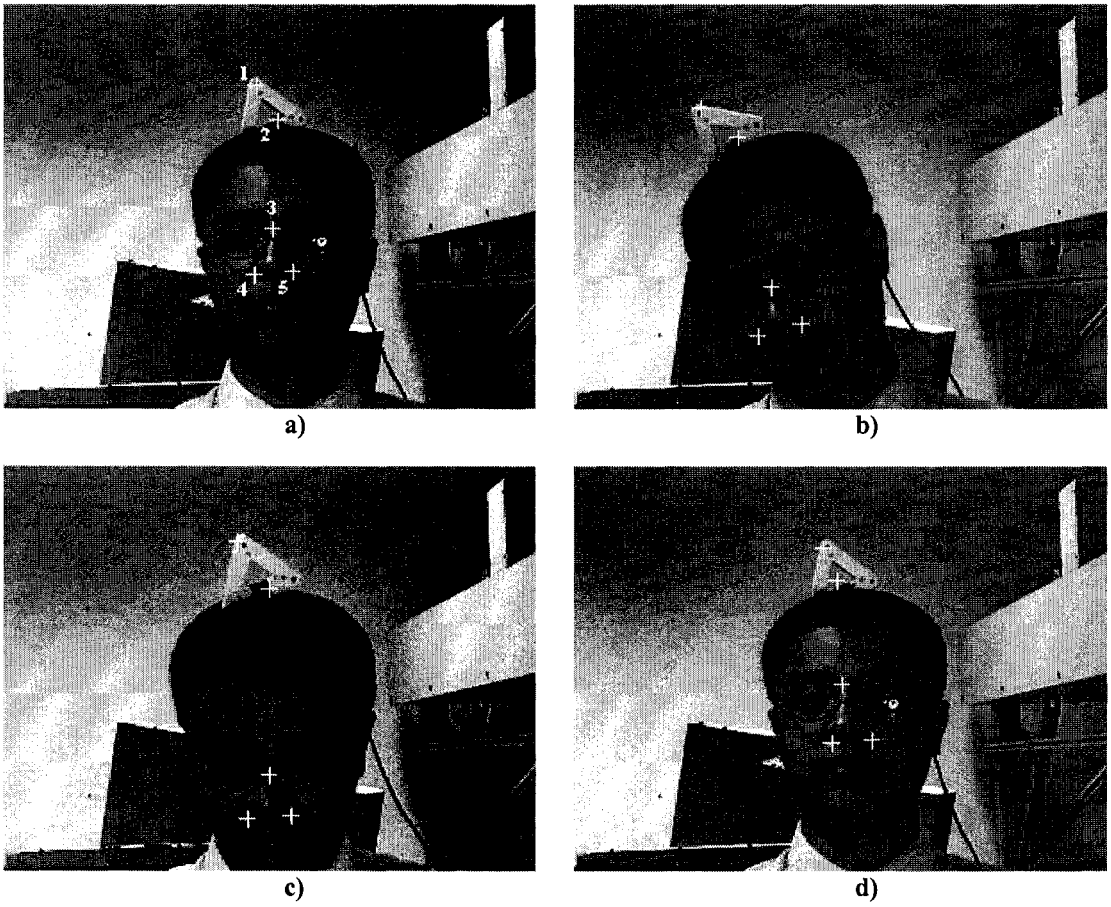


Figure A.1. Première utilisation simple du SSD. En a), les 5 points sur I_0. En b), ces points à la 21^e image. En c), ces points à la 66^e image. En d), ces points à la 179^e image.

Numéro du point	21^e image		66^e image		179^e image	
	Erreur en X	Erreur en Y	Erreur en X	Erreur en Y	Erreur en X	Erreur en Y
1	1	1	1	0	0	0
2	0	0	4	0	0	0
3	0	0	0	0	0	0
4	1	4	1	1	0	0
5	-1	2	-2	3	-1	-1

Tableau A.2. Résultats de le première séquence avec la SSD et la NCC.

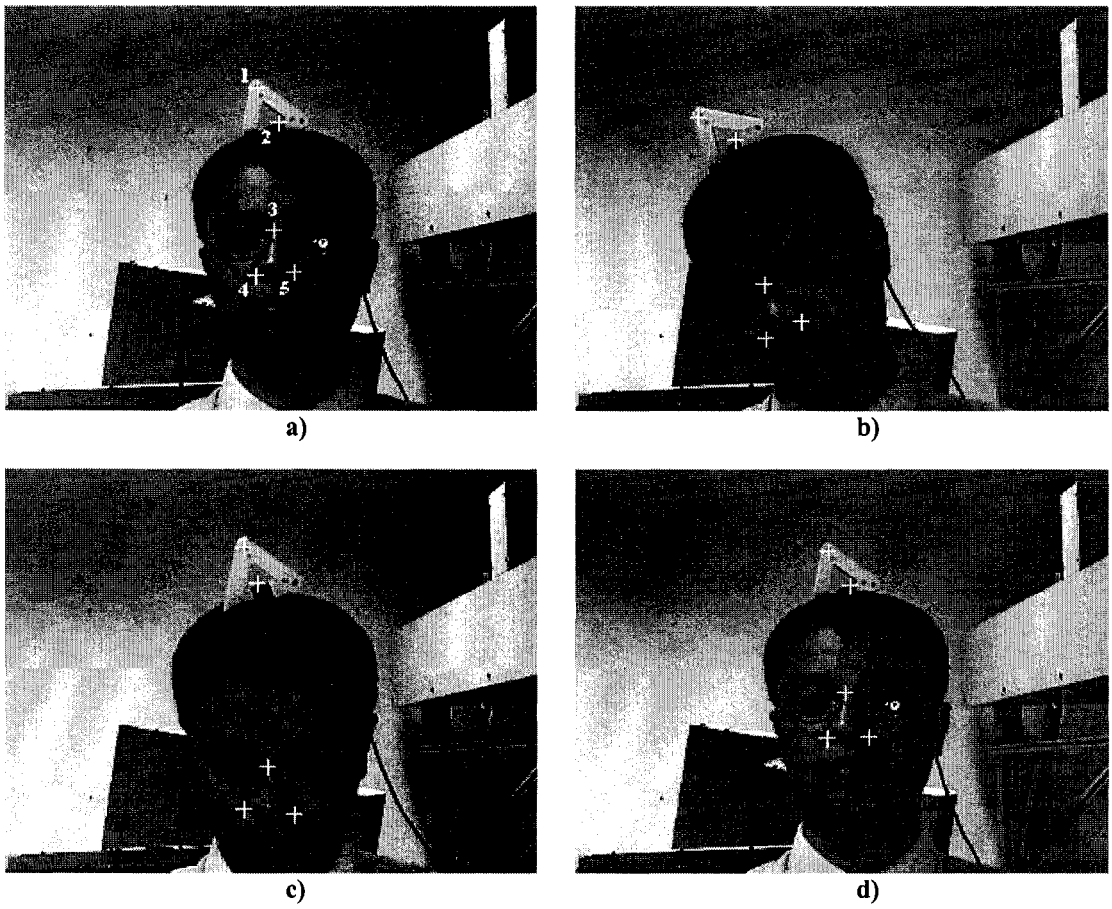


Figure A.2. Première utilisation de l'algorithme proposé utilisant le SSD et le NCC. En a), les 5 points sur I_0. En b), ces points à la 21^e image. En c), ces points à la 66^e image. En d), ces points à la 179^e image.

Les résultats obtenus au tableau A.2 sont beaucoup plus précis que ceux du tableau A.1. À la 179^e image (lors du retour à la 1ère image), toutes les positions sont parfaites sauf pour le point 5 où une très petite erreur a été obtenue.

Une analyse similaire a été effectuée à l'aide de 10 points sur une autre séquence. Sur la figure A.3, le suivi n'utilise que la SSD. La séquence va encore de la 1ère image à la 90^e pour ensuite revenir à la première. Sur la 179^e image, soit la figure A.3.d), les

imprécisions sont surtout obtenues sur les points 3, 6, 7 et 9. Le tableau A.3 illustre les résultats obtenus.

Numéro du point	26e image		55e image		179e image	
	Erreur en X	Erreur en Y	Erreur en X	Erreur en Y	Erreur en X	Erreur en Y
1	1	0	0	-2	-1	0
2	0	-1	6	-1	-2	-1
3	-2	3	5	3	3	9
4	1	-2	2	-1	-1	1
5	2	0	4	2	1	3
6	-1	1	4	-4	6	0
7	0	-1	2	1	-2	2
8	0	0	-2	4	2	0
9	-1	2	-2	4	-7	0
10	-1	-2	3	1	0	1

Tableau A.3. Résultats de la deuxième séquence avec la SSD simple.

Numéro du point	26e image		55e image		179e image	
	Erreur en X	Erreur en Y	Erreur en X	Erreur en Y	Erreur en X	Erreur en Y
1	0	-1	2	-1	0	0
2	0	0	1	1	0	0
3	0	0	0	1	0	0
4	-1	0	0	0	0	0
5	-2	-1	0	0	0	0
6	-1	-1	0	-3	0	0
7	1	2	2	1	0	0
8	0	0	-1	1	0	0
9	0	-1	1	0	0	0
10	1	0	1	0	0	0

Tableau A.4. Résultats de la deuxième séquence avec la SSD et la NCC.

