
Mémoire de fin d'études

Détection de visages 2D dans des poses frontales et non-frontales

Etudiant : PHAM Van Viet

pvviet@ifi.edu.vn

Encadrant : Dijana PETROVSKA

Dijana.Petrovska@it-sudparis.eu

*Stage réalisé au Département Electronique et Physique (EPH) du 01/04/2010 au
30/09/2010*

17 septembre 2010

TABLE DES MATIERES

Liste des figures	ii
Liste des tables	iii
Remerciements	iv
Résumé	v
Abstract	v
1. INTRODUCTION.....	1
1.1. DETECTION DE VISAGE ET SON APPLICATION.....	1
1.2. PROBLEMATIQUES	2
1.3. OBJECTIFS DE STAGE	3
1.4. STRUCTURE DU MEMOIRE	5
2. ETAT DE L'ART	7
2.1. DETECTION DU VISAGE FRONTAL	7
2.1.1. INTRODUCTION (La littérature).....	7
2.1.2. DETECTEUR DU VISAGE EN UTILISANT L'ADABOOST (DETECTEUR DE VIOLA&JONES)	10
2.2. DETECTION DU VISAGE DE PROFILE.....	22
2.3. BASES DE DONNEES PUBLIQUES.....	24
2.3.1. BANCA	24
2.3.2. MBGC	25
2.3.3. FRGC	27
2.3.4. FPI.....	28
3. EVALUATION DU SYSTEME DE DETECTION DE VISAGE DE FACE	29
3.1. EVALUATION DU SYSTEME EN UTILISANT LES PARAMETRES PAR DEFAULT	29
3.2. EVALUATION DU SYSTEME AVEC D'AUTRES PARAMETRES	30
3.2.1. DIMINUTION DE FAUSSES ALARMES PAR LA QUALITE DE REGION DETECTEE (CV_HAAR_DO_CANNY_PRUNING)	30
3.2.2. TERMINAISON DE DETECTION DES OBJETS (CV_HAAR_ROUGH_SEARCH)	34
3.2.3. UTILISATION DE CONNAISSANCES A PRIORI (CV_HAAR_FIND_BIGGEST_OBJECT)	36
3.3. CONCLUSION.....	39
4. EVALUATION DU SYSTEME DE DETECTION DU VISAGE DE PROFILE.....	41
4.1. EN UTILISANT LES CARACTERISTIQUES DE HAAR.....	41
4.2. CONCLUSION.....	43
5. CONCLUSION ET PERSPECTIVES.....	44
6. BIBLIOGRAPHIE	46

Liste des figures

Figure 1 - Détection du visage de face	5
Figure 2 - Filtres rectangulaires dits de Haar mis en œuvre par la cascade avec les poids w associé avec aux rectangles blancs entiers et aux rectangles gris qu'ils incluent.....	11
Figure 3 - Un filtre rectangulaire est défini par sa forme, sa taille et sa position dans la rétine de recherche [23].....	12
Figure 4 - $II(x_0, y_0)$ est, par définition, la somme des niveaux de gris situés sous le rectangle A. La somme des pixels situés sous le rectangle Rect est la somme de ceux situés sous les quatres rectangles moins la somme des pixels sous les rectangles A et B, moins la somme des pixels sous les rectangles A et C, plus la somme des pixels sous le rectangle A	15
Figure 5 - Cascade de classifieurs boosté : les classifieurs les plus simple et les plus rapides rejettent un maximum de fausses alarmes, les classifieurs les plus complexes se concentrent alors sur les zones de l'image les plus proche de l'apparence d'un visage .	17
Figure 6 - Nombre de classifieurs faibles contenus dans chacun des 25 classifieurs boostés de la cascade 24x24 [23]	18
Figure 7 - Les 9 filtres de Haar du premier classifieur fort F_0 de la cascade [23]	19
Figure 8 - Les filtres de Haar du second étage F_1 de la cascade [23].....	21
Figure 9 – Deux cascades en parallèle.....	24
Figure 10 - Quelques exemples de la base BANCA.....	25
Figure 11 - Quelques exemples de la base MBGC d'un homme qui marche à travers un portail ...	26
Figure 12 - Exemple d'une session du sujet dans la base FRGC	27
Figure 13 - Quelques exemples de la base FERET et PIE	28
Figure 14 - Bons résultats de détection	32
Figure 15 - Résultats avec des fausses alarmes.....	33
Figure 16 - Exemples de non-détection	33
Figure 17 - Résultats de détection avec des fausses alarmes	35
Figure 18- Bons résultats de détection	37
Figure 19 - Position de détection correcte.....	37
Figure 20- Résultats de détection avec la position fausse	38
Figure 21 - Comparaison de la performance de trois paramètres	40
Figure 22 - Résultats de détection correcte avec l'image de profile.....	42
Figure 23 - Résultats de détection avec des fausses alarmes	42
Figure 24- Exemples de non-détection	43
Figure 25 - Courbe ROC de détection du visage de profile en utilisant les caractéristiques de Haar	43

Liste des tables

Table 1 - L'algorithme d'apprentissage par AdaBoost	20
Table 2 – Approche rétinienne du détecteur de visage implémenté par OpenCV	22
Table 3 - Résultats de 2 expérimentations avec des paramètres par défaut	30
Table 4 - Résultat de 2 expérimentations avec le paramètre CV_HAAR_DO_CANNY_PRUNING ...	31
Table 5 - Résultat de chaque catégorie de la base BANCA avec le paramètre CV_HAAR_DO_CANNY_PRUNING.....	34
Table 6 - Résultat de 2 expérimentations avec le paramètre CV_HAAR_DO_ROUGH_SEARCH	35
Table 7 - Résultat de 2 expérimentations avec le paramètre CV_HAAR_BIGGEST_OBJECT.....	36
Table 8 - Résultat de chaque catégorie de la base BANCA avec le paramètre CV_HAAR_BIGGEST_OBJECT.....	39
Table 9 - Résultat de la détection de visage de profile	41

Remerciements

En préambule de cette mémoire, je souhaite adresser ici tous mes remerciements aux personnes qui m'ont apporté leur aide et qui ont ainsi contribué à l'élaboration de cette mémoire.

Tout d'abord, Mme Dijana Petrovska-Delacrétaz, directeur de ce stage, et M. NGUYEN Quoc Dinh, pour leur direction, leurs orientations et leur compréhension.

Tous mes professeurs d'informatique à l'Institut de la Francophonie pour l'Informatique pour leurs enseignements.

Résumé

Ce travail porte sur la détection de visage de face et de visage de profile. L'application de détection de visage est diverse, par exemple : la biométrie, la détection de sourires dans les appareils photos récents, les outils de suivi de visages par des webcams motorisées, etc. La détection du visage a reçu beaucoup d'attention des chercheurs. Bien que la recherche en détection faciale se poursuive depuis de nombreuses années, ce problème est encore largement non résolu, notamment sur la détection du visage profile. L'objectif de ce stage est d'étudier et d'estimer l'algorithme Adaboost et des caractéristiques de Haar pour la détection du visage. Dans un premier temps, nous utilisons l'Adaboost et des caractéristiques de Haar pour la détection du visage de face et nous réalisons aussi des étapes de post-traitement pour l'amélioration de la performance de détection. En fin, nous utilisons Adaboost et des caractéristiques de Haar pour la détection du visage de profile.

Mots-clés : Détection du visage de face, Détection du visage de profile, AdaBoost, Caractéristiques de Haar

Abstract

This work focuses on the frontal and profile face detection. The application of face detection is diverse, for example : biometrics, detection of smiles in modern cameras, tools for tracking faces by motorized webcams, etc. The face detection has received much attention from researchers. Although research in face detection will continue for many years, this problem is still largely unresolved, in particular the profile face detection. The objective of this work is to study et to estimate the algorithm AdaBoost and Haar features for the face detection. Initially, we use the AdaBoost and Haar features for the frontal face detection and we also carry out post-processing steps to improve the detection performance. In the end, we use the AdaBoost and Haar features for the profile face detection

Key words : Frontal face detection, Profile face detection, AdaBoost, Haar features

1. INTRODUCTION

1.1. DETECTION DE VISAGE ET SON APPLICATION

Au cours des 40 dernières années, depuis que Sakai et *al.* [7] ont commencé à développer un système pour détecter la présence ou l'absence d'un visage dans une image, de nombreux chercheurs, dans des disciplines différentes telles que le traitement d'image, la reconnaissance de formes, les réseaux de neurones, la vision par ordinateur et la psychologie, ont effectué des recherches sur le traitement du visage. Les progrès ont été ralentis par la complexité de la relation entre le visage comme objet et son image, complexité impliquant la forme, les propriétés de surface, la pose et l'éclairage, etc.

La détection du visage est une tâche préliminaire nécessaire à la plupart des techniques d'analyse du visage, elle est donc un champ d'investigations primordial dans le domaine de la vision par ordinateur et du traitement d'images. Ses applications sont diverses, citons par exemple, la biométrie, l'Interface Homme machine (IHM), la reconnaissance de gestes, la détection de sourires dans les appareils photos récents, les outils de suivi de visages par des webcams motorisées, la recherche de visages dans des grandes bases de données, etc.

Plus précisément ses applications peuvent être citées comme la suivant :

- ✓ **En vidéo-surveillance** : Son intérêt est de déterminer la présence d'êtres humains pour surveiller automatiquement des zones sensibles et puis remplacer des systèmes de sécurité actuelles comme ceux de clefs, de badges ou de mots de passe [4]. La détection du visage est l'étape préliminaire dans le cadre d'un système de contrôle d'accès à un bâtiment, à une porte d'embarquement ou à un ordinateur.
- ✓ **En biométrie faciale** : Parmi des développements récents dans ce domaine, on trouve des applications de la détection du visage dans la reconnaissance du visage en trois dimensions ou dans l'utilisation d'infrarouges pour analyser la cartographie vasculaire du visage.

- ✓ **En jeux vidéo** : Dans les domaines déjà accessibles au grand public, outre les webcams standard, certains jeux vidéo vont bientôt intégrer une nouvelle technologie permettant d'incruster le visage du joueur dans les jeux. C'est-à-dire tout d'abord l'utilisateur prend une photo de face et une de profil. Ensuite, un modèle 3D du visage est reconstruit à partir ces photos, puis l'sauvegardé dans le jeu.
- ✓ **En indexation d'images** : La croissance exponentielle de base de données numériques a rendu nécessaire l'utilisation de la détection du visage, pour la détermination des images contenant des personnes (ou pas, par exemple si l'on cherche des images de paysages) ou pour la recherche des images contenant une personne particulière.
- ✓ **En interaction homme-machine** : Son application se trouve dans des commandes par le visage pour l'assistance aux personnes handicapées, ou dans la perception par un robot de son environnement.
- ✓ **En études comportementales** : En couplant à une reconnaissance de geste, de posture, ou à un suivi de regard, la détection du visage permettra des études automatiques sur l'apprentissage de parole chez un bébé ou encore sur les zones d'attention de regard sur des affiches publicitaires.
- ✓ Les dernières générations d'appareil photos ont une option de réglage de la mise au point sur les visages quand on prend des photos.

1.2. PROBLEMATIQUES

Pour la tâche de détection du visage, on peut résumer les problèmes typiques rencontrés dans la littérature :

- ✓ **La pose** : Dans une image, l'orientation d'un visage n'est pas toujours frontale (dans le plan). Le plus souvent, le visage subit de grandes rotations en profondeur qui peuvent se produire dans l'un ou les deux axes de l'image. Ces rotations introduisent des transformations non-linéaires (ou déformations). Le visage devient alors un objet non-rigide : son contour et les contours des éléments qui le composent varient, modifiant ainsi ses

formes et pouvant provoquer une occultation partielle dans l'image. Le changement du sens de réflexion de la lumière sur un visage lui-même lié à une modification de la pose est responsable du changement de son apparence.

- ✓ **L'expression du visage** : qui est causé par la parole et les émotions est difficile à traiter par les méthodes de suivi décrites dans la littérature. D'abord, l'expression modifie la forme du visage indépendamment de la pose. C'est ce que l'on a appelé la déformation intrinsèque. En outre, l'expression du visage est causée par le mouvement des muscles du visage qui modifie la surface du visage dynamiquement et fabrique différents bruits comme les rides d'expressions, les ombres, etc.
- ✓ **Les conditions d'acquisition de l'image** : quand une image est acquise par caméra par exemple, il y a des facteurs affectant sur l'apparition du visage dans l'image. Plus précisément l'éclairement de la scène : les sources lumineuses, le cas d'ombres portées sur le visage, le cas de source de surexposition lumineuse; les différences de l'appareil capturé de l'image : sensibilité, résolution de l'image.
- ✓ **D'autres défis** : l'occlusion où des parties du visage peut être occlues par des autres objets (par exemple : dans une image ayant plus d'une personne, des visages peuvent être occlues partiellement par des autres visages), couleur de la peau (noir, blanche, etc.).

1.3. OBJECTIFS DE STAGE

La détection du visage dans images consiste à déterminer la présence ou l'absence de visages dans une ou plusieurs images et en cas de présence à déterminer sa localisation. Les techniques utilisées sont généralement issues du domaine de la reconnaissance des formes. En effet, le problème peut être vu comme la détection de caractéristiques communes à l'ensemble des visages humaines : il s'agit de comparer une image à un modèle générique de visage et d'indiquer s'il y a ou non ressemblance. Ces méthodes seront donc fortement conditionnées par le choix effectué pour modéliser un visage.

La sortie d'un détecteur de visage indique le nombre de visage présent dans l'image. De plus, la plupart des détecteurs de visage actuel sont aussi des localisateurs de visages : ils renvoient une localisation des visages détectés (une boîte englobante par exemple). La figure 1 montre un exemple de la détection du visage de face dans une image.

En raison de son importance pratique, la détection du visage a reçu beaucoup d'attention de la part des chercheurs. Bien que la recherche en détection faciale se poursuive depuis de nombreuses années, ce problème est encore largement non résolu, notamment sur la détection du visage profile. Ces dernières années ont vu des progrès significatifs dans le domaine (plusieurs travaux sur la détection ont été réalisés, mais la détection faciale profile reste un grand défi pour les chercheurs en vision par ordinateur et en reconnaissance des formes.

Un état de l'art sera expliqué dans le chapitre 2, il est à noter qu'une technique populaire a été développée dans [3]. Il s'agit de la méthode, d'après notre connaissance, qui marche en temps réel et donne une meilleure performance par rapport à l'autre méthode. La méthode est implantée dans le toolbox OpenCv (les détails de cette méthode dans OpenCv se trouvent sur la ligne <http://note.sonots.com/SciSoftware/haartraining.html>) mais l'information de base de données, les étapes de prétraitement et les étapes de post-traitement ne sont pas publiés.

Dans la carte de ce stage, nous voulons tester plusieurs paramètres du détecteur dans le toolbox OpenCv, l'amélioration de la performance avec l'étape de post-traitement et de l'entraînement. Ensuite, nous nous proposons d'aborder le problème de la pose du visage. En particulier, nous nous attachons au problème d'utilisation des caractéristique de Haar travers le détecteur de Viola & Jones [3].



Figure 1 - Détection du visage de face

1.4. STRUCTURE DU MEMOIRE

Ce stage vise essentiellement à construire des systèmes de détection du visage (de face et de profile). Il est important de souligner que nous nous basons notre travail sur le travail de Viola [3]. Ce mémoire est organisé comme suit :

- ✓ **Chapitre 2 :** Un état de l'art de la détection du visage est présenté en premier. Puis, une présentation détaillée de l'algorithme du détecteur de visages de Viola&Jones. En fin, une présentation des bases de données utilisées dans l'évaluation par exemple : FRGC, MBGC, BANCA et PFI (pie, feret et autres images à partir de sources Internet).
- ✓ **Chapitre 3 :** Nous allons montrer le résultat d'évaluation d'AdaBoost sur la base de données de test BANCA en utilisant les paramètres par défaut. Puis

nous parlons du résultat d'évaluation d'AdaBoost avec des paramètres qui permettent de diminuer les fautes d'alarme.

- ✓ **Chapitre 4** : Nous allons présenter des expérimentations sur la détection de visage de profile.
- ✓ **Chapitre 5** : Conclusion et perspectives

2. ETAT DE L'ART

Le problème du traitement du visage s'est avéré plus difficile que prévu et a attiré l'attention de nombreux chercheurs. Ce chapitre ne souhaite pas couvrir la totalité des méthodes de traitement du visage (les lecteurs sont invités à se référer à des articles plus détaillés : [1] et [2]). Nous voulons seulement présenter ici les principales méthodes de détection du visage.

2.1. DETECTION DU VISAGE FRONTAL

2.1.1. INTRODUCTION (La littérature)

La détection de visage a pour objectifs de décider si une image contient un ou plusieurs visages, et de donner leur localisation. Deux principales métriques caractérisent les performances d'un système de détection : le taux de détection, c'est-à-dire, le pourcentage de visages correctement détectés, et le taux de faux positifs, correspondant à des détections dans des régions ne contenant pas de visage. Un système performant est associé à la fois à un taux de détection important et à un taux de faux positifs faible. D'autre part, la localisation faciale, consiste à trouver la position exacte du visage dans une image.

Comme la détection du visage a de très nombreuses applications, diverses techniques ont déjà été proposées pour extraire des informations visuelles associées au visage. Selon les types d'informations extraits par ces techniques ainsi que le processus de développement de ces techniques (dans le domaine de la vision par ordinateur en général ou de la détection du visage particulier), nous les classons en 2 grandes catégories : les approches locales et celles globales.

2.1.1.1. LES APPROCHES LOCALES

Les approches locales, les plus anciennes, reposent sur une analyse de bas niveau de l'image par l'étude des coins, de l'intensité, de la couleur [1], [8]. Une des approches qui est beaucoup utilisée est la couleur, une caractéristique de base qui peut être efficace et discriminante en segmentation ou en classification du visage.

Elle est aussi facile à comprendre et robuste vis-à-vis de certains changements géométriques (e.g. la rotation, l'échelle, ou la translation). L'espace couleur le plus fréquemment employé est le RVB (Rouge - Vert - Bleu) classique. Mais, le YCbCr (i.e. Y est la luminance, Cb et Cr sont des composantes chromatiques), le HSV (i.e. Teinte, Saturation, Luminance) et les espaces perceptifs (Luv, Lab) sont aussi utilisés régulièrement pour faire ressortir l'information de teinte [23].

La couleur a été couramment utilisée pour la détection du visage dans la littérature. Elle peut être une caractéristique satisfaisante pour une détection du visage dans de bonnes conditions. En général, en raison de sa sensibilité aux variations d'éclairage, les changements d'illumination (chromatiques en particulier) peuvent être préjudiciables à certaines méthodes [24]. De plus, l'absence de contrainte sur la forme fait que ce type de méthode ne donne en général que des formes approximatives des visages et est particulièrement sensible au bruit.

2.1.1.2. LES APPROCHES GLOBALES

Les méthodes globales, c'est-à-dire basées sur la détection du visage dans sa globalité, ont été introduites dans le but d'être appliquées à des images contenant plusieurs visages et/ou en présence d'un fond complexe. Ces techniques évitent les problèmes d'imprécision dans la détection des caractéristiques faciales par le biais de l'apprentissage des règles intrinsèques des visages. Pour cela, il est nécessaire de disposer d'un volume de données important et présentant une grande variabilité dans les conditions d'éclairage, de prise de vue, etc. Ces dernières approches sont en général plus robustes au bruit et à des déformations de la région faciale que les approches locales. La majorité de ces approches traitent la détection comme un problème de classification binaire, où les deux classes sont les visages et les non visages. Les approches globales les plus connues reposent sur l'utilisation de techniques d'analyse statistique multivariées ([9], [10], [11], [12], [13], [14]) pour la représentation des visages et/ou leur classification. Cette phase d'analyse est éventuellement suivie de classifieurs Bayésiens ou basés sur des réseaux de neurones. D'autres approches reposent sur l'utilisation directe de réseaux de

neurones ou de Machines à Vecteurs de Support pour la classification et la détection de visages [15], [16], [17], [3], [19].

Depuis plusieurs années, plusieurs chercheurs ont consacré des efforts considérables à l'amélioration des techniques qui pourraient conduire à un système d'apprentissage avec une forte capacité de généralisation. Une des techniques ayant rencontré le plus de succès est le boosting proposé par Freund en 1995 [5]. Contrairement aux approches d'apprentissage ordinaires qui tentent de générer un classifieur unique à partir de données, la méthode de boosting essaie de construire un ensemble de classifieurs faibles puis cherche à les combiner. Les classifieurs faibles sont généralement produits à partir des données par règle de classification qui peut être un réseau de neurones [22], des SVMs ou d'autres types de règle [6], [3]. La méthode de boosting est capable de "booster" les classifieurs faibles, un tant soit peu meilleurs que le hasard, pour fabriquer un classifieur fort et efficace. Freund and Schapire [5] ont ensuite proposé un algorithme de boosting adaptatif, nommé l'Adaboost, qui permet de pondérer les classifieurs.

Viola et Jones [3] ont proposé une cascade de détecteurs où chaque détecteur est basé sur Adaboost. Cette méthode a constitué un pas important dans le domaine de la détection de visages. C'est le premier détecteur à présenter des capacités en temps réel tout en ayant des performances comparables à celles de l'état de l'art. Elle se base sur une sélection automatique des ondelettes de Haar appliquée à l'image monochrome, et à une évaluation itérative de l'importance accordée à chaque exemple durant l'apprentissage, grâce à la procédure Adaboost. Pour la détection du visage, Lienhart et al [6] ont amélioré la dernière méthode en proposant un arbre de détecteurs. Les auteurs ont montré un résultat sur le suivi de bouche dont la précision de détection est égale ou supérieure à un classificateur simple ou multiple en cascade, tout en exigeant moins de calculs. Cette méthode sera détaillée dans la prochaine section.