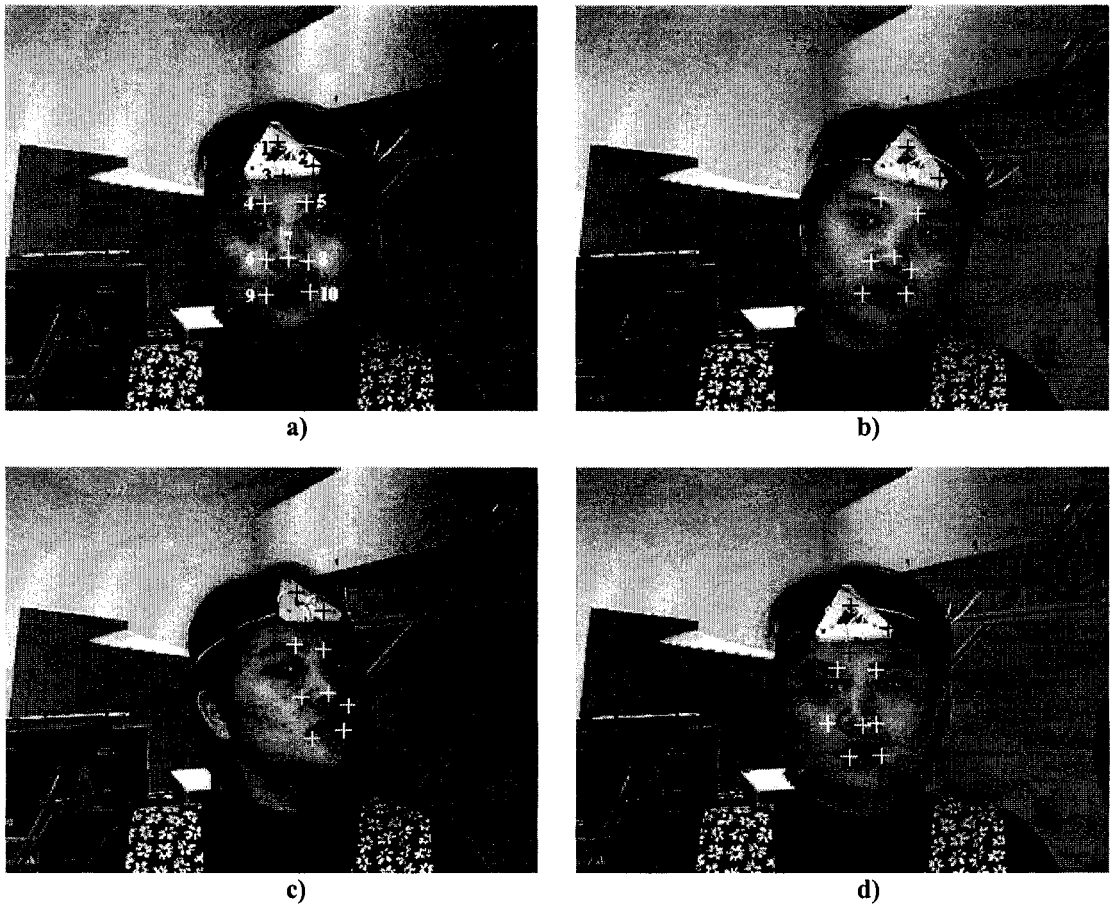


Figure A.3. Deuxième utilisation simple du SSD. En a), les 10 points sur I_0. En b), ces points à la 26^e image. En c), ces points à la 55^e image. En d), ces points à la 179^e image.

La figure A.4 utilise les mêmes points initiaux et la même séquence d'images que la figure A.3 mais le suivi a été effectué en utilisant l'algorithme proposé basé sur la SSD et la NCC. D'après la figure A.4.d), la précision est beaucoup plus grande et le tableau 3.4 illustre les résultats obtenus. À la 179^e image (lors du retour à la 1ère image), toutes les positions sont parfaites.

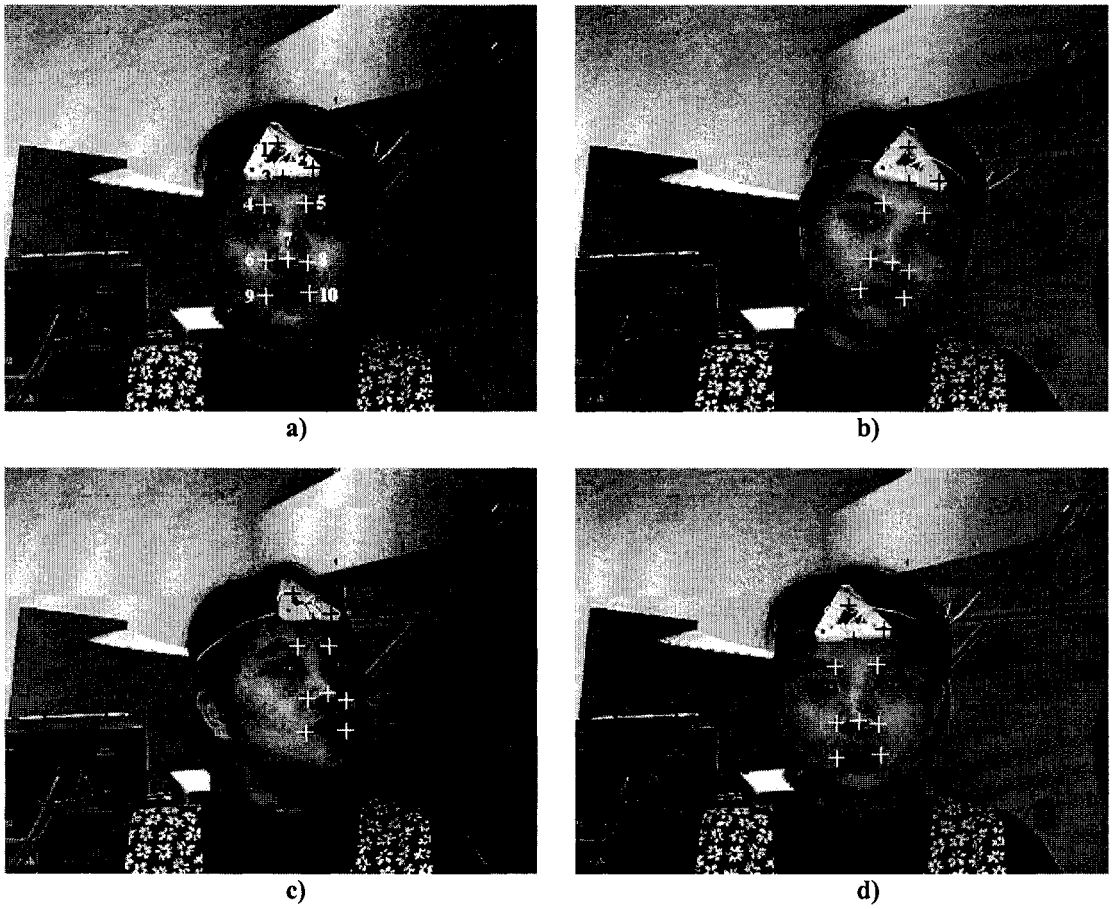


Figure A.4. Deuxième utilisation de l'algorithme proposé utilisant le SSD et le NCC. En a), les 10 points sur I_0. En b), ces points à la 26^e image. En c), ces points à la 55^e image. En d), ces points à la 179^e image.

Discussion des résultats

D'après les résultats obtenus, la méthode implantée permet un suivi très stable des points comparativement à une méthode n'utilisant que la SSD [84]. Cette amélioration est principalement due à une comparaison avec l'image initiale. Pour ce faire, la NCC a été utilisée mais des essais auraient pu aussi être effectués en utilisant de nouveau la SSD. Il semble cependant peu probable qu'une telle approche aurait donné d'aussi bons résultats lors de grandes variations d'intensité entre I_0 et I_t puisque la NCC est reconnue

pour corriger ce genre de faiblesse rencontrée avec la SSD. Il a d'ailleurs été noté que la vitesse d'exécution est suffisante avec la NCC, grâce aux algorithmes mentionnés et implantés, afin d'effectuer les traitements en temps réel sans compromettre d'une façon significative la qualité des résultats.

Des imprécisions peuvent survenir lorsqu'une région d'image est différente de celle de l'image précédente ou bien de la première image. Ces imprécisions sont souvent recorrigées et se produisent généralement lors d'occlusion (un des coins du nez devenant invisible par exemple). Ce suivi doit utiliser beaucoup de mémoire de l'ordinateur mais cela est largement compensé par le gain de rapidité. Cette méthode semble donc appropriée dans ce projet pour le suivi de plusieurs points sur de longues séquences d'images.

Annexe II

Analyse des résultats de l'estimation du mouvement rigide

L'analyse de l'estimation des mouvements rigides est qualitative et consiste à observer la configuration d'un axe tridimensionnelle sur les images de diverses séquences. Les objectifs sont atteints lorsque le mouvement estimé de la tête d'un usager est visuellement précis et stable à long terme.

L'analyse est effectuée sur 4 séquences d'images. Sur des images spécifiques de chaque séquence, des échantillons de résultats ont été récupérés permettant de bien visualiser les mouvements rigides. Les valeurs numériques des paramètres sont recueillies dans des tableaux et une discussion des résultats est fournie.

Figure A.5. Résultats du mouvement rigide pour la 1ère séquence. En a), 1ère image. En b), 30e image. En c), 45e image. En d), 60e image. En e), 75e image. En f), 115e image.

Figure A.6. Résultats du mouvement rigide pour la 2^{e} séquence. En a), 1ère image. En b), 25^{e} image. En c), 75^{e} image. En d), 95^{e} image. En e), 120^{e} image. En f), 175^{e} image.

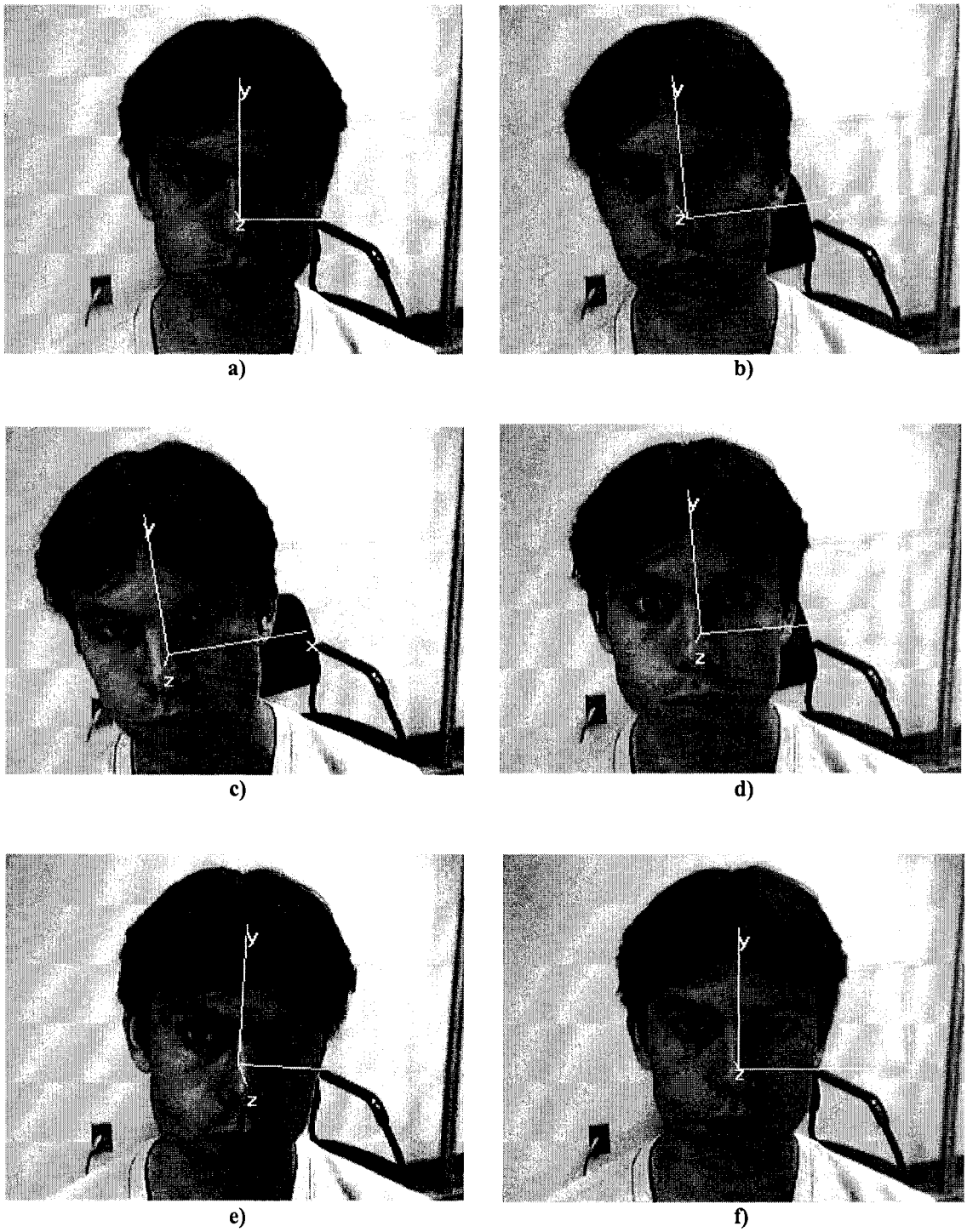


Figure A.7. Résultats du mouvement rigide pour la 3^{e} séquence. En a), 1ère image. En b), 20^{e} image. En c), 45^{e} image. En d), 90^{e} image. En e), 150^{e} image. En f), 200^{e} image.

Figure A.8. Résultats du mouvement rigide pour la 4^{e} séquence. En a), 1ère image. En b), 80^{e} image. En c), 110^{e} image. En d), 170^{e} image. En e), 205^{e} image. En f), 240^{e} image.

Première séquence :

Paramètres	Image 1	Image 30	Image 45	Image 60	Image 75	Image 115
t_x	0.000	-31.587	-45.844	-19.147	-1.503	0.046
t_y	0.000	-2.906	-3.286	2.328	-3.940	-0.585
t_z	0.000	0.000	0.000	0.000	0.000	0.000
θ_x	0.000	0.109	-0.095	-0.009	0.164	0.109
θ_y	0.000	0.224	0.469	0.354	0.095	0.051
θ_z	0.000	-0.328	-0.352	-0.129	-0.022	0.004

Tableau A.5. Paramètres obtenus pour la première séquence.

Deuxième séquence :

Paramètres	Image 1	Image 25	Image 75	Image 95	Image 120	Image 175
t_x	0.000	3.037	-5.530	-11.665	-10.693	-7.582
t_y	0.000	-6.196	-7.755	2.370	5.810	-9.541
t_z	0.000	0.000	0.000	0.000	0.000	0.000
θ_x	0.000	0.407	0.137	-0.108	0.046	-0.444
θ_y	0.000	-0.074	-0.085	0.128	0.109	-0.477
θ_z	0.000	0.030	0.011	-0.090	-0.049	-0.324

Tableau A.6. Paramètres obtenus pour la deuxième séquence.

Troisième séquence :

Paramètres	Image 1	Image 20	Image 45	Image 90	Image 150	Image 200
t_x	0.000	-31.095	-48.732	-24.276	2.107	1.718
t_y	0.000	-0.837	-12.159	-1.614	-2.869	-2.831
t_z	0.000	0.000	0.000	0.000	0.000	0.000
θ_x	0.000	0.083	-0.093	-0.036	-0.115	0.048
θ_y	0.000	0.096	0.031	0.019	-0.055	0.031
θ_z	0.000	-0.110	-0.165	-0.066	0.033	-0.004

Tableau A.7. Paramètres obtenus pour la troisième séquence.

Quatrième séquence :

Paramètres	Image 1	Image 80	Image 110	Image 170	Image 205	Image 240
t_x	0.000	-6.798	-7.579	-12.119	-15.671	-37.058
t_y	0.000	2.038	-2.168	-9.003	-4.140	1.913
t_z	0.000	0.000	0.000	0.000	0.000	0.000
θ_x	0.000	-0.047	-0.018	0.108	0.110	-0.157
θ_y	0.000	-0.079	-0.018	0.071	0.230	0.400
θ_z	0.000	0.196	-0.005	-0.065	-0.022	-0.176

Tableau A.8. Paramètres obtenus pour la quatrième séquence.

Discussion des résultats

D'après les essais effectués et une analyse qualitative, les résultats semblent stables pour de faibles mouvements de la tête. En effet, si les mouvements sont de trop grande amplitude, des points ne pourront pas être retrouvés pour le suivi, ceci nuit à l'estimation. Les paramètres t_x, t_y, θ_z semblent souvent plus précis que θ_x et θ_y d'après la configuration des axes sur les images. Cela peut-être causé par l'emplacement des points du suivi. Près des yeux, par exemple, les imprécisions sur les points pourraient causer de l'instabilité pour θ_y. Et les imprécisions sur les points près du nez pourraient causer de l'instabilité pour θ_x. Sur la 175^e image de la 2^e séquence, le mouvement estimé est totalement faux ; cela étant causé par le suivi erroné des points causé par les mouvements de trop grande amplitude de la tête. Sur la 240^e image de la 4^e séquence, les résultats sont cependant très bons, malgré la grande amplitude du mouvement selon θ_y et l'imprécision pour θ_x à la 205^e image. Étant donné le peu d'informations disponibles pour estimer le mouvement rigide, les résultats obtenus paraissent satisfaisants.

Le système manque cependant d'efficacité dans certaines situations rencontrées lors de l'expérimentation. Il a été remarqué que ces situations ont généralement au moins l'une des caractéristiques suivantes :

1. Une estimation imprécise des points 3D du visage à l'aide du modèle virtuel ;
2. Le suivi de points 2D est imprécis.

La première caractéristique a été rencontrée lorsque des usagers avaient les proportions du visage trop différentes du modèle virtuel : par exemple, le visage trop

allongé [84]. Dans d'autres cas on retrouve aussi une mauvaise pose initiale de l'usager, une mauvaise détection du centre des yeux, etc.

La deuxième caractéristique a été rencontrée lorsque les points 2D du suivi n'ont pas été positionnés initialement sur des endroits riches en texture, engendrant ainsi une instabilité. Des occlusions vis-à-vis des points du suivi, engendrées par de larges mouvements de la tête, ont quelquefois donné de mauvais résultats pour tout le reste de la séquence.

Annexe III

Analyse des résultats de la localisation des éléments non rigides

Pour bien commenter les résultats de ce présent ouvrage, voici les objectifs recherchés en terme de performances pour la localisation des éléments non rigides :

- Avoir une localisation précise des modèles (par exemple, que l'iris d'un modèle d'œil soit de même rayon et centré au même endroit que l'œil correspondant sur l'image) ;
- Bien adapter les modèles sur l'image initiale (ceci est surtout important pour les yeux et la bouche car de mauvais résultats auront une influence pour tout le reste de la séquence) ;
- Bien adapter les modèles aux diverses configurations probables (par exemple, un œil peut être ouvert ou fermé, il peut regarder dans plusieurs directions, etc) ;
- Bien adapter les modèles malgré le mouvement rigide de la tête (par exemple, l'image de la bouche perd de sa symétrie lorsque la tête est trop tournée et ceci peut affecter la localisation) ;
- Une stabilité à long terme (même si un élément est mal localisé sur une image, il est souhaitable que la précision de la localisation ne soit pas nécessairement décroissante en fonction du temps).

Sur chaque séquence, 18 images sont disponibles et les résultats ont été recueillis à chaque intervalle de 10 images pour ne pas alourdir inutilement, ceci correspond environ à chaque tiers de seconde.

Analyse de la première séquence

Dans cette séquence, dont 18 images sont présentées à la figure A.9, on retrouve un grand mouvement de la tête vers l'arrière, l'ouverture et la fermeture de la bouche et un mouvement des yeux. Initialement en a), le contour extérieur de la lèvre inférieure a été localisé trop bas à cause de la forte teneur en arêtes en cet endroit et le manque de visibilité de la lèvre. Cette imprécision affecte donc tout le reste de la séquence mais n'empêche pas de bien représenter l'ouverture de la bouche. En f), la tête de l'usager est beaucoup plus éloignée qu'au début mais le suivi est toujours précis. En i), l'usager commence à ouvrir la bouche et l'intérieur de la lèvre supérieure est mal localisé à cause de la forte teneur en vallées et en arêtes à l'intérieur de la bouche. Cette imprécision n'a cependant pas affectée les images suivantes où une très bonne précision est obtenue pour la bouche ouverte. En k), l'usager regarde vers la droite et les yeux ont bien été localisés. En m), le contour de la paupière supérieure est mal localisé étant donné l'imprécision obtenue sur l'ouverture de l'œil. Cette imprécision est causée par le reflet lumineux dans l'iris qui a engendré des arêtes indésirables sur l'image. En q), l'œil gauche est considéré fermé car l'iris ne forme plus un cercle bien visible, la détection a donc échoué. Sur l'œil droit, l'imprécision du contour a été causée par la faible teneur en vallées et par la mauvaise détection sur l'œil gauche. Il est à noter que les sourcils sont toujours bien localisés sur ces images. En général, un bon suivi des mouvements non rigides a été obtenu sur cette séquence.

a)

b)

c)

d)

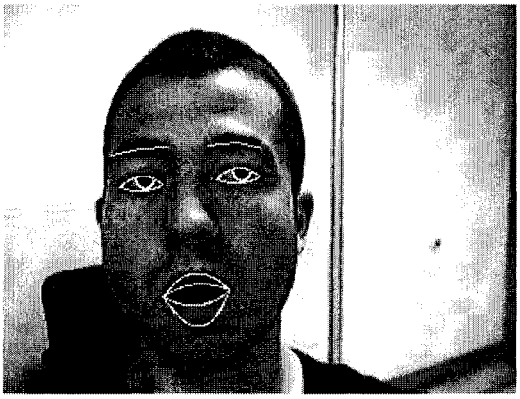
e)

f)

g)

h)

i)



j)

k)

l)

m)

n)

o)

p)

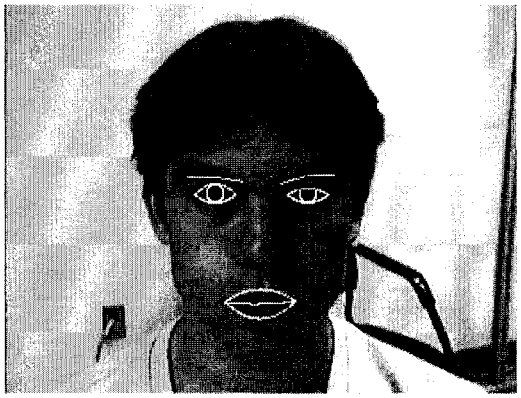
q)

r)

Figure A.9. Résultats sur la première séquence.