2.1.2. DETECTEUR DU VISAGE EN UTILISANT L'ADABOOST (DETECTEUR DE VIOLA&JONES)

Viola&Jones [3] a développé un algorithme appelé *Adaboost* comme un détecteur de visages qui se base sur des filtres rectangulaires de Haar illustrés par la figure 2. Cet algorithme est amélioré plus tard par Lienhart et *al.* [20]. Afin de bien comprendre cet algorithme, on va examiner un exemple suivant. Pour une rétine de taille 24x24 pixels, on peut utiliser plus de 180.000 filtres en multipliant indépendamment la hauteur et la largeur des rectangles par un nombre entier, et en comptant les différentes positions de l'image où ces filtres peuvent être centrés dans la rétine. Le nombre de filtres est déterminé dans l'intervalle m^4 où $m = 24$ en se basant sur le nombre de positions possibles multiplié par le nombre de tailles. Après avoir défini la taille de rétine et le nombre de filtres, on applique un classifieur statistique sur une base d'images monochromes de visages et de non-visages de taille fixe 24x24 (ou 20x20) par une procédure de *boosting* adaptative. La sélection des filtres rectangulaires mis en œuvre par chacun des classifieurs boostés se fait automatiquement durant l'apprentissage. Le schéma de la figure 5 montre que ce classifieur est composé d'une cascade attentionnelle composée de classifieurs boostés. La détection de visages se fait par une approche basée sur un codage rétinien : une fenêtre glissante de taille 24x24 scanne l'image et chaque sous-image est classifiée visage ou non-visage par la cascade. La fenêtre est ensuite agrandie d'un facteur f_q , et scanne à nouveau l'image à plus grande échelle, l'image gardant sa résolution initiale : moins de vignettes sont alors visées par le classifieur.



Figure 2 - Filtres rectangulaires dits de Haar mis en œuvre par la cascade avec les poids ω associé avec aux rectangles blancs entiers et aux rectangles gris qu'ils incluent

2.1.2.1. REPRESENTATION DE L'IMAGE : FILTRES SPACIAUX ET CLASSIFIEUR FAIBLES ASSOCIES

La figure 3 donne un exemple de vignette sur laquelle est appliqué un filtre de *Haar*. Chaque filtre est défini par :

- ✓ Un ou deux rectangles gris inclus dans un rectangle blanc comme le montre la figure 2.
- ✓ Ses dimensions, largeur l_0 et hauteur h_0 à la première échelle, où l_0 et h_0 sont des multiples entiers des dimensions des filtres rectangulaires de base illustrés par la figure 2. La taille $l \times h$ du filtre à l'échelle n est alors : $l = \overline{f_q^{n-1} l_0}$ et $h = \overline{f_q^{n-1} h_0}$ (la barre indiquant l'arrondi entier le plus proche).
- ✓ Sa position dans la rétine : si elle est définie en (x_0, y_0) à l'échelle un, elle devient $(\overline{f_q^{n-1} x_0}, \overline{f_q^{n-1} y_0})$ à l'échelle n .

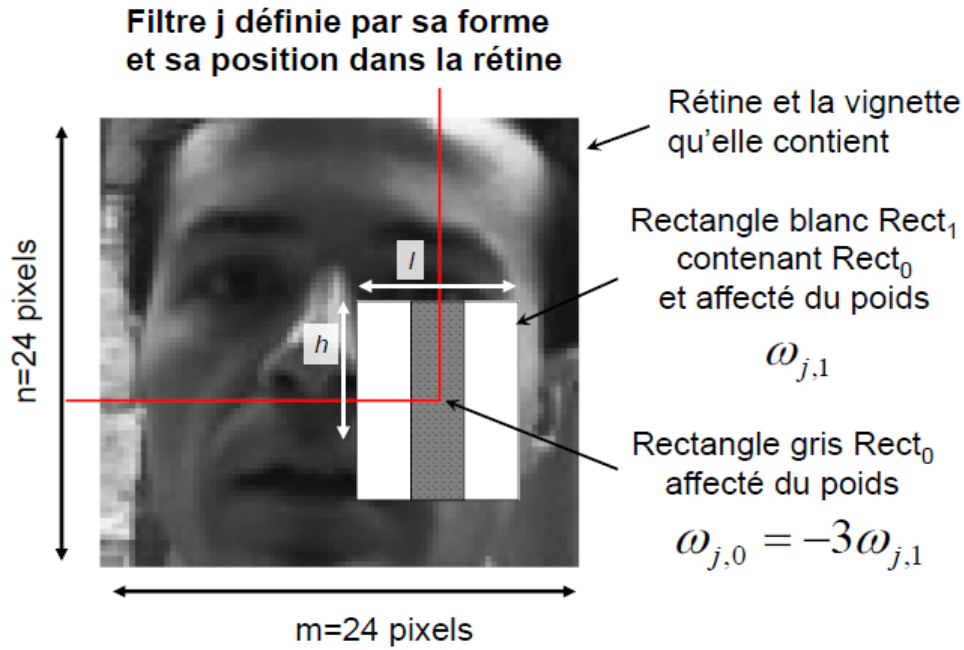


Figure 3 - Un filtre rectangulaire est défini par sa forme, sa taille et sa position dans la rétine de recherche [23]

Les filtres sont constitués de deux à trois rectangles : les filtres contours et les filtres lignes en ont deux, les filtres diagonales trois. A chaque filtre est associé un classifieur faible ayant pour entrée une imagerie et pour sortie la classe, visage ou non-visage, de l'entrée; sa fonction discriminante est définie par l'équation 1.3. Un classifieur faible seul est insuffisant pour détecter un visage avec un taux de réussite acceptable. Le boosting permet de combiner des classifieurs faibles pour obtenir un classifieur fort dit *boosté*. Alors, un classifieur boosté est une somme pondérée des classifieurs faibles dont chacun est associé avec un poids particulier à chaque itération par une procédure d'*AdaBoost*. Plus tard, plusieurs classifieurs boostés sont mis en cascade pour obtenir un classifieur efficace avec la vitesse et la performance. Pour chaque filtre, un opérateur différentiel est appliqué entre la portion d'image située sous le rectangle entier et l'image sous le rectangle grisé est calculée. La valeur s_j issue du filtre j est ainsi la somme pondérée de deux termes : la somme des pixels situés dans le rectangle entier (blanc + noir), indexé par 1 dans la formule (1), et la somme des pixels inscrit dans le rectangle gris (indexé par 0).

$$s_j = \omega_{j,1} \sum_{(x,y) \in Rect_1} I(x,y) + \omega_{j,0} \sum_{(x,y) \in Rect_0} I(x,y) \quad (1)$$

Le rectangle gris se voit affecté un poids inversement proportionnel à l'aire qu'il occupe dans le rectangle entier. La figure 3 illustre le cas d'un filtre constitué d'un rectangle gris inscrit dans un rectangle blanc d'aire trois fois plus élevée.

La formule (1) montre aussi la différence entre la somme des niveaux de gris situés sous le rectangle gris et celui des deux rectangles blancs qui lui sont adjacents en utilisant un même poids moyen à chaque rectangle. Une amélioration proposée par Lienhart [6] dans OpenCv est la normalisation de ces poids lorsque la vignette parcourt l'image à une échelle qui est supérieur à 1. En effet, les filtres rectangulaires sont agrandis d'un facteur f_Q^{n-1} , mais la somme des pixels est calculée en faisant référence à des points de coordonnées entières d'une image intégrale. L'arrondi vers l'entier le plus proche est exécuté pour déterminer chacun des quatre points de l'image intégrale (et de l'image auxiliaire permettant le calcul de la variance, pour la correction de luminosité décrite plus bas). [6] a montré que l'opération d'arrondi a un impact considérable sur les performances du détecteur final. En fait, le rapport entre l'aire du rectangle blanc et celui du rectangle gris qu'il contient peut différer de sa valeur à l'échelle $n = 1$ quand les classifieurs ont été appris. Les poids de chaque rectangle, pour un filtre donné, sont corrigés à l'échelle n de manière à conserver le même rapport d'aire qu'à l'échelle 1. De l'ordre de dix fois plus de faux positifs, pour un même taux de détection, sont obtenus sans cette correction.

Un classifieur f_j est associé au filtre j . Appelé classifieur faible, il est défini par une sur la valeur s_j . Il s'agit d'un arbre de décision à deux nœuds terminaux :

$$f_j = \begin{cases} 1 & \text{si } s_j \geq \tau_j & \text{visage} \\ 0 & \text{si } s_j < \tau_j & \text{non-visage} \end{cases} \quad (2)$$

Un nombre important de classifieurs faibles sont à combiner, et la différence calculée par l'équation (2) pour chaque filtre et à chaque échelle est consommatrice en temps de calcul. Viola & Jones utilisent une représentation intermédiaire de

l'image, l'image intégrale II qui permet de calculer ces valeurs de manière rapide sur l'image monochrome I , en une seule passe.

$$II(x, y) = \sum_{\substack{x' \leq x \\ y' \leq y}} I(x', y') \quad (3)$$

L'image intégrale peut être calculée en une passe sur l'image I par récurrence, en parcourant chaque ligne de l'image selon les x croissants et en changeant de ligne selon les y croissants. En effet :

$$II(x, y) = \sum_{x' \leq x} I(x', y) - \sum_{x' \leq x-1} I(x', y) \quad (4)$$

Ou

$$\sum_{x' \leq x} I(x', y) = \sum_{\substack{x' \leq x \\ y' \leq y}} I(x', y') - \sum_{\substack{x' \leq x \\ y' \leq y-1}} I(x', y') \quad (5)$$

Et

$$\sum_{x' \leq x-1} I(x', y) = \sum_{\substack{x' \leq x-1 \\ y' \leq y}} I(x', y') - \sum_{\substack{x' \leq x-1 \\ y' \leq y-1}} I(x', y') \quad (6)$$

D'où l'on déduit la formule à appliquer pour calculer l'image intégrale :

$$II(x, y) = I(x, y) + II(x-1, y) + II(x, y-1) - II(x-1, y-1) \quad (7)$$

En posant $II(-1, y) = II(x, -1) = 0$ ce qui fait en particulier que l'image intégrale est de dimensions $(W+1) \times (H+1)$ pour une image de taille $W \times H$. La somme des niveaux de gris des pixels inclus dans un rectangle $Rect(x_0, y_0, w, h)$ de dimensions $w \times h$ et dont le coin haut gauche est à la position (x, y) se déduit alors de l'image intégrale par quatre opérations de lecture de l'image intégrale :

$$\sum_{(x,y) \in Rect} I(x, y) = II(x_0 + w - 1, y_0 + h - 1) - II(x_0 + w - 1, y_0) + II(x_0, y_0 + h - 1) - II(x_0, y_0) \quad (8)$$

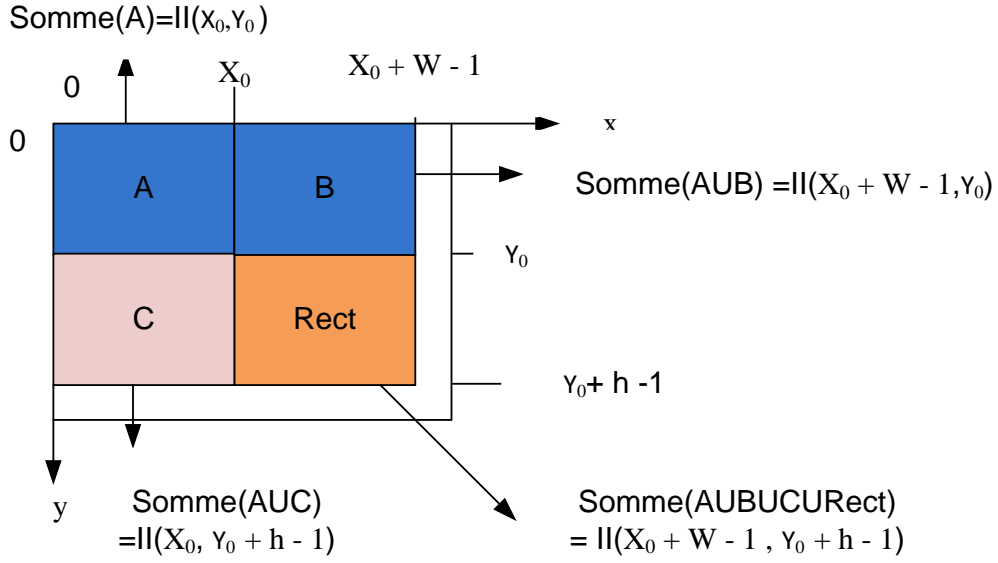


Figure 4 - $I(x_0, y_0)$ est, par définition, la somme des niveaux de gris situés sous le rectangle A. La somme des pixels situés sous le rectangle Rect est la somme de ceux situés sous les quatre rectangles moins la somme des pixels sous les rectangles A et B, moins la somme des pixels sous les rectangles A et C, plus la somme des pixels sous le rectangle A