Analyse de la deuxième séquence

Dans cette séquence, dont 18 images sont présentées à la figure A.10, on retrouve un grand mouvement de la tête vers la gauche, l'ouverture et la fermeture de la bouche et un mouvement des yeux. Initialement en a), tous les éléments non rigides ont été bien localisés. En b), l'iris a été mal localisé étant donné le manque de visibilité causé par l'ombrage. Le contour de l'œil a donc été automatiquement affecté. En e), un grand mouvement de la tête vers la gauche a été effectué et le suivi est toujours précis sauf pour le contour de la paupière supérieure de l'œil gauche. Cela est causé par la présence du sourcil trop près de l'œil et le manque de visibilité. En f), le côté gauche du sourcil gauche est mal localisé à cause de la forte teneur en vallées dans l'œil et de l'inclinaison du sourcil par rapport à l'œil. En k), les yeux sont grands ouverts et la localisation est toujours précise sauf pour le côté droit du sourcil droit. Cela est causé par une inclinaison imprévue du sourcil. En n), les iris des yeux ne ressemblent pas assez à des cercles et n'ont pas pu être détectés. La localisation des yeux est donc très imprécise car le système a tenté de localiser des yeux fermés. Les sourcils sont donc directement affectés. Ces mauvaises localisations ont affecté tout le reste de la séquence, l'œil gauche a pu être relocalisé mais est trop large car l'œil droit est localisé trop loin. En p), les contours de la bouche sont mal localisés car les frontières entre les lèvres et la peau sont presque invisibles. Cela a affecté tout le reste de la séquence et a engendré d'étranges configurations de la bouche. En général, de bons résultats ont été obtenus jusqu'à m). Pour le reste de la séquence, les détections n'ont pas pu être replacées correctement.



a)

b)

c)

d)

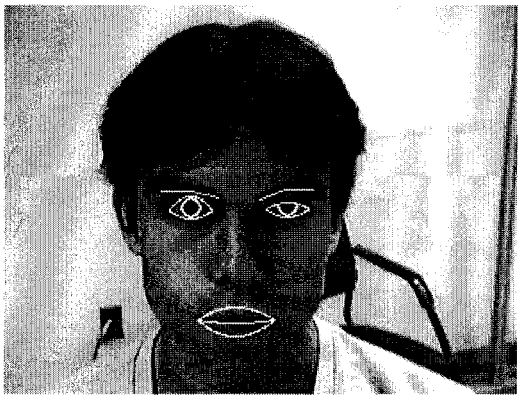
e)

f)

g)

h)

i)



j)

k)

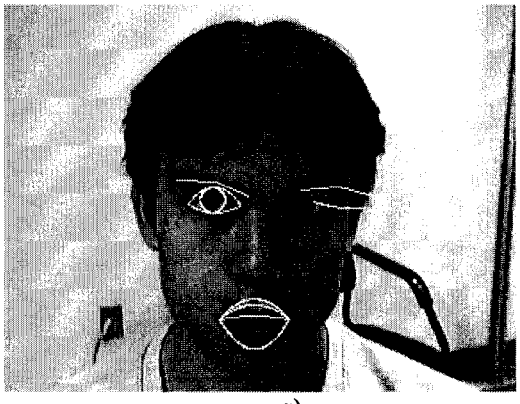
l)

m)

n)

o)

p)



q)

r)

Figure A.10. Résultats sur la deuxième séquence.

Analyse de la troisième séquence

Dans cette séquence, dont 18 images sont présentées à la figure A.11, on retrouve un grand mouvement de la tête vers la droite, un sourire et la fermeture des yeux. Initialement en a), le contour extérieur de la lèvre inférieure a été localisé trop bas à cause de la faible teneur en arêtes entre la peau et la lèvre. Cela affecte donc tout le reste de la séquence. En d), l'œil droit a été considéré fermé étant donné le manque de visibilité du cercle de l'iris. La position a tout de même été précise et l'œil gauche n'a donc pas été affecté. En f), il y a un grand mouvement de la tête vers la droite et la position des éléments non rigides est toujours bonne. Les cercles des deux iris sont peu visibles et les yeux ont donc été considérés fermés. Les sourcils sont bien localisés malgré qu'ils soient peu visibles. En g), l'iris de l'œil droit n'a pas été détecté et l'œil a une très mauvaise position car le système a tenté de localiser un œil fermé. Le sourcil droit a donc été automatiquement affecté ainsi que plusieurs images de la séquence. En k), l'ouverture de la bouche a été bien obtenue pendant que l'usager faisait un sourire. En m), l'iris droit n'est pas encore détecté mais la position de l'œil droit s'est améliorée. En o), le suivi de l'œil droit est de nouveau précis et permet d'obtenir de meilleurs résultats pour tout le reste de la séquence. En q), la fermeture des yeux est bien localisée. En général, de bons résultats ont été obtenus pour le contour intérieur de la bouche et les sourcils mais de mauvaises localisations ont souvent été rencontrées sur l'œil droit.

a)

b)

c)

d)

e)

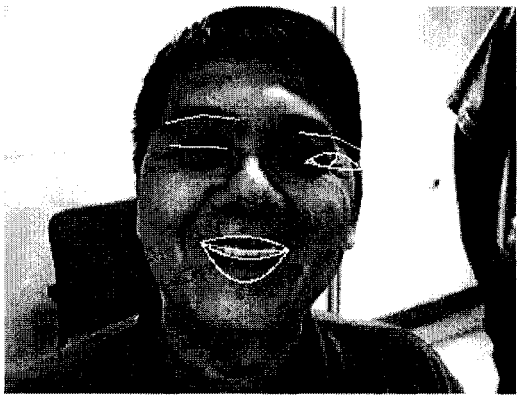
f)

g)

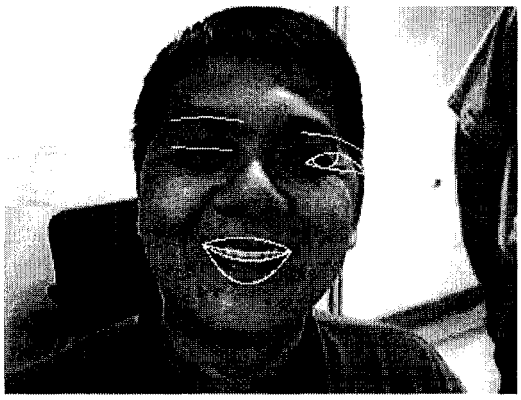
h)

i)

j)



k)



l)

m)

n)

o)

p)

q)

r)

Figure A.11. Résultats sur la troisième séquence.

Analyse de la quatrième séquence

Dans cette séquence, dont 18 images sont présentées à la figure A.12, on retrouve un grand mouvement de la tête vers la gauche, un mouvement des yeux, l'ouverture et la fermeture de la bouche et des yeux. Initialement en a), tous les éléments non rigides ont été bien localisés. Sur ces images, de très bons résultats ont été obtenus malgré le manque de visibilité des cercles des iris en f). Des imprécisions mineures, surtout au centre de la bouche comme en f), ont cependant été rencontrées.

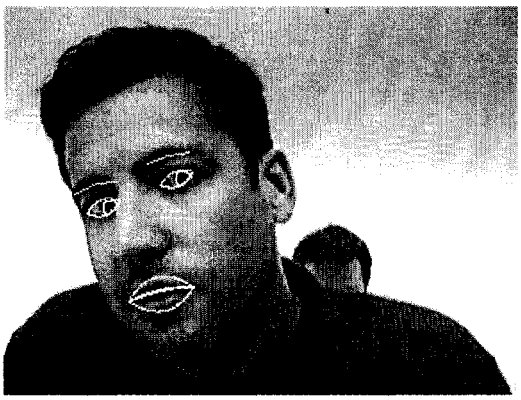
a)

b)

c)

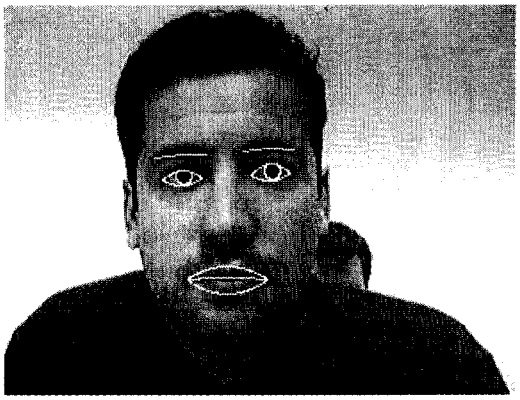
d)

e)

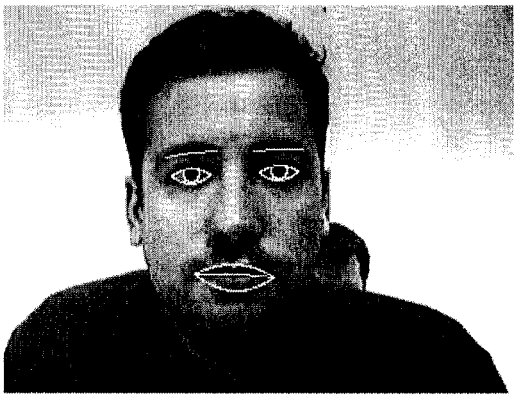


f)

g)