L'image intégrale est calculée une seule fois par image et un classifieur faible j donné y fait référence pour le calcul de la différence s_j en un temps constant quelque soit l'échelle. Toutes les vignettes visées par la cascade en test, de même que tous les exemples d'apprentissage, sont normalisés par la variance, afin de diminuer la sensibilité du détecteur aux conditions d'éclaircissement. Soit ainsi **Rect** la vignette courante (i.e. sa position et ses dimensions $w \times h$), l'imagette appliquée en entrée de la cascade est celle dont les niveaux de gris sont corrigés par la formule (9) :

$$I'(x, y) = \frac{I(x, y) - \mu}{2\sigma} \quad (9)$$

Où μ et σ sont respectivement la moyenne et l'écart-type des niveaux gris sur la vignette. Le niveau moyen de gris est la somme des pixels dans la vignette divisé par son aire. La valeur σ en revanche nécessite de calculer la somme du carré des niveaux de gris :

$$\sigma^2 = \frac{\sum_{(x,y) \in Rect} I(x, y)^2}{w \times h} - \mu^2 \quad (10)$$

Cette somme est calculée par la même astuce de l'image intégrale. La somme intégrale du carré des niveaux de gris correspondante est obtenue en remplaçant le terme $I(x, y)$ par $I(x, y)^2$ dans l'équation (7).

2.1.2.2. APPRENTISSAGE D'UNE CASCADE DE CLASSIFIEURS FORTS

Un classifieur plus complexe F_k , dit classifieur boosté ou fort, agrège n classifieurs faibles :

$$F_k = \text{sign} \left(\sum_{i=1}^n c_i f_i - \text{seuil} \right) \quad (11)$$

La somme pondérée (11) des classifieurs faibles est obtenue itérativement par une procédure de boosting décrite au tableau 1. A chaque itération, le seuil de décision τ_i d'un classifieur faible f_i est appris et un poids c_i lui est assigné en fonction de son taux d'erreur sur la base d'apprentissage : plus l'erreur de f_i est petite et plus c_i est élevé. La sélection du classifieur f_i est faite par la procédure *d'Adaboost* (Adaptive boosting [21], [5]) : tous les classifieurs f_k sont en pratique appris, et le classifieur faible retenu f_i est celui qui produit la plus petite erreur sur la base d'apprentissage. Le classifieur ainsi appris est ajouté à la somme (11), et le poids de chaque exemple d'apprentissage est mis à jour de manière à mettre en valeur les exemples mal-classés par la nouvelle somme F_k lors de l'apprentissage du classifieur faible suivant. [21] montre que le taux de détection de F_k peut être arbitrairement grand avec un taux de faux positifs arbitrairement petit en utilisant un nombre suffisant de classifieurs faibles f_i .

A condition que les classifieurs faibles soient un tant soit peu meilleurs que le hasard, c'est-à-dire, comme explicité dans le tableau 1, qu'ils soient capables de classifier correctement au moins la moitié des exemples d'apprentissage. Mais le nombre de classifieurs faibles nécessaire, et donc le temps de calcul, seraient alors excessivement élevés. L'idée introduite par Viola & Jones [2], [3] est d'implémenter une « cascade attentionnelle » : N classifieurs boostés de complexité croissante sont mis en série. Ces classifieurs F_k utilisent un nombre de filtres croissant avec k sur le

schéma de la figure 5. L'apprentissage de la cascade se fait en y ajoutant des classifieurs forts de manière itérative. Pour ce faire, une base de validation, ne contenant pas les exemples de la base d'apprentissage, est employée. Un classifieur fort est appris selon la procédure de boosting décrite dans le tableau 1 : le seuil est choisi de manière à classifier correctement 99.9% des visages et 50% des non-visages de la base de validation. Le nombre T d'itérations est incrémenté jusqu'à ce que ces taux soient atteints. Le classifieur boosté est alors agrégé en série aux précédents. Les exemples d'apprentissage classifiés comme visage par tous les étages précédents constituent la base d'apprentissage du classifieur fort suivant, dont le seuil et le nombre de classifieurs faibles sont aussi déterminés sur la base de validation. Par voie de conséquence, les exemples d'apprentissage sont de plus en plus complexes (i.e. les visages et non-visages sont plus similaires) à mesure que l'on avance dans la construction de la cascade : il en résulte un nombre nécessaire de classifieurs faibles plus important, avec beaucoup plus de filtres « diagonales ». Le nombre de classifieurs forts dans la cascade est alors déterminé par les performances que l'on cherche à obtenir sur la base de validation. En choisissant d'arrêter l'apprentissage de chaque F_k lorsqu'il atteint un taux de détection de 99.9% pour un taux de faux positifs de 50%, pour $k = 1$ à N , alors un ensemble de $N = 25$ classifieurs boostés permet d'atteindre, sur la base de validation, un taux de détection de $99.9\%^{25} = 97.5\%$ pour un taux de faux positifs de l'ordre de 10^{-8} ($50\%^{25}$).

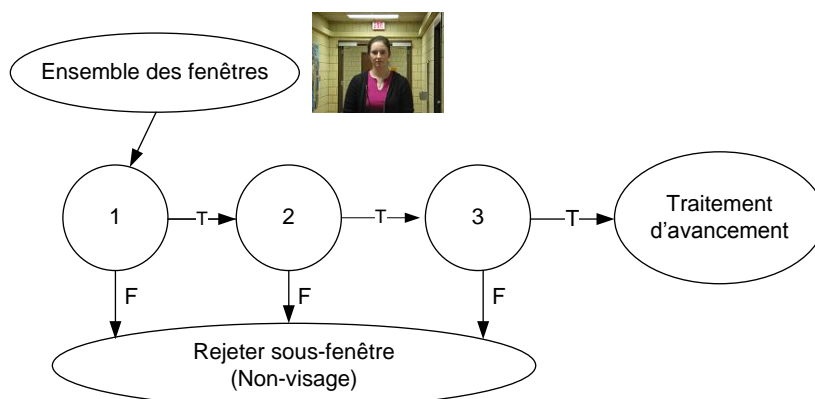


Figure 5 - Cascade de classifieurs boosté : les classifieurs les plus simple et les plus rapides rejettent un maximum de fausses alarmes, les classifieurs les plus complexes se concentrent alors sur les zones de l'image les plus proches de l'apparence d'un visage

Le classifieur le plus simple est le premier auquel les vignettes sont mises en entrée et sert à rejeter les régions les plus simples à discriminer, avant d'appeler les classifieurs suivants, plus complexes, et utilisant des filtres différents. Ainsi, 70 à 80% des candidats sont rejetés par les deux premiers classifieurs boostés, utilisant les filtres les plus simples (contours et lignes) illustrés par les figures 7 et 8. La cascade employée ici est constituée de 25 classifieurs boostés contenant chacun entre 9 et 211 classifieurs faibles, comme le montre le graphique de la figure 6. Le premier étage de la cascade en contient neuf, ils sont illustrés par la figure 7. Ces filtres sont interprétables physiquement. Le premier (7-a) mesure la différence entre la luminosité d'une région contenant les yeux et celle des régions contenant les sourcils et le nez. Le filtre suivant (7-b) analyse la différence entre la luminance de la cloison nasale et celle des yeux, tandis que le troisième compare le nez avec les yeux et le haut de la bouche.

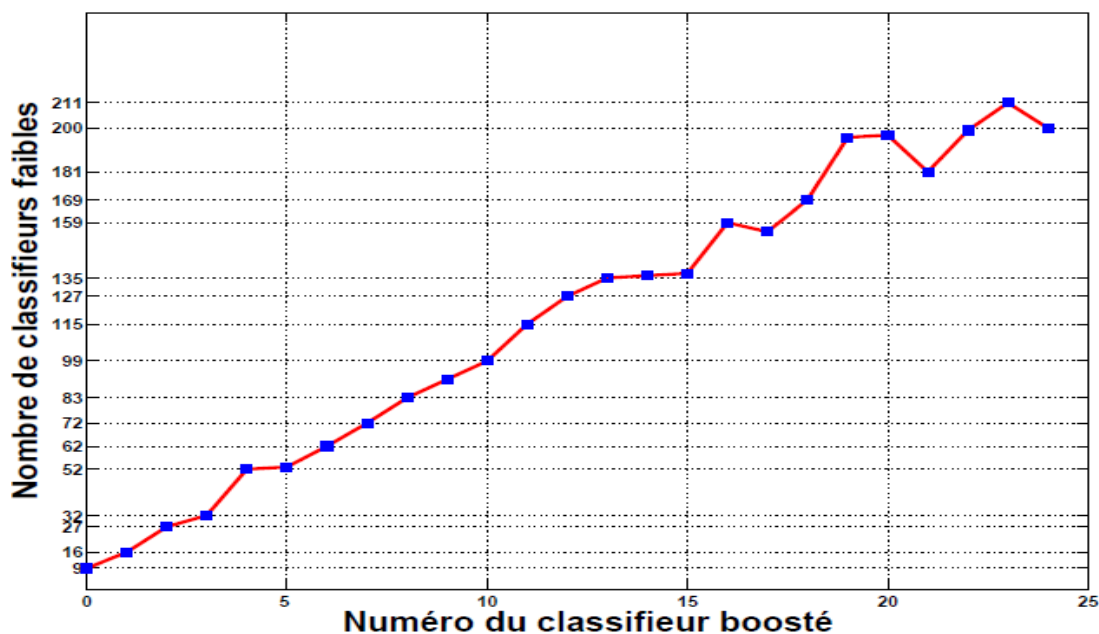


Figure 6 - Nombre de classifieurs faibles contenus dans chacun des 25 classifieurs boostés de la cascade 24x24 [23]

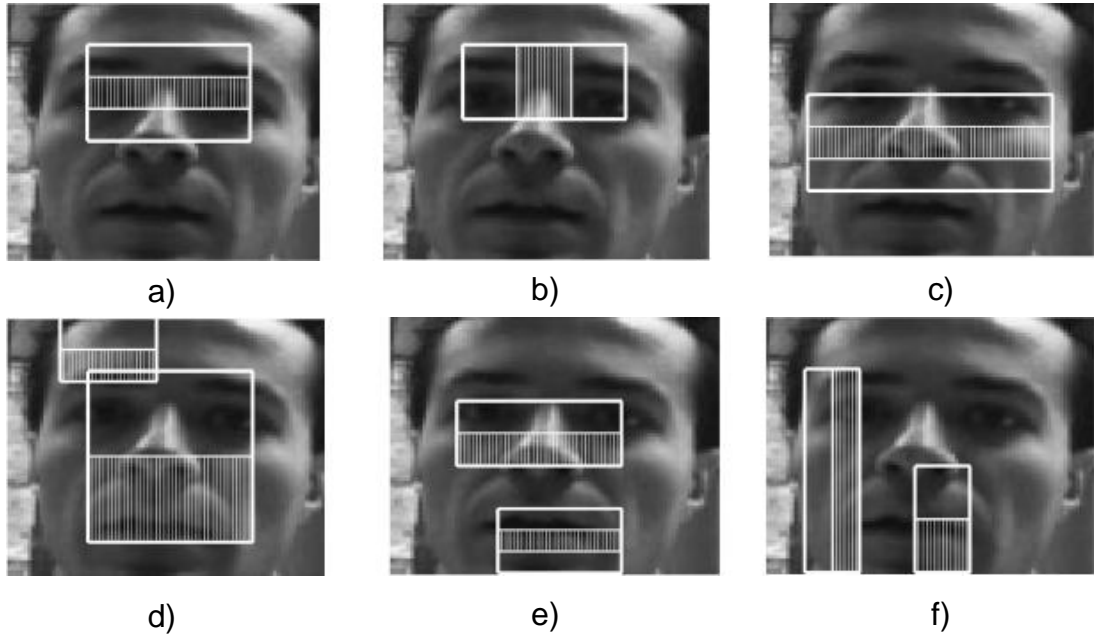


Figure 7 - Les 9 filtres de Haar du premier classifieur fort F_0 de la cascade [23]

Soit une base d'apprentissage $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ de $N = m+1$ exemples x_i étiquetés y_i , m étant le nombre d'exemples négatif ($y_i=0$) et le nombre de positifs ($y_i=1$).

Chaque exemple x_i est une image au niveau de gris de taille 24×24 corrigée en luminosité : $x_i \in \mathbb{R}^{576}$ car $24 \times 24 = 576$

✓ Initialisation des poids : $w_{1,i} = \frac{1}{m}, \frac{1}{l}$ pour $y_i = 0, 1$ respectivement

✓ Pour $t = 1, \dots, T$:

1. Normaliser des poids : $w_{t,i} \leftarrow \frac{w_{1,i}}{\sum_{k=1}^N w_{t,k}}$

2. Apprendre chaque filtre j , le classifieur faible f_j qui lui est associé de manière à minimiser l'erreur $\epsilon_j = \sum_i w_{t,i} |f_j(x_i) - y_i|$ (ϵ_j est nécessairement inférieur à 1 vu la normalisation des poids, et à 0.5 si le classifieur faible f_j est un peu meilleur que le hasard)

3. Sélectionner le classifieur faible f_t ayant la plus petite erreur w_t , lui

affecter un poids $\alpha_t = \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$ (α_t est d'autant plus petit que w_t est grand et est nécessairement ≥ 0)

4. Mettre à jour les poids en pénalisant les exemples les mieux classifiés

$$\text{par } f_t : w_{t+1,i} = \begin{cases} w_{t,i} e^{-\alpha_t} & \text{si } f_t(x_i) = y_i \quad x_i \text{ correctement classifié} \\ w_{t,i} & \text{si } f_t(x_i) \neq y_i \quad x_i \text{ mal classifié} \end{cases}$$

✓ Le classifieur boosté final est alors défini par :

$$F_k = \begin{cases} 1 & \text{si } \sum_{t=1}^T \alpha_t f_t(x) \geq \text{seuil} \\ 0 & \text{si non} \end{cases}$$

Table 1 - L'algorithme d'apprentissage par AdaBoost

2.1.2.3. CLASSIFICATION ET ARBITRAGE ENTRE DETECTION MULTIPLES

En pratique, la rétine à l'échelle n parcourt une ligne sur $\max(2, f_Q^{n-1})$ de l'image, et se déplace horizontalement avec un même pas sauf si la cascade détecte un visage. Dans ce dernier cas la position horizontale suivante n'est pas évaluée par la cascade (voir tableau 2). Le pas varie comme l'échelle à laquelle l'image est parcourue : le visage doit être centré approximativement dans la rétine, comme l'étaient les exemples positifs d'apprentissage, pour être détecté. Ainsi, à plus petite échelle, un pas trop grand pourrait déplacer le détecteur dans l'image de sorte que le visage soit en trop grande partie en dehors de la rétine. A plus grande échelle en revanche un pas trop petit multiplierait le nombre de vignettes à évaluer, et donc le temps de calcul, pour un gain en précision qui n'a pas de sens étant donné l'étape d'arbitrage entre détections multiples qui est opérée par la suite (cet arbitrage est décrit en bas).

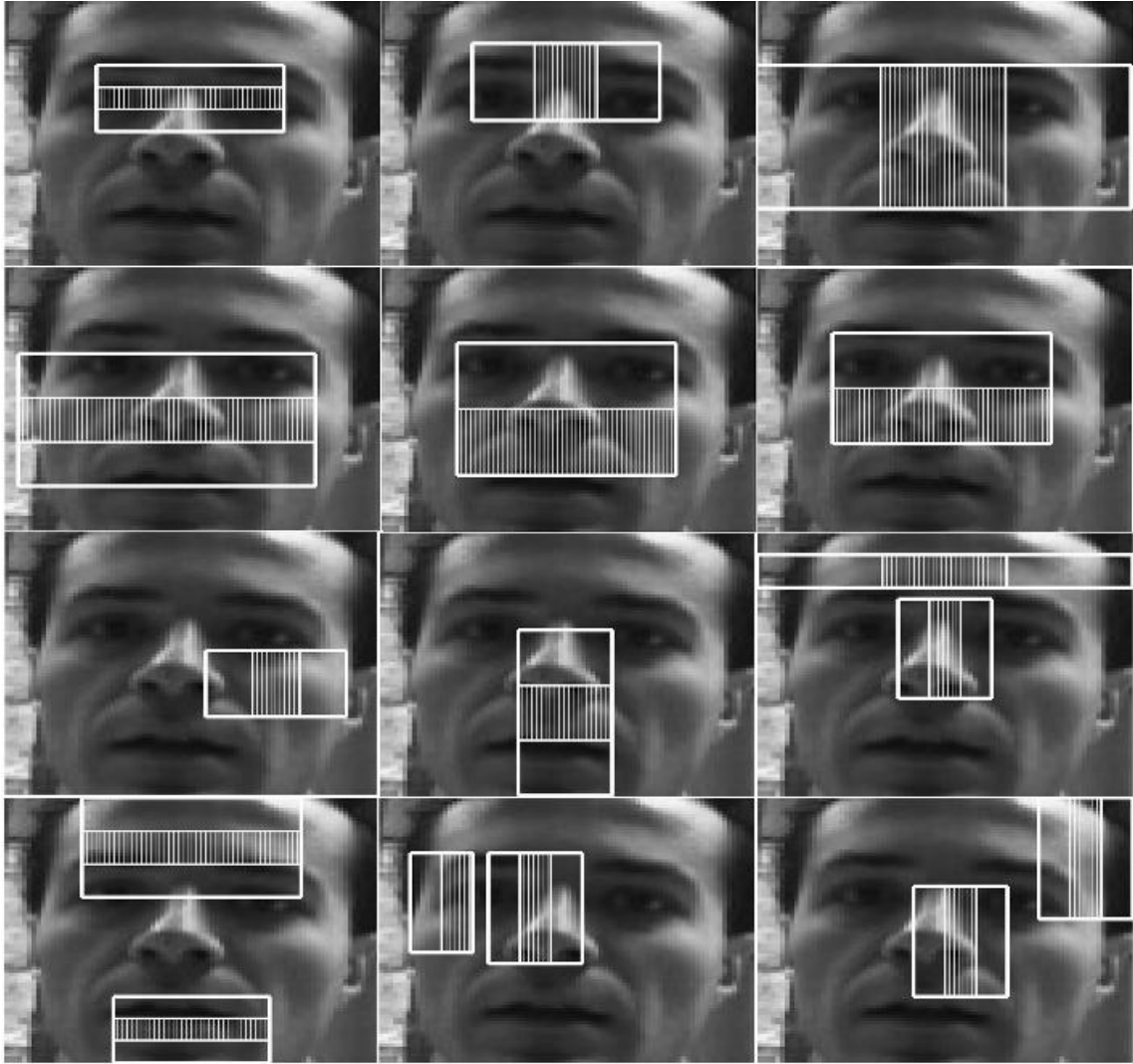


Figure 8 - Les filtres de Haar du second étage F1 de la cascade [23]

Le pas avec lequel l'image est parcourue, horizontalement et verticalement, dépend en particulier de la complexité de l'image : l'image d'un fond blanc, par exemple, sera parcourue à plus de positions qu'une image contenant un visage, mais sera traitée beaucoup plus rapidement comme ayant peu de chance de passer le premier étage de la cascade. Le nombre total de vignettes appliquées en entrée du détecteur dépend donc de la complexité de l'image, mais aussi du nombre et de la taille des visages éventuellement présents dans la scène.

Soient $x_{orig} \times h_{orig}$ la taille de la rétine à l'échelle 1 (24x24 ou 20x20), et $W \times H$ – la taille de l'image. On pose $facteur = f_Q^{n-1}$.

Tant que $x_{orig} \times facteur < W$ et $h_{orig} \times facteur < H$:

- ✓ Agrandir la taille de la rétine, ainsi que celle des filtres de Haar :

$$w_n = \overline{w_{orig} \times facteur} \text{ et } h_n = \overline{h_{orig} \times facteur}$$

- ✓ Poser $step = \max(2, facteur)$, $x_{\max} = \left\lceil \frac{W - w_n}{step} \right\rceil$ et $y_{\max} = \left\lceil \frac{H - h_n}{step} \right\rceil$

- ✓ Pour $y = 0, \dots, y_{\max}$ (par pas de 1)

$$X = \overline{x \times step} \text{ et } x_{step} = 1$$

- ✓ Pour $x = 0, \dots, x_{\max}$ par pas de x_{step} :

$$X = \overline{x \times step}$$

Appliquée la vignette de coin haut gauche (X, Y) et de taille $w_n \times h_n$ en entrée des deux premiers étages F_0 et F_1 de la cascade : $x_{step} = 2$ et passer la vignette au reste de la cascade sinon $x_{step} = 1$

- ✓ Incrémenter n .

Table 2 – Approche rétinienne du détecteur de visage implémenté par OpenCV

2.2. DETECTION DU VISAGE DE PROFILE

La détection du visage de profile d'une personne est une tâche complexe. Le système de détection doit être capable de détecter de façon précise la position du visage dans un environnement non contrainte sur les variations de trois principaux facteurs : la pose, la lumière et les expressions du visage. En effet, l'aspect visuel du visage peut être très variable en fonction du changement de ces facteurs. C'est un problème très intéressant dans le domaine de vision par ordinateur car des algorithmes de détection doivent faire face à la complexité de différentes apparences du visage, surtout dans un environnement non contrainte.

La forme du visage qui est changée selon le mouvement des lèvres et des muscles du visage quand on parle et quand on manifeste ses émotions est très difficile à traiter par des méthodes de détection décrites dans la littérature. Tout d'abord, l'expression du visage modifie sa forme indépendamment de la pose. C'est ce qu'on appelle la déformation intrinsèque. En outre, l'expression du visage est causée par

le mouvement des muscles du visage qui modifie la surface du visage dynamiquement et fabrique de différents bruits comme les rides d'expressions, les ombres, etc.