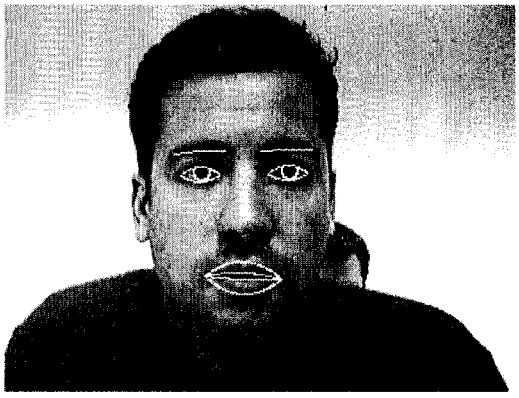


h)

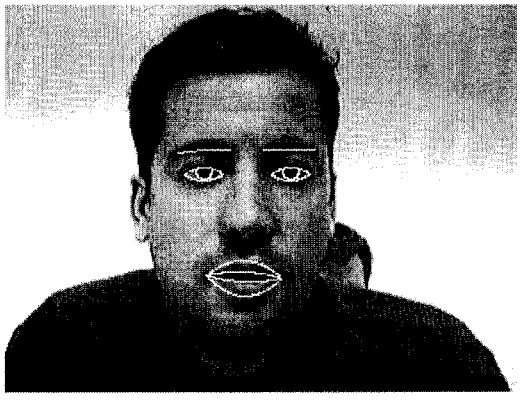


i)

j)



k)



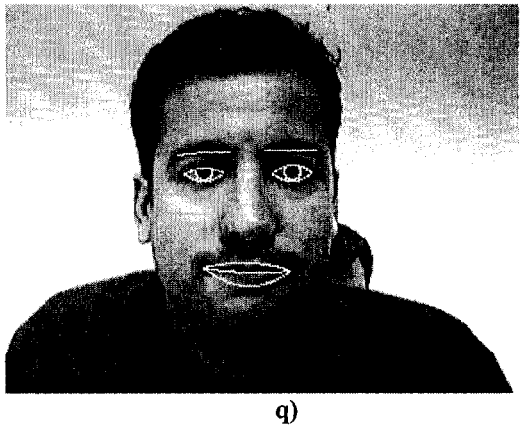
l)

m)

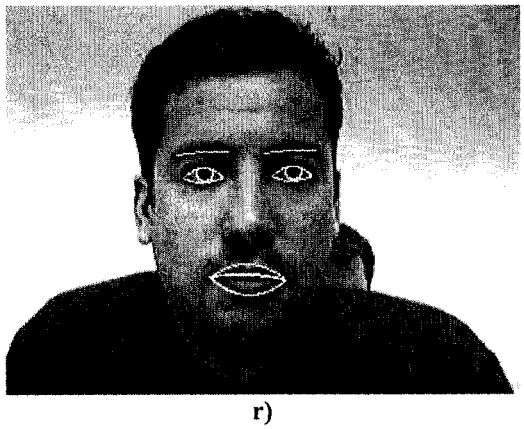
n)

o)

p)



q)

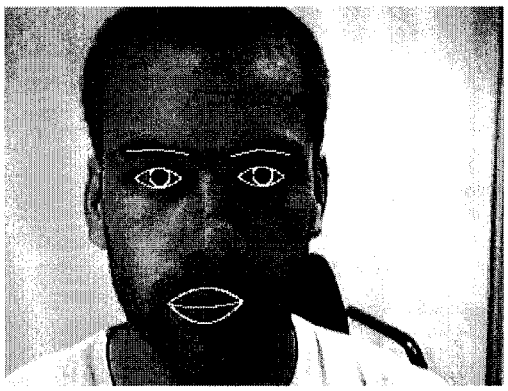


r)

Figure A.12. Résultats sur la quatrième séquence.

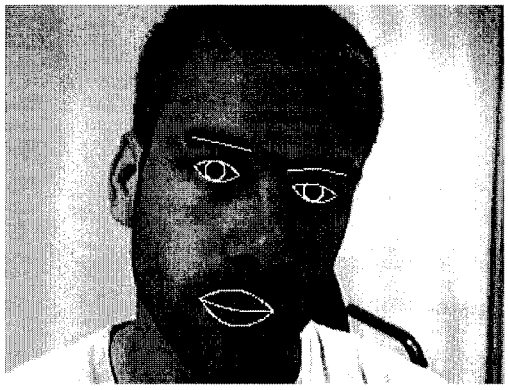
Analyse de la cinquième séquence

Dans cette séquence, dont 18 images sont présentées à la figure A.13, on retrouve diverses inclinaisons de la tête. Initialement en a), le coin droit de la bouche a été mal localisé à cause de la présence de la barbe, ceci a engendré des vallées non désirées. Cela a affecté tout le reste de la séquence, provoquant une mauvaise localisation du contour de la bouche. En b), d), f), g), p) et r), les images des contours des iris sont trop floues, ceci a diminué les arêtes et certains iris n'ont pas été détectés. Le système a donc tenté de positionner des fermetures d'œil. En f), de fortes vallées ont été rencontrées près du contour supérieur de l'œil gauche et le sourcil gauche a donc mal été localisé. En p), à cause de l'inclinaison de la tête, le sourcil gauche est trop éloigné vers la gauche de l'œil et n'a pas été localisé correctement : la concentration d'arêtes au centre du sourcil n'a pas été trouvée. En q), les yeux sont fermés et les fermetures ont été bien localisées. En général, les résultats sur cette séquence étaient très imprécis à cause de la mauvaise localisation initiale et des images trop floues vis-à-vis des yeux. Cette séquence illustre bien l'impact d'une mauvaise localisation initiale des coins de la bouche. Lorsqu'un coin est localisé sur la fermeture de la bouche au lieu du vrai coin, le suivi de ce coin est instable à cause de la mauvaise portion d'images retenue initialement. Contrairement aux yeux, la bouche ne possède pas un système de correction pour les coins.



a)

b)

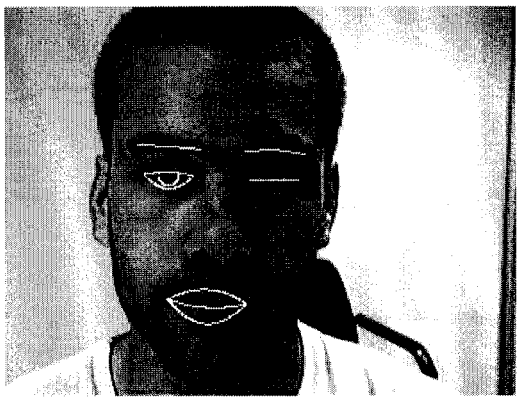


c)

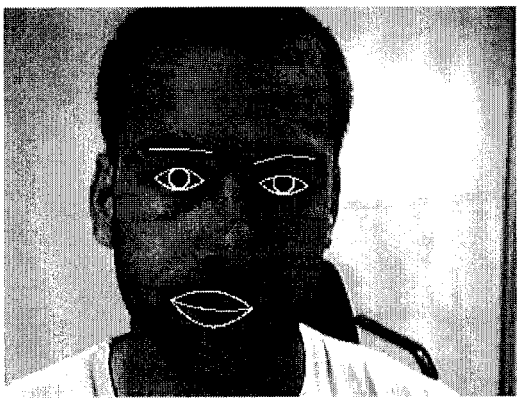
d)

e)

f)



g)

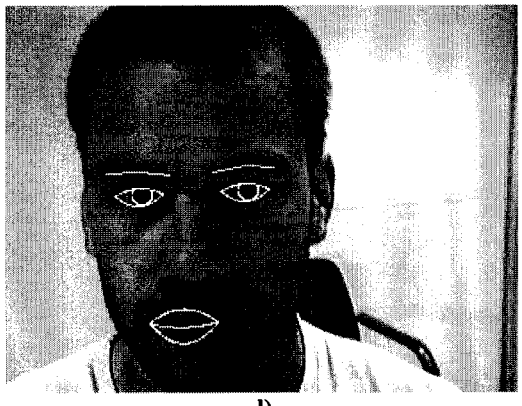


h)

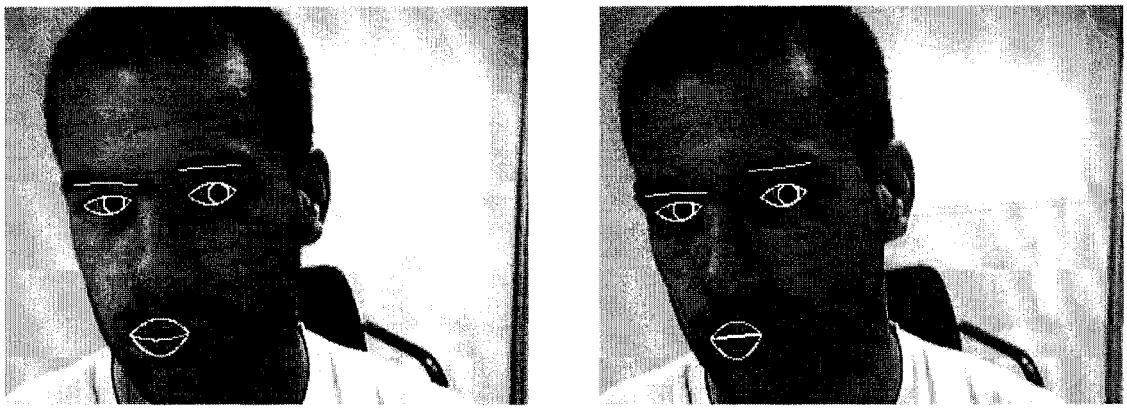
i)

j)