L'éclairage d'une scène est souvent constant. Par contre, des réflexions et des ombres de la lumière sur un visage peuvent causer des erreurs de détection. Un changement de ces deux facteurs sera une source de modification de l'apparence du visage. Le changement du sens de réflexion de la lumière sur un visage lié à une modification de la pose est également responsable du changement de l'apparence du visage. L'approche appelée "Boosting" a été proposée pour réduire les effets de la lumière et des expressions du visage. Alors, elle est plus robuste aux grandes variations de l'apparence du visage. Cependant, le problème de la pose n'est pas encore résolu dans cette approche.

L'orientation d'un visage dans une image n'est pas toujours frontale. Le plus souvent, le visage subit de grandes rotations en profondeur qui peuvent se produire selon l'un ou les deux axes de l'image. Ces rotations causent des transformations non-linéaires (ou déformations) du visage et peuvent provoquer une occultation partielle du visage dans l'image. Le visage devient donc un objet non-rigide : son contour et les contours des éléments qui le composent varient.

En résumé, dans ce stage, pour la détection du visage de profile, nous voulons construire un système Adaboost en parallèle, le schéma du système est montré dans la figure 9. On trouve que le visage est symétrique et le visage de face a différent apparence avec le visage de profile. C'est la raison pour laquelle dans ce schéma du système, un cascade est entraîné sur les visages de face dont l'orientation varie entre $[-45, +45]$ degrés et un cascade les visages de profile dont l'orientation varie entre $[-90, -45]$ et $[+45, +90]$ degrés.

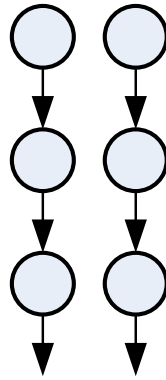


Figure 9 – Deux cascades en parallèle

2.3. BASES DE DONNEES PUBLIQUES

Nous avons utilisé quelques bases d'images de visage dans ce stage. Voici quelques informations détaillées de ces bases.

2.3.1. BANCA

La base de données BANCA (Biometric Access Control for Networked and e-Commerce Applications) sur la ligne <http://www.ee.surrey.ac.uk/CVSSP/banca/> est une base qui contient deux modalités (visage et voix). Deux types d'acquisition ont été effectués pour cette base : une acquisition des images et de la voix en haute définition et une acquisition en mode dégradé. Les personnes dans la base ont été enregistrées selon trois scénarios différents : contrôlés, dégradés et non contrôlés. On a au total 12 sessions différentes avec 4 sessions par scénario. Ces acquisitions s'étalent sur trois mois. Au total, les modalités de 208 personnes ont été capturées, la moitié sur des hommes et la moitié sur des femmes. Les données disponibles pour la communauté scientifique sont celles acquises sur 52 personnes, dont 26 sont des femmes et 26 sont des hommes.

Un nouveau point de fonctionnement a été introduit avec la base BANCA. C'est le point de fonctionnement WER (Weighted Error Rate), donné par la formule (12), correspondant à différentes conditions de fonctionnement. Ces conditions à 3 différentes valeurs du rapport de coût de fausse acceptation (C_{FA}) et de faux rejet

(C_{FR}), $r = \frac{C_{FA}}{C_{FR}}$, avec $r = 0.1$, $r = 1$ et $r = 10$.

$$WER(r) = \frac{P_{FA} + r * P_{FR}}{1 + r} \quad (12)$$

P_{FA} : Probabilité de (FA) et P_{FR} : probabilité de (FR).

Voici quelques exemples de cette base dans trois conditions (contrôlée, dégradée et non-contrôlée).



Figure 10 - Quelques exemples de la base BANCA

2.3.2. MBGC

L'objectif principal de construction de la base d'images MBGC (Multiple Biometric Grand Challenge) sur la ligne <http://face.nist.gov/mbgc> est d'étudier, de tester et d'améliorer la performance de la reconnaissance de visage et d'iris des yeux dans des images fixes et des vidéos. Le MBGC atteint cet objectif grâce à l'utilisation des résultats obtenus de plusieurs domaines de technologie :

- ✓ La reconnaissance de visage sur des images fixes de face frontale à haute et à basse résolution
- ✓ La reconnaissance d'iris des yeux dans des séquences de vidéo et des images hors-angle
- ✓ La fusion de la face et de l'iris
- ✓ La reconnaissance des visages sans contrainte dans des images fixes et des vidéos
- ✓ La reconnaissance dans des vidéos de proche infrarouge (NIR – Near Infrared) et de haute définition (HD – High Definition) prises par des portails

Le MBGC se compose d'un ensemble de problèmes visant à faire progresser l'état actuel de la technologie. Ces problèmes se concentrent sur trois grands domaines:

- ✓ Reconnaissance de visage et d'iris à partir des Vidéos de Portail : le but est de développer des algorithmes qui reconnaissent les gens dans des séquences d'images de NIR et des séquences de vidéo HD. Ces séquences seront acquises quand les gens marchent à travers un portail.
- ✓ Reconnaissance de visage et d'iris à partir des Images Contrôlées : le but est d'améliorer la performance sur des images d'iris et de visage. Des données de visage sont des images de face frontale à haute et à basse résolution. Des images d'iris se composent des séquences d'images fixes et de vidéo.
- ✓ Des images fixes et des vidéos de face : l'objectif est de faire progresser la reconnaissance de visage à partir des séquences de vidéo sans contrainte à l'extérieur et des images fixes.



Figure 11 - Quelques exemples de la base MBGC d'un homme qui marche à travers un portail

2.3.3. FRGC

Les données de FRGC (Face Recognition Grand Challenge) <http://www.frvt.org/FRGC/> se composent de 50 000 enregistrements divisés en deux parts : une part utilisée pour des algorithmes d'entraînement et une part de validation pour évaluer la performance des algorithmes. La part de validation se compose des images de 4003 sessions du sujet. Une session du sujet est l'ensemble des images d'une personne prises à chaque fois quand les données biométriques de cette personne sont collectées. Chaque session se compose de quatre images fixes contrôlées, deux images fixes non-contrôlées et une image en trois dimensions. Les images contrôlées sont prises dans un studio avec des conditions d'éclairage définies et elles sont des faces frontales avec deux expressions : souriant et neutre. Les images non-contrôlées ont été prises dans de diverses conditions d'éclairage, par exemple dans des couloirs, dans des atriums ou à l'extérieur. Les images non-contrôlées ont aussi deux expressions faciales : souriant et neutre. La figure suivante montre un exemple d'une session du sujet dans cette base.

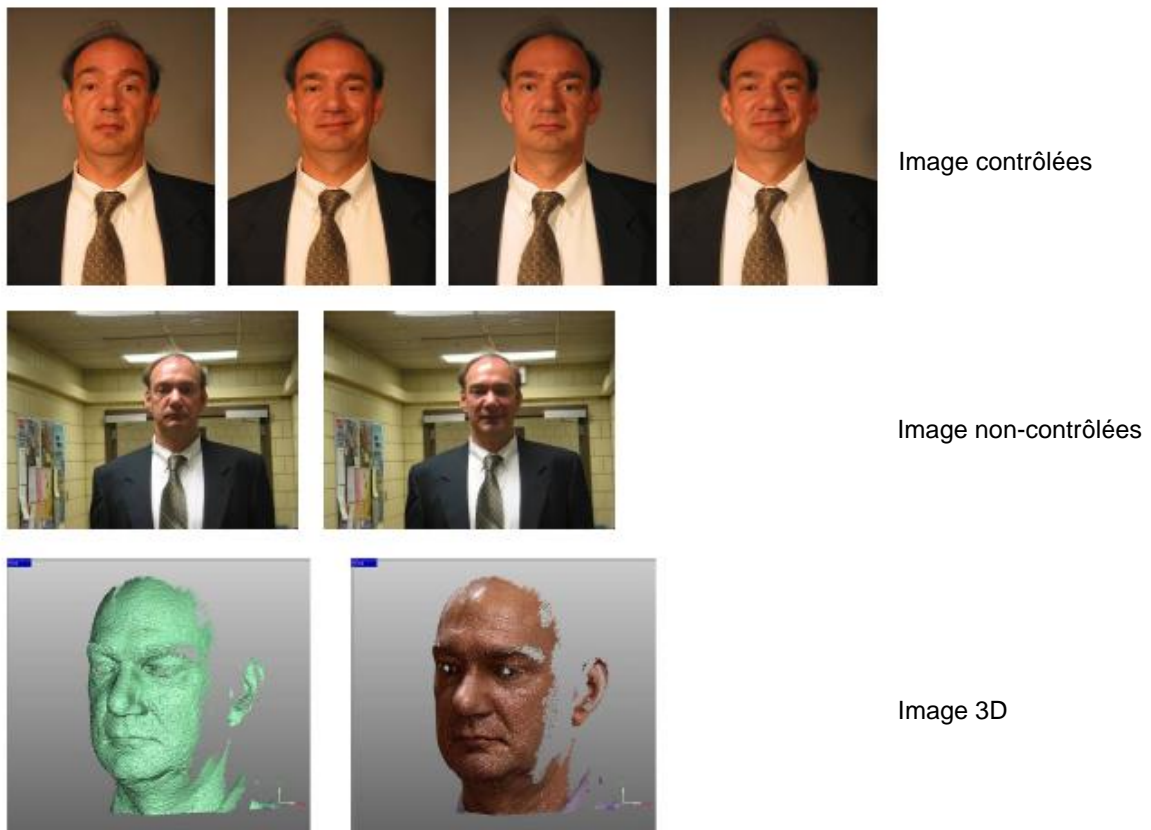


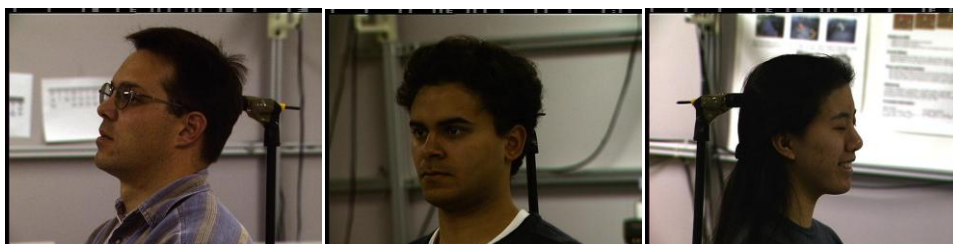
Figure 12 - Exemple d'une session du sujet dans la base FRGC

2.3.4. FPI

FPI (FERET – PIE – Internet) est une base de profile utilisée pour la détection du visage de profile. La plupart des images de cette base vient de la base des images de profile FERET (FacE REcognition Technology) qui se trouve sur la ligne <http://www.frvt.org/FERET/>. La base FERET est créée de façon indépendante par des développeurs d'algorithme de détection de visage et des images de cette base sont collectées dans un environnement demi-contrôlé. Cette base contient 1564 ensembles d'image au total de 14126 images dans lesquelles il y a 1199 personnes et de 365 ensembles doubles d'images. Un ensemble double est un second ensemble d'images d'une personne qui est déjà trouvée dans la base et qui est prise à un autre moment. L'objectif de la construction des ensembles doubles est d'étudier des changements d'une apparence d'une personne qui ont lieu plus d'un an. Une autre partie de la base FPI est prise de la base de données PIE et l'autre est à partir de l'Internet. La base PIE <http://vasc.ri.cmu.edu/idb/html/face/> est fournie pour le but d'évaluer des algorithmes de détection de visages humains de la vue frontale et profile. Cependant, pour la détection de visages de profile, nous avons utilisé seulement une partie de cette base concernant des visages de profile.