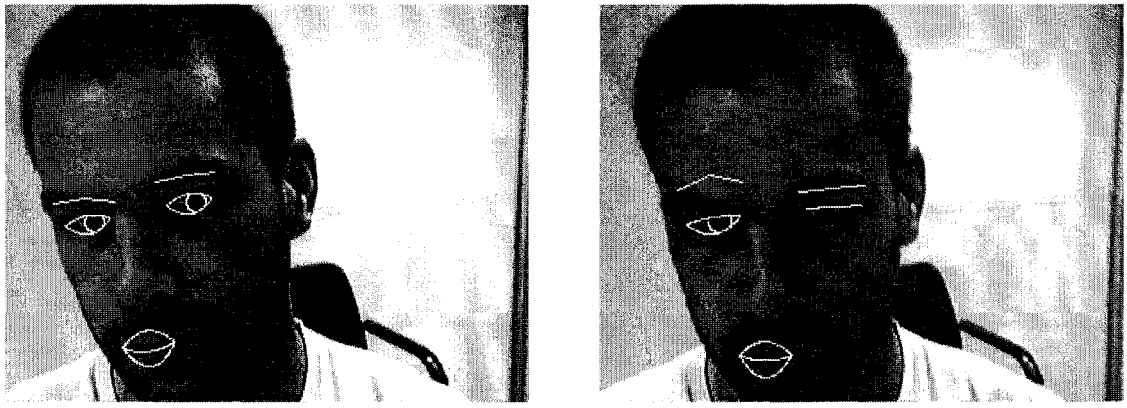
k)



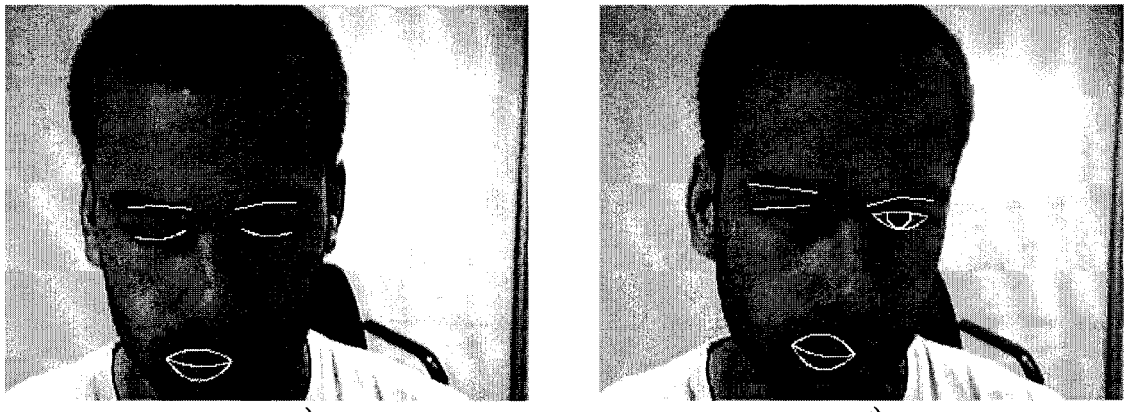
l)



m) n)



o) p)



q) r)

Figure A.13. Résultats sur la cinquième séquence.

Annexe IV

Intégration numérique de courbes ou surfaces sur les images

Pendant ce projet, plusieurs intégrations de courbes sur diverses images ont dû être effectuées. Lors des descriptions de certaines méthodes utilisées pour les détections ou les suivis, les intégrations des courbes étaient illustrées d'une façon plutôt symbolique afin d'en simplifier la représentation et d'éviter la redondance dans les calculs. Ces intégrations ont surtout été utilisées pour l'évaluation des mouvements non rigides du visage, c'est-à-dire sur les modèles géométriques utilisés pour localiser ou effectuer le suivi des yeux, des sourcils et de la bouche. Dans cette section, une description plus détaillée est effectuée afin de mieux expliquer comment ces intégrations ont été implantées. Pour chaque type d'intégration, la représentation symbolique sera illustrée ainsi que l'algorithme réel effectué à l'aide de traitements numériques. Il est à noter que toutes les intégrations sont normalisées en fonction des longueurs des courbes puisque dans ce projet, les intégrations étaient utilisées afin de déterminer des concentrations (vallées, sommets, gradients, etc.) et les résultats obtenus doivent donc être invariants selon la longueur des courbes. De plus, les intégrations implantées sont approximatives puisque les images sont discrètes et qu'il peut être difficile d'implanter des méthodes très précises demandant peu de traitements. Cependant, les intégrations obtenues sont assez précises pour que la perte de précision soit négligeable. Lors des implantations détaillées, la fonction "arrondie(α)" signifie que la valeur de α est arrondie à la valeur entière la plus proche.

Intégration d'un segment de droite

Cette intégration a principalement été utilisée pour les sourcils. L'équation A.1 illustre la représentation symbolique de l'intégration d'un segment de droite S sur une image I tandis que l'équation A.2 décrit l'implantation détaillée.

$$score = \int_S I(\vec{p}) \cdot \overrightarrow{dp} \tag{A.1}$$

Où $\vec{p}$ est un vecteur de position sur S. Voici maintenant l'implantation détaillée :

$$IntegSeg(x1, y1, x2, y2, I) \tag{A.2}$$

```
{
    acc = 0
    nb_pt = 0
    Δx = x2 − x1
    Δy = y2 − y1
    L = √(Δx² + Δy²)
    u.x = Δx / L
    u.y = Δy / L
    pas = 1
    Pour  i = 0  jusqu'à  L,  i = i + pas
    {
        v.x = i·u.x
        v.y = i·u.y
        nb_pt = nb_pt + 1
        acc = acc + I(arrondie(x1 + v.x), arrondie(y1 + v.y))
    }
    score = acc / nb_pt
}
```

Où (*x1*, *y1*) et (*x2*, *y2*) sont les deux points qui définissent S et *I* est l'image utilisée.

Intégration de l'aire d'un cercle

Cette intégration a principalement été utilisée pour l'iris des yeux. L'équation A.3 illustre la représentation symbolique de l'intégration de l'aire *A* d'un cercle sur une image *I* tandis que l'équation A.4 décrit l'implantation détaillée.

$$score = \int_A I(\vec{p}) \cdot \overrightarrow{dp} \tag{A.3}$$

Où $\vec{p}$ est un vecteur de position sur *A*. Voici maintenant l'implantation détaillée :

$$IntegAireCercle(x, y, r, I) \tag{A.4}$$

```
{
    acc = 0
    nb_pt = 0
    x1 = x − r
    y1 = y − r
    x2 = x + r
    y2 = x + r
    Pour  i = x1  jusqu'à  x2,  i = i + 1
    {
        Pour  j = y1  jusqu'à  y2,  j = j + 1
        {
            Δx = i − x
            Δy = j − y
            Si  (Δx² + Δy² ≤ r²)  alors
            {
                nb_pt = nb_pt + 1
                acc = acc + I(arondie(i), arondie(j))
            }
        }
```

}

$$score = \frac{acc}{nb\_pt}$$

}

Où x, y, r et I sont la coordonnée du centre, le rayon et l'image utilisés respectivement.

Intégration de la circonférence d'un cercle

Cette intégration a principalement été utilisée pour l'iris des yeux. L'équation A.5 illustre la représentation symbolique de l'intégration de la circonférence C d'un cercle sur une image I tandis que l'équation A.6 décrit l'implantation détaillée.

$$score = \int_C I(\vec{p}) \cdot \overrightarrow{dp} \tag{A.5}$$

Où $\vec{p}$ est un vecteur de position sur C. Voici maintenant l'implantation détaillée :

$$IntegCirconCercle(x, y, r, I) \tag{A.6}$$

{

$acc = 0$

$nb\_ang = \lceil 2 \cdot \pi \cdot r \rceil$

$\Delta\theta = \frac{2 \cdot \pi}{nb\_ang}$

$Pour \quad \theta = \Delta\theta \quad jusqu'à \quad 2 \cdot \pi, \quad \theta = \theta + \Delta\theta$

{

$x1 = x + r \cdot \cos(\theta)$

$y1 = y + r \cdot \sin(\theta)$

$acc = acc + I(arondie(x1), arondie(y1))$

}

$$score = \frac{acc}{nb\_ang}$$

}

Où x, y, r et I sont la coordonnée du centre du cercle, le rayon et l'image utilisés respectivement.

Intégration d'une parabole

Cette intégration a principalement été utilisée pour le contour des yeux et de la bouche. L'équation A.7 illustre la représentation symbolique de l'intégration d'une parabole P sur une image I tandis que l'équation A.8 décrit l'implantation détaillée. Il est à noter que les segments de paraboles sont symétriques.

$$score = \int_P I(\vec{p}) \cdot \vec{dp} \tag{A.7}$$

Où $\vec{p}$ est un vecteur de position sur P. Voici maintenant l'implantation détaillée :

$$IntegParabole(x1, y1, x2, y2, h, I) \tag{A.8}$$

{

$acc = 0$

$nb\_pt = 0$

$\Delta x = x2 - x1$

$\Delta y = y2 - y1$

$L = \sqrt{\Delta x^2 + \Delta y^2}$

$\theta = \arcsin\left(\frac{\Delta y}{L}\right)$

$c\_ang = \cos(\theta)$

$s\_ang = \sin(\theta)$

$a = \frac{4 \cdot h}{L^2}$

pas = 1

$Pour \quad x = -\frac{L}{2} \quad jusqu'à \quad \frac{L}{2}, \quad x = x + pas$

{

$$nb\_pt = nb\_pt + 1$$
$$y = -a \cdot x^2 + h$$
$$vx.x = x \cdot c\_ang$$
$$vx.y = x \cdot s\_ang$$
$$vy.x = -y \cdot s\_ang$$
$$vy.y = y \cdot c\_ang$$
$$acc = acc + I(arondie(x + vx.x + vy.x), arondie(y + vx.y + vy.y))$$

}

$$score = \frac{acc}{nb\_pt}$$

}

Où $(x1, y1)$ et $(x2, y2)$ représentent les deux points aux extrémités de la courbe de P tandis que h et I représentent la hauteur de P et l'image utilisée respectivement.

Intégration de l'aire obtenue de l'intersection entre un cercle et deux paraboles

Cette intégration a principalement été utilisée pour la portion d'aire visible de l'iris selon la direction du regard et l'ouverture de l'oeil. L'équation A.9 illustre la représentation symbolique de l'intégration de l'aire *AI1* sur un cercle entre deux paraboles se croisant en deux points tandis que l'équation A.10 décrit l'implantation détaillée. Il est à noter que les segments de paraboles sont symétriques.

$$score = \int_{AI1} I(\vec{p}) \cdot \overrightarrow{dp} \tag{A.9}$$

Où $\vec{p}$ est un vecteur de position sur *AI1*. Voici maintenant l'implantation détaillée:

$IntegAireCercleParaboles1(x1, y1, x2, y2, hu, hd, xc, yc, r, I)$ (A.10)

{

$acc = 0$

$nb\_pt = 0$

$\Delta x = x2 - x1$

$\Delta y = y2 - y1$

$xp = \dfrac{\Delta x}{2}$

$yp = \dfrac{\Delta y}{2}$

$Lp = \sqrt{\Delta x^2 + \Delta y^2}$

$Lc = \sqrt{(x - xp)^2 + (y - yp)^2}$

$\theta 1 = \arcsin\left(\dfrac{\Delta y}{Lp}\right)$

$\theta 2 = \arcsin\left(\dfrac{y - yp}{Lc}\right)$

$c\_ang = \cos(\theta 1)$

$s\_ang = \sin(\theta 1)$

$xc = Lc \cdot \cos(\theta 2 - \theta 1)$

$yc = Lc \cdot \sin(\theta 2 - \theta 1)$

$au = \dfrac{4 \cdot hu}{L^2}$

$ad = \dfrac{4 \cdot hd}{L^2}$

$\text{pas} = 1$

$Pour \quad i = -\dfrac{L}{2} \quad jusqu'à \quad \dfrac{L}{2}, \quad i = i + pas$

{

$yu = -au \cdot i^2 + hu$

$yd = -ad \cdot i^2 + hd$

$Pour \quad j = yd \quad jusqu'à \quad yu, \quad j = j + pas$

{

$Si \quad \left((i - xc)^2 + (j - yc)^2 \le r^2\right) \quad alors$

{

$$nb\_pt = nb + pt + 1$$
$$vx.x = i \cdot c\_ang$$
$$vx.y = i \cdot s\_ang$$
$$vy.x = -j \cdot s\_ang$$
$$vy.y = j \cdot c\_ang$$
$$pt.x = arondie(x + vx.x + vy.x)$$
$$pt, y = arondie(y + vx.y + vy.y)$$
$$acc = acc + I(pt.x, pt.y)$$

}

}

$$score = \frac{acc}{nb\_pt}$$

}

Où (*x1*, *y1*) et (*x2*, *y2*) représentent les deux points communs aux extrémités des deux paraboles et hu et hd représentent les hauteurs de ces paraboles. Les paramètres (*xc*, *yc*) et *r* représentent le centre et le rayon du cercle respectivement tandis que *I* est l'image utilisée.

Intégration de l'aire obtenue entre deux paraboles et à l'extérieur d'un cercle

Cette intégration a principalement été utilisée pour la portion d'aire visible du blanc de l'oeil selon la direction du regard et l'ouverture de l'oeil. L'équation A.11 illustre la représentation symbolique de l'intégration de l'aire *AI2* entre deux paraboles se croisant en deux points et à l'extérieur d'un cercle tandis que l'équation A.12 décrit l'implantation détaillée. Il est à noter que les segments de paraboles sont symétriques.

$$score = \int_{AI2} I(\vec{p}) \cdot \overrightarrow{dp} \qquad \text{(A.11)}$$

Où $\vec{p}$ est un vecteur de position sur *AI2*. Voici maintenant l'implantation détaillée:

$IntegAireCercleParaboles2(x1, y1, x2, y2, hu, hd, xc, yc, r, I)$ (A.12)

{

$acc = 0$

$nb\_pt = 0$

$\Delta x = x2 - x1$

$\Delta y = y2 - y1$

$xp = \frac{\Delta x}{2}$

$yp = \frac{\Delta y}{2}$

$Lp = \sqrt{\Delta x^2 + \Delta y^2}$

$Lc = \sqrt{(x - xp)^2 + (y - yp)^2}$

$\theta 1 = \arcsin\left(\frac{\Delta y}{Lp}\right)$

$\theta 2 = \arcsin\left(\frac{y - yp}{Lc}\right)$

$c\_ang = \cos(\theta 1)$

$s\_ang = \sin(\theta 1)$

$xc = Lc \cdot \cos(\theta 2 - \theta 1)$

$yc = Lc \cdot \sin(\theta 2 - \theta 1)$

$au = \frac{4 \cdot hu}{L^2}$

$ad = \frac{4 \cdot hd}{L^2}$

$\text{pas} = 1$

$Pour \quad i = -\frac{L}{2} \quad jusqu'à \quad \frac{L}{2}, \quad i = i + pas$

{

$yu = -au \cdot i^2 + hu$

$yd = -ad \cdot i^2 + hd$

$Pour \quad j = yd \quad jusqu'à \quad yu, \quad j = j + pas$

{

$Si \quad \left((i-xc)^2+(j-yc)^2 > r^2\right) \quad alors$

{

$nb\_pt = nb + pt + 1$

$vx.x = i \cdot c\_ang$

$vx.y = i \cdot s\_ang$

$vy.x = -j \cdot s\_ang$

$vy.y = j \cdot c\_ang$

$pt.x = arondie(x + vx.x + vy.x)$

$pt, y = arondie(y + vx.y + vy.y)$

$acc = acc + I(pt.x, pt.y)$

}

}

$$score = \frac{acc}{nb\_pt}$$

}

}

Où (*x1*, *y1*) et (*x2*, *y2*) représentent les deux points communs aux extrémités des deux paraboles et hu et hd représentent les hauteurs de ces paraboles. Les paramètres (*xc*, *yc*) et *r* représentent le centre et le rayon du cercle respectivement tandis que *I* est l'image utilisée.

Annexe V

Calculer les images des vallées, des sommets et des arêtes

Les images des vallées et des sommets ont souvent été mentionnées dans cet ouvrage et étaient nommées $I_{\text{vallées}}$ et I_{sommets} respectivement. Une vallée est une région foncée dans l'image qui est entourée d'une région pâle. Les sourcils, par exemple, forment de fortes vallées puisque ces derniers sont plus foncés que la peau qui les entoure. Un sommet est l'opposé d'une vallée, c'est-à-dire une région pâle dans l'image qui est entourée d'une région foncée. Le blanc des yeux, par exemple, forme de forts sommets. Plusieurs méthodes ont été étudiées pour calculer les vallées et les sommets [52][53][65][68] et une version légèrement modifiée de [53] a été utilisée afin d'être mieux adaptée aux images utilisées. Un score de vallée est calculé de la façon suivante :

$$ScoreVallée(x,y) = \max_{k=1 \text{ à } \Delta} \left\{ \begin{array}{l} dU(x,y,k) + dR(x,y,k) \\ -\alpha 1 \cdot \left| dU(x,y,k) - dD(x,y,k) \right| - \alpha 2 \cdot \left| dR(x,y,k) - dL(x,y,k) \right| \end{array} \right\}$$

Où
$$\begin{aligned} dU(x,y,k) &= I(x,y-k) - I(x,y) \\ dD(x,y,k) &= I(x,y+k) - I(x,y) \\ dR(x,y,k) &= I(x+k,y) - I(x,y) \\ dL(x,y,k) &= I(x-k,y) - I(x,y) \end{aligned}$$

Pour de bons résultats, les valeurs Δ, $\alpha 1$ et $\alpha 2$ ont été fixées à 5, 1 et 1 respectivement. Les sommets sont obtenus de la façon suivante qui est très similaire :

$$ScoreSommet(x,y) = \max_{k=1 \text{ à } \Delta} \left\{ dR(x,y,k) - \alpha 3 \cdot \left| dR(x,y,k) - dL(x,y,k) \right| \right\}$$

Pour de bons résultats, les valeurs Δ et $\alpha 3$ ont été fixées à 5 et 1 respectivement. Les valeurs obtenues sur l'image I sont donc insérées à la position (x,y) correspondantes dans $I_{\text{vallées}}(x,y)$ ou $I_{\text{sommets}}(x,y)$ selon si ce sont les vallées ou les sommets qui sont calculés.

Pour obtenir l'image des arêtes $I_{\text{arêtes}}$, l'algorithme de Canny est utilisé. Cet algorithme consiste à calculer les gradients dans l'image I et ensuite à parcourir les frontières engendrées par de forts changements de gradients. Plusieurs autres techniques avaient été utilisées (filtres de Sobel, etc.) mais Canny semblait plus efficace pour mettre en évidence le contour du bas de l'iris. Et c'est seulement pour traiter ce cas que les arêtes sont utilisées dans ce présent ouvrage. Pour de bons résultats, les coefficients σ, *tl* et *th* de Canny ont été fixés à 1.5, 0.4 et 0.8 respectivement. Le coefficient σ est utilisé pour la déviation standard du filtre passe-bas initialement appliqué sur l'image tandis que *tl* et *th* permettent de définir les frontières d'un filtre d'hystérésis permettant d'éliminer un voisinage de pixels. Ceci permet de contrôler la concentration d'arêtes dans l'image.

Les images $I_{vallées\_bin}$ et $I_{sommets\_bin}$ sont souvent utilisées dans ce projet et correspondent à des images binaires obtenues après un seuillage sur $I_{\text{vallées}}$ et I_{sommets} respectivement. Le seuillage varie selon l'utilisation.

Annexe VI

Étude sur l'anthropométrie du visage

Dans [2], une description statistique tridimensionnelle de la structure du visage est donnée. Trois types de visage ont été étudiés : 24 Asiatiques, 29 Noirs et 29 Blancs. D'après l'article, la tête et le visage sont probablement les régions du corps les plus difficiles à décrire en utilisant les techniques de mesures traditionnelles. La raison provient du fait que les diverses dimensions varient grandement d'un individu à l'autre et sont très peu corrélées. Pour obtenir cette description statistique, l'approche consiste en 3 étapes.

Étape 1 : l'échelle ostéométrique

Consiste à décrire la transformation des mesures d'un spécimen par rapport à celles d'un autre en analysant les variations des proportions.

Étape 2 : l'accumulation des spécimens normalisés

Consiste à construire, grâce à la première étape, un modèle normalisé afin de contenir les mesures selon des moyennes et des déviations standard. S'il y a trop de variation, il est préférable que les sujets soient groupés en plusieurs catégories.

Étape 3 : le test statistique

Consiste à déterminer les ressemblances entre les différentes catégories. Les mesures sur les individus furent prises en fonction de leurs utilités dans la construction d'accessoires divers tels des casques, des masques, etc.