Base FERET



Base PIE

Figure 13 - Quelques exemples de la base FERET et PIE

3. EVALUATION DU SYSTEME DE DETECTION DE VISAGE DE FACE

Dans des prochaines parties, nous allons présenter le résultat d'évaluation sur la base de test BANCA. Cette base de données contient 6240 images de face. Chaque image se compose d'un visage de face. La performance d'un système de détection du visage est caractérisée par le taux de détection correcte (Hits Rate) (ou le taux de détection incorrecte Missed) et le taux de fausse alarme (FAR – False Alarm Rate). Dans [3], l'auteur utilise une courbe *Receiver Operating Characteristic* (ROC) qui donne le taux de détection correcte en fonction du nombre total de fausses alarmes sur toutes les images du corpus pour présenter le résultat. Le taux de détection correcte est calculé par la proportion entre le nombre de visages correctement détectés et le nombre d'images dans la base. En plus, le taux de fausse alarme correspond à la proportion du nombre des résultats donnés dans des images mais ils ne sont pas de visage avec le nombre total d'images de la base.

3.1. EVALUATION DU SYSTEME EN UTILISANT LES PARAMETRES PAR DEFAUT

Nous avons évalué la performance du détecteur de Viola & Jones dans OpenCv sur la base BANCA (des détails de cette méthode dans OpenCv se trouvent sur la ligne <http://note.sonots.com/SciSoftware/haartraining.html>). Cette base possède 6240 images dont chacune contient un visage de face. Nous avons fait deux expérimentations différentes sur la même base par le même détecteur de Viola & Jones mais quelques paramètres sont changés. L'étape d'entraînement de l'expérimentation **expt_1** utilise une collection de 785 images négatives (paramètre $-bg$) et une taille d'échantillon 20×20 ($-w\ 20\ -h\ 20$). Des images négatives sont des images utilisées dans l'étape d'entraînement mais elles ne contiennent aucun visage humain. Et l'étape d'entraînement de l'expérimentation **expt_2** utilise une collection de 3804 images négatives et une taille d'échantillon 24×24 ($-w\ 24\ -h\ 24$). Voici le résultat obtenue de ces deux expérimentations avec des paramètres par défaut du système :

	Hits	Missed	False Alarm	Hits Rate	False Alarm Rate
expt_1	6012	228	3063	96.35 %	49.08 %
expt_2	6104	136	2208	97.82 %	35.38 %

Table 3 - Résultats de 2 expérimentations avec des paramètres par défaut

La valeur de détection correcte (Hits) de ces deux expérimentations est élevée (6012 d'**expt_1** et 6104 d'**expt_2**). Cette valeur correspond à une bonne performance de détection : $6012/6240 = 96,35\%$ d'**expt_1** et $6104/6240 = 97,82\%$ d'**expt_2**. Par contre, le nombre de fausses alarmes est beaucoup élevé, 3063 et 2208.

On peut conclure quelques résultats comme les suivants :

- ✓ La valeur de Hits d'expérimentation **expt_2** est supérieure à celle d'**expt_1**.
- ✓ Le système de détection des visages de face abordé dans [3] obtient une bonne performance en fonction du taux de détection correcte.
- ✓ Le nombre de fausses alarmes de deux expérimentations est encore beaucoup élevé.

Dans les parties suivantes, nous allons essayer de diminuer le nombre de fausses alarmes en ajoutant l'étape de post-traitement. Dans cette étape, nous allons utiliser d'autres paramètres comme CV_HAAR_DO_CANNY_PRUNING, CV_HAAR_ROUGH_SEARCH et CV_HAAR_FIND_BIGGEST_OBJECT car ces paramètres peuvent influencer directement la performance du système, surtout le nombre de fausses alarmes.

3.2. EVALUATION DU SYSTEM AVEC D'AUTRES PARAMETRES

3.2.1. DIMINUTION DE FAUSSES ALARMES PAR LA QUALITE DE REGION DETECTEE (CV_HAAR_DO_CANNY_PRUNING)

Avec le paramètre CV_HAAR_DO_CANNY_PRUNING, un détecteur des contours de Canny est utilisé pour éliminer des régions d'image qui possèdent très peu ou trop de contours et elles ne peuvent pas donc contenir des objets et plus précisément dans notre cas, des visages. Par ailleurs, avec ce paramètre, quelques seuils particuliers sont fixés pour la détection de visage et dans ce cas, la diminution

(PRUNING) permet d'accélérer le traitement. Le résultat obtenu sur la base de test BANCA est présenté dans la table suivante :

	Hits	Missed	False Alarm	Hits Rate	False Alarm Rate
expt_1	6069	171	2270	97.26 %	36.38 %
expt_2	6145	95	1669	98.48 %	26.75 %

Table 4 - Résultat de 2 expérimentations avec le paramètre CV_HAAR_DO_CANNY_PRUNING

Avec ce paramètre, la valeur de Hits de deux nouveaux tests est un peu meilleure que celle de paramètre par défaut. C'est parce que des régions qui ont très peu ou trop de contours ne peuvent plus influencer de façon négative sur le résultat de détection. De plus, la valeur de Hits d'**expt_1** est aussi supérieure à celle d'**expt_2**. On va analyser plus concrètement le résultat obtenu dans la table ci-dessus. Pour le test **expt_2**, le nombre de visages de face détectés correctement est augmenté ($6145/6240 = 98,47\%$). Alors, on peut conclure que la caractéristique de Haar est appropriée pour détecter le visage de face. En plus, le nombre de visages de face mal détectés (False Alarm ou fausses alarmes) est également diminué par rapport au cas des paramètres par défaut (1669 par rapport à 2208). En conséquence, avec le paramètre CV_HAAR_DO_CANNY_PRUNING, la performance générale du système est améliorée : le taux de détection correcte est augmenté et le taux de fausse alarme est diminué.

En fait, on trouve qu'autour d'un visage de face dans l'image, il apparaît plusieurs résultats de détection à différentes positions et à différentes valeurs du facteur d'échelle car le classifieur de face n'est pas vraiment sensible avec de petites erreurs de localisation. A l'inverse du cas de bonne détection de visage, il y a beaucoup de détections dans l'arrière-plan quand on change le facteur d'échelle mais elles ne se trouvent pas autour d'une position particulière. Cette idée permet de combiner de bonnes détections et de réduire le nombre de fausses alarmes. Plus précisément, après avoir obtenu des résultats de détection à tous les niveaux d'échelle possibles de l'image, on doit réaliser l'étape de les combiner. En général, cette étape est traitée comme la suivante :

- ✓ Premièrement, si deux régions détectées chevauchent par un pourcentage donné t ou plus de t , ces deux détectations appartiennent à un groupe. Un groupe est un candidat pour la deuxième étape si son nombre de détectations est plus grand qu'un seuil s . Si un groupe n'est pas de candidat, toutes ses détectations sont supprimées.
- ✓ Deuxièmement, la position de détection de visage de face finale est calculée par la moyenne des boîtes englobantes de chaque région détectée. Par ailleurs, chaque boîte englobante est attachée par un poids de sa confiance.

Des valeurs t et s ne sont pas faciles à choisir. Si la valeur t est petite, les détectations d'un même groupe peuvent être appartenues à différents groupes. Si la valeur t est grande, deux groupes de détection peuvent être appartenues à un groupe. On constate le même problème pour le choix de valeur s . Si elle est petite, les fausses alarmes chevauchées peuvent être considérées comme un groupe candidat. Si elle est grande, de vrais groupes candidats peuvent être abandonnés. Alors, le choix des valeurs de ces deux paramètres est difficile et il peut influencer de façon significative la performance du système de détection de face en termes de la vitesse de détection, du nombre de fausses alarmes de la première étape et de la précision de détection de la deuxième étape.

La figure 14 représente la position de quelques détectations correctes. Le résultat de détection est une boîte englobante (un rectangle rouge) qui marque la position du visage dans chaque image.

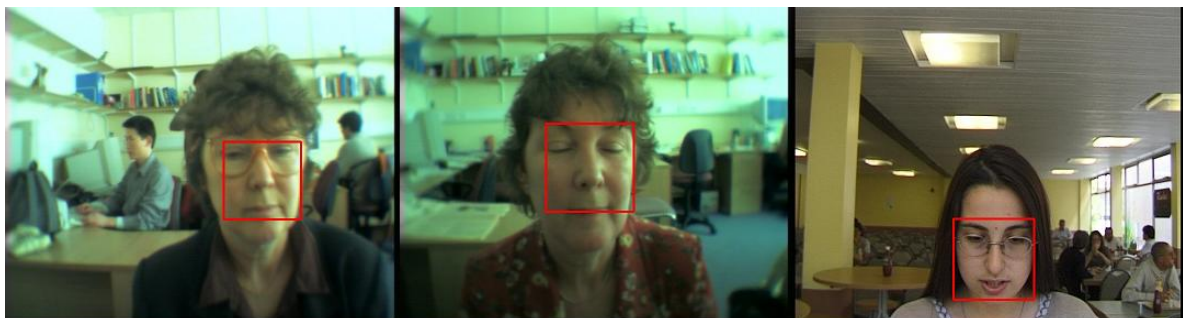


Figure 14 - Bons résultats de détection

A l'inverse, la figure 15 représente le cas où plusieurs fausses alarmes apparaissent. Dans la figure 15a, il y a plusieurs résultats de détection autour du visage correct. Dans les figures 15b et 15c, des fausses alarmes ne se trouvent pas seulement autour du visage correct mais aussi dans l'arrière plan.

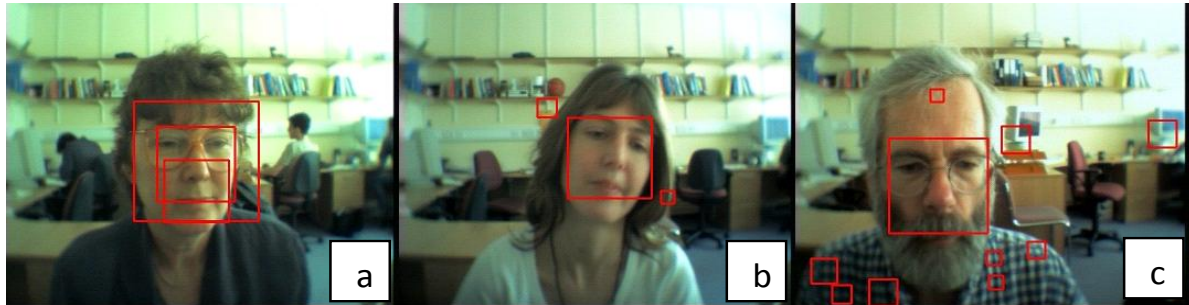


Figure 15 - Résultats avec des fausses alarmes

Dans la figure 16, ce sont des images dans lesquelles le système ne peut détecter aucun visage même des images contrôlées de la base BANCA comme la figure 16a. La raison principale est que toutes les personnes dans ces images se courbent (leurs têtes sont baissées) et elles portent des lunettes. C'est la raison pour laquelle leurs visages se déforment et des caractéristiques des visages sont totalement changées. Cependant, on peut constater que l'homme qui porte des lunettes dans les figures 15c et 16c est le même. Alors, dans ce cas, sa tête baissée est le facteur le plus décisif qui influence la précision de détection.