Dans cet article, les différences entre les catégories d'individus sont d'ailleurs analysées de la façon suivante :

- 26 points de mesures sur la tête et le visage.
- 24 spécimens asiatiques
- 29 spécimens blancs
- 29 spécimens noirs

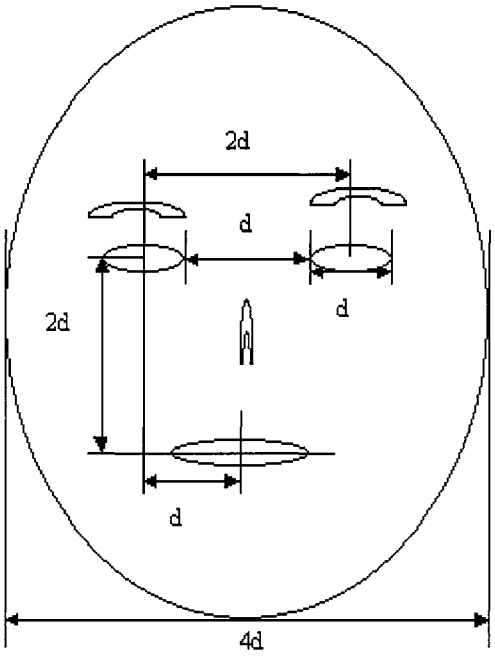


Figure A.14. Certaines proportions idéales selon [84].

Voici les principales différences observées entre les races :

- La largeur du nez des Noirs est plus grande que celle des Blancs et des Asiatiques
- Les Blancs ont un nez plus long que celui des Noirs
- Les Noirs ont une bouche plus grande et plus large que celle des Blancs

Dans [84], certaines proportions idéales du visage sont utilisées telles que présentées à la figure A.14.

Dans [1], des visages potentiels sont détectés sur une image en ton de gris. D'abord, l'image des gradients est obtenue. Ensuite, la transformée de Hough est appliquée pour localiser des ellipses (pour les contours de tête potentielle). Pour renforcer la localisation des ellipses, la distribution des éléments du visage en position frontale est exploitée à l'aide d'un masque. Ce masque contient les propriétés anthropométriques du visage à l'aide de mesures effectuées sur plusieurs individus. Le masque est donc positionné vis-à-vis des ellipses détectées sur l'image des gradients. Les éléments du masque (yeux, bouche, etc) sont vérifiés selon les propriétés des gradients sur le visage (le nez a des gradients horizontaux, la bouche a des gradients verticaux, etc). Plusieurs essais sont effectués (changement de taille de l'ellipse, de position et d'orientation). Mais l'algorithme ne vise qu'à localiser la tête, les positions des éléments du visage (yeux, bouche, etc) ne sont pas recherchées avec précision. La figure A.15 représente les proportions statistiques selon [1].

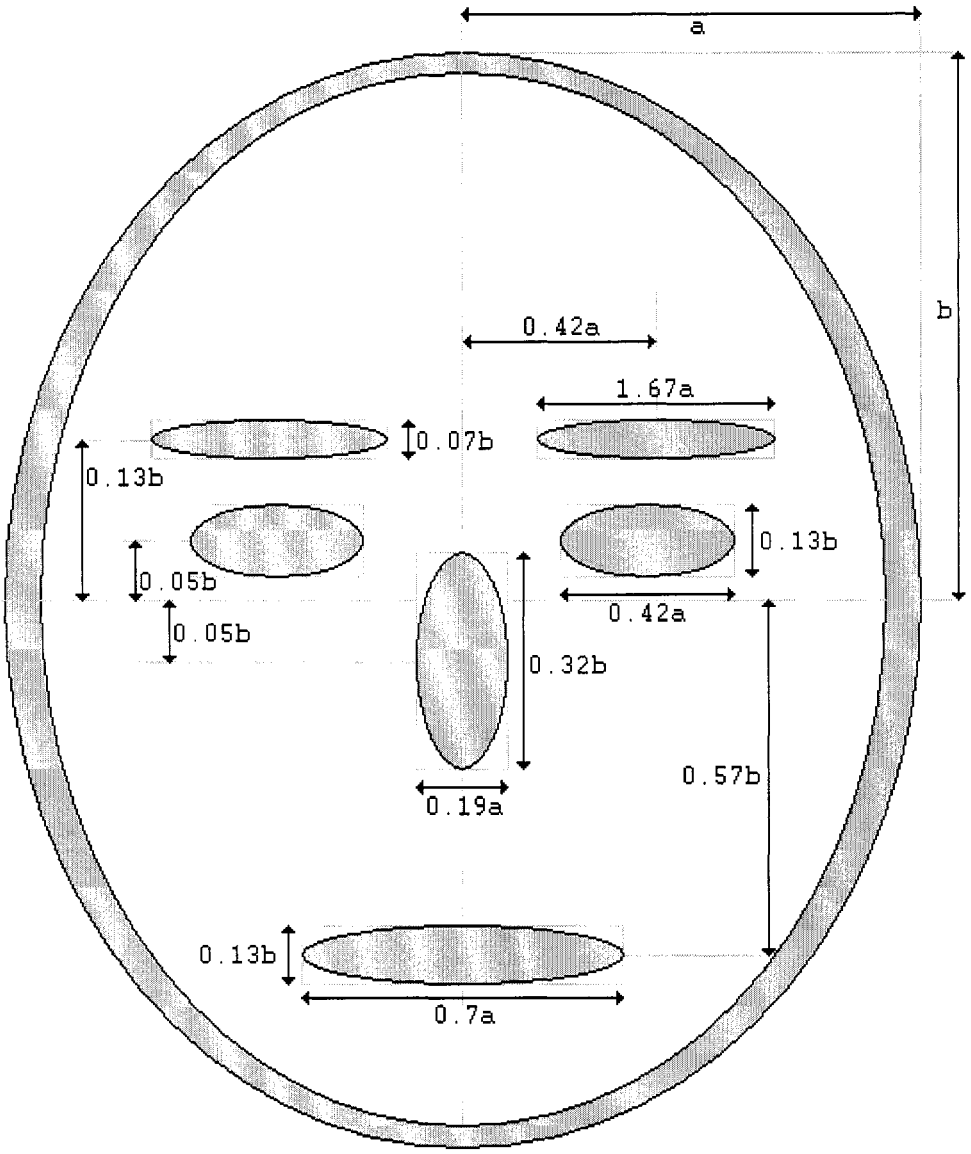


Figure A.15. Proportions du visage utilisées selon [1].

Dans les articles [4] et [9], le suivi est effectué sur 5 points sur le visage : les 2 coins des 2 yeux et le bout du nez. La séquence d'images est monoculaire et la position relative du visage par rapport à la caméra est effectuée grâce à certaines caractéristiques du visage telles la symétrie des yeux et les statistiques des mesures anthropométriques.

La forme exacte de la tête et du visage n'est pas connue au départ. Pour estimer l'orientation de la tête, des distances calculées à partir des 5 points sont utilisées. Lorsque la tête a une position frontale, les 4 points situés sur les coins des yeux sont presque colinéaires en 3D et cette droite engendrée est presque parallèle au plan de l'image. Ces 4 points permettent donc d'évaluer les rotations autour des axes Y et Z. Le 5^{e} point sur le bout du nez est utilisé pour calculer la rotation autour de l'axe X mais l'information de la profondeur n'est pas disponible. Pour régler ce problème, l'usager est classifié selon l'âge, la race et le genre. Ensuite, la rotation autour de l'axe X est estimée en tenant compte des statistiques des mesures anthropométriques de l'usager. La position des 5 points est censée être acquise sur la première image. La figure A.16 illustre la distribution de ces points.

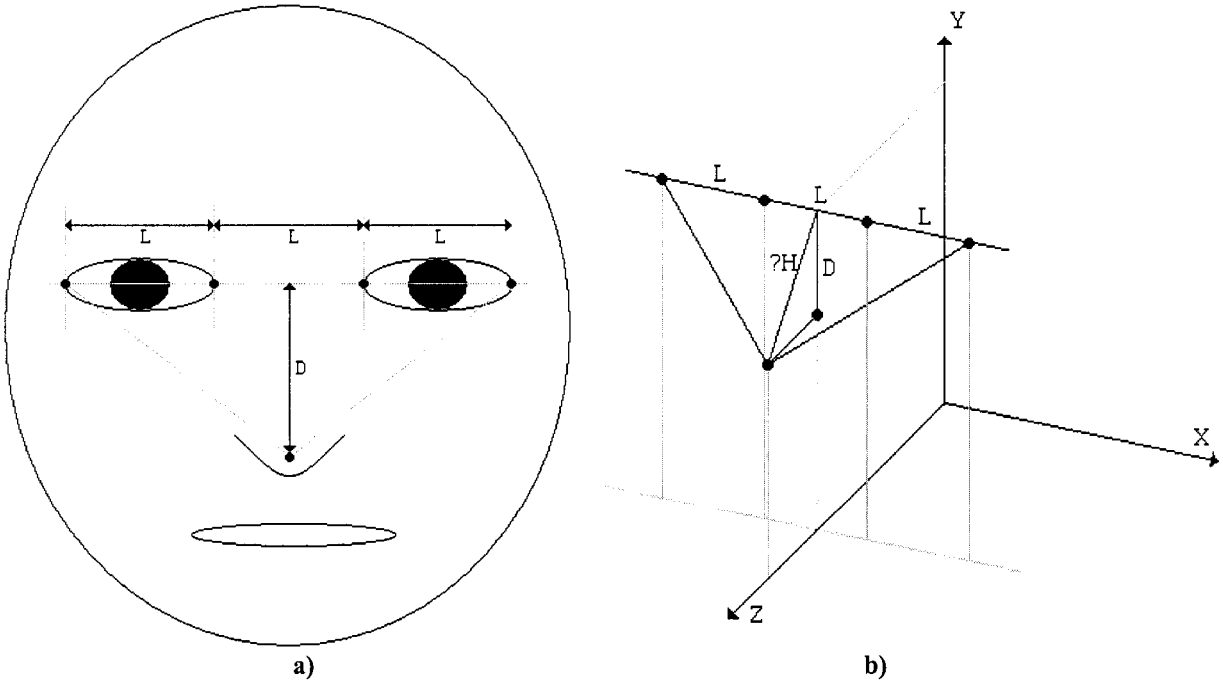


Figure A.16. Géométrie de quelques éléments du visage. En a), en 2D. En b), la valeur H doit être estimée par des statistiques anthropométriques afin d'avoir une information 3D.