Figure 16 - Exemples de non-détection

Pour montrer le changement de la performance du système quand on modifie la qualité des images et l'environnement de la prise des images, nous avons testé la détection de visage du système sur trois parties différentes de la base BANCA : contrôlées, dégradées et non-contrôlées comme nous avons abordé dans la partie [2.3.1](#). Chaque partie contient 2080 images de 52 personnes avec des conditions de

l'environnement différentes. Voici le résultat obtenu pour chaque partie en utilisant des paramètres d'**expt_2**.

Images	Hits	Missed	False Alarm	Hits Rate	False Alarm Rate
Contrôlées	2062	18	170	99.13 %	8.17 %
Dégradées	2060	20	1222	99.04 %	58.75 %
Non-Contrôlées	2023	57	277	97.26 %	13.32 %

Table 5 - Résultat de chaque catégorie de la base BANCA avec le paramètre CV_HAAR_DO_CANNY_PRUNING

A partir de ces résultats, on peut constater la participation de chaque partie de la base dans le résultat total d'**expt_2** dans le tableau précédent. On trouve que le taux de détection correcte est diminué quand la qualité d'images est basse ($2060/2080 = 99,04\%$ des images dégradées) et surtout quand des conditions de l'environnement ne sont pas contrôlées ($2023/2080 = 97,26\%$ des images non-contrôlées). Cependant, on peut voir un grand changement dans le taux de fausse alarme. Pour des images contrôlées, l'arrière-plan est un rideau bleu et la lumière est définie. C'est la raison pour laquelle le taux de fausse alarme est très bas car le système ne commet pas beaucoup d'erreurs dans la détection de visage. Si on change l'arrière-plan des images et la lumière des scènes est combinée de l'éclairage des lampes et la lumière naturelle comme dans le cas des images non-contrôlées, le nombre de fausses alarmes est augmenté. Par contre, si la qualité d'images est diminuée dans le cas des images dégradées, des zones dans des images deviennent plus floues et alors des contours des objets dans l'image ne sont pas bien détectés, la méthode utilisant le détecteur des contours de Canny ne représente pas d'avantages par rapport à d'autres approches. En conséquence, son taux de fausse alarme est fortement augmenté.

3.2.2. TERMINAISON DE DETECTION DES OBJETS (CV_HAAR_ROUGH_SEARCH)

La différence entre le paramètre CV_HAAR_ROUGH_SEARCH et le paramètre précédent CV_HAAR_DO_CANNY_PRUNING est que pour ce paramètre, l'étape de détection des objets dans l'image est terminée immédiatement après avoir détecté le premier candidat d'objet en se basant sur un nombre suffisant de voisins d'objet

pour considérer un candidat comme un objet. Voici le résultat obtenu de deux expérimentations pour ce paramètre :

	Hits	Missed	False Alarm	Hits Rate	False Alarm Rate
expt_1	6086	154	5055	97.53 %	81.01 %
expt_2	6152	88	3584	98.59 %	57.43 %

Table 6 - Résultat de 2 expérimentations avec le paramètre CV_HAAR_DO_ROUGH_SEARCH

Avec ce paramètre, on peut dire que le résultat de détection de face est plus mauvais que celui du paramètre CV_HAAR_DO_CANNY_PRUNING dans le tableau 4 et celui des paramètres par défaut dans le tableau 3. Plus précisément, le nombre de fausses alarmes du paramètre CV_HAAR_DO_ROUGH_SEARCH est très élevé mais la valeur de Hits reste presque la même pour toutes les deux expérimentations. Ça montre que la performance générale du système est diminuée. La raison principale est la façon de choisir immédiatement le premier objet détecté à chaque niveau de balayage qui ne représente pas exactement la position du visage dans l'image. La figure 17 montre quelques mauvais résultats de la détection de visage par cette méthode.

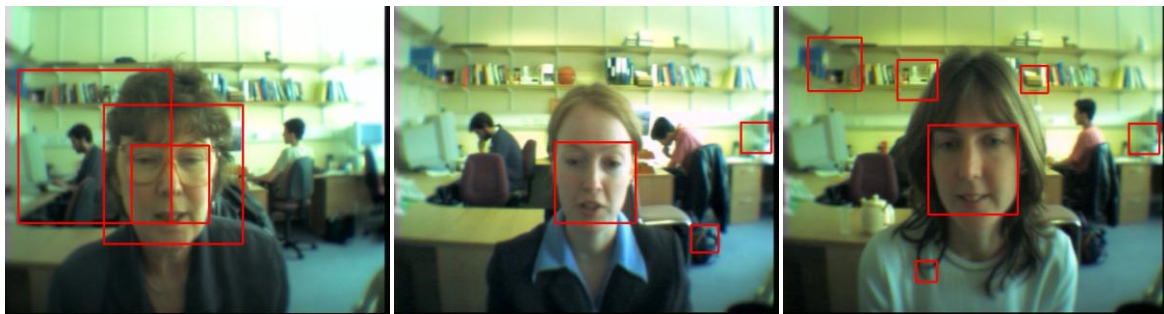


Figure 17 - Résultats de détection avec des fausses alarmes

On trouve que pour ces exemples, il apparaît plusieurs fausses alarmes dans le résultat final à différents niveaux de balayage. En résumé, la performance générale du système en utilisant ce paramètre n'est pas améliorée, voire même diminuée, et ce paramètre n'est pas approprié à la détection de visage de face.

3.2.3. UTILISATION DE CONNAISSANCES A PRIORI (CV_HAAR_FIND_BIGGEST_OBJECT)

Si ce paramètre est utilisé, le résultat de détection final de chaque image est le seul objet dont la taille est la plus grande. Pour les images qui contiennent plus d'un objet et dans notre cas, des visages, ce paramètre ne peut pas alors être utilisé. Cependant, pour des images qui ne contiennent qu'un objet, ce paramètre nous permet d'obtenir une performance de détection très élevée et le temps de traitement beaucoup diminué par rapport à d'autres paramètres. La table suivante montre le résultat obtenu en utilisant le paramètre CV_HAAR_FIND_BIGGEST_OBJECT.

	Hits	Missed	False Alarm	Hits Rate	False Alarm Rate
expt_1	6060	180	119	97.11 %	1.91 %
expt_2	6137	103	61	98.35 %	0.98 %

Table 7 - Résultat de 2 expérimentations avec le paramètre CV_HAAR_FIND_BIGGEST_OBJECT

Bien que le taux de détection correcte de ce paramètre ne soit pas vraiment augmenté en comparaison de celui des paramètres CV_HAAR_DO_CANNY_PRUNING et CV_HAAR_DO_ROUGH_SEARCH, la performance de cette approche est bien représentée par le taux de fausse alarme. Grâce au choix de la taille des objets détectés dans chaque image, ce paramètre permet d'enlever de petits objets de non-visage qui apparaissent dans l'arrière-plan et de prendre l'objet le plus grand dans l'ensemble des objets détectés autour du visage réel dans chaque image. C'est la raison pour laquelle le taux de fausse alarme est très bas dans le résultat final.

La figure 18 représente quelques exemples de détection correcte. Ces trois images contiennent un seul visage de face et l'arrière-plan est assez simple : l'arrière-plan des images 18.a et 18.b n'a qu'une couleur et l'arrière-plan de l'image 18.c est un peu plus complexe mais en général, cet arrière-plan et le visage de cette image sont différents en termes de la taille, la couleur et la texture. En plus, les trois visages frontaux dans cette figure sont vraiment symétriques : deux joues, deux yeux, le nez, la bouche, etc.

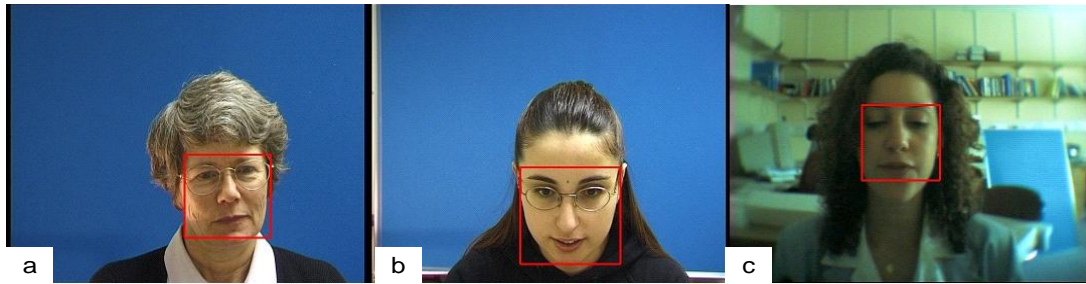


Figure 18- Bons résultats de détection

La figure 19 illustre l'effet du paramètre CV_HAAR_FIND_BIGGEST_OBJECT pour diminuer le nombre de fausses alarmes. Dans cette image, il n'y a qu'un visage mais comme on peut voir dans l'image 19a, le système a détecté 7 visages dans le cas normal. Chaque visage est représenté par une boîte englobante en rouge. Autrement dit, le nombre de fausses alarmes est 6. Si on utilise ce paramètre, la taille de la boîte englobante qui se trouve exactement à la position du visage est la plus grande. La figure 19b est le résultat final de détection sous l'effet de post-traitement avec le paramètre CV_HAAR_FIND_BIGGEST_OBJECT. Le résultat final de détection est vraiment très bon et il n'y a pas de fausses alarmes.

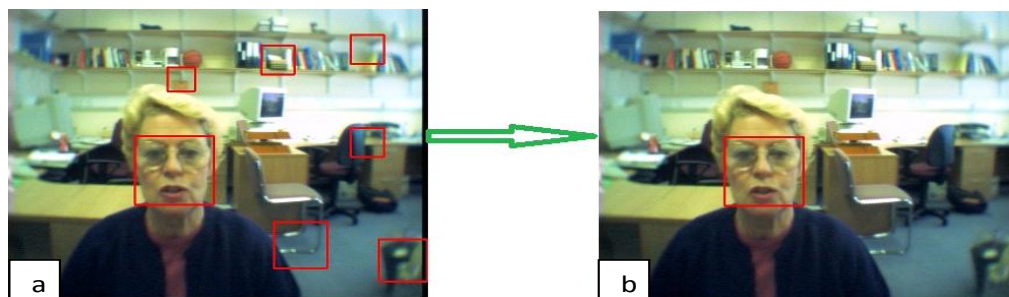


Figure 19 - Position de détection correcte

Comme le terme BIGGEST marqué dans le nom de ce paramètre, le résultat final de détection est la boîte englobante la plus grande dans l'ensemble des boîtes englobantes obtenues de différents niveaux de balayage. Dans l'image à gauche de la figure 20a, il y a 7 boîtes englobantes au total mais la boîte la plus grande ne se situe pas correctement à la position exacte du visage. Par conséquent, après avoir appliqué l'étape de post-traitement avec le paramètre CV_HAAR_FIND_BIGGEST_OBJECT pour diminuer le nombre de fausses alarmes, le résultat final est la boîte englobante la plus grande en bas du visage comme on peut la voir dans l'image à droite de la figure 20.a. Un autre exemple du

résultat de détection incorrecte se trouve dans la figure 20b. On peut voir dans l'image à gauche de la figure 20b, il apparaît deux boîtes englobantes mais la boîte la plus grande se situe à côté du visage. Alors, dans le résultat final, la position de la boîte de détection ne coïncide pas avec la position exacte du visage.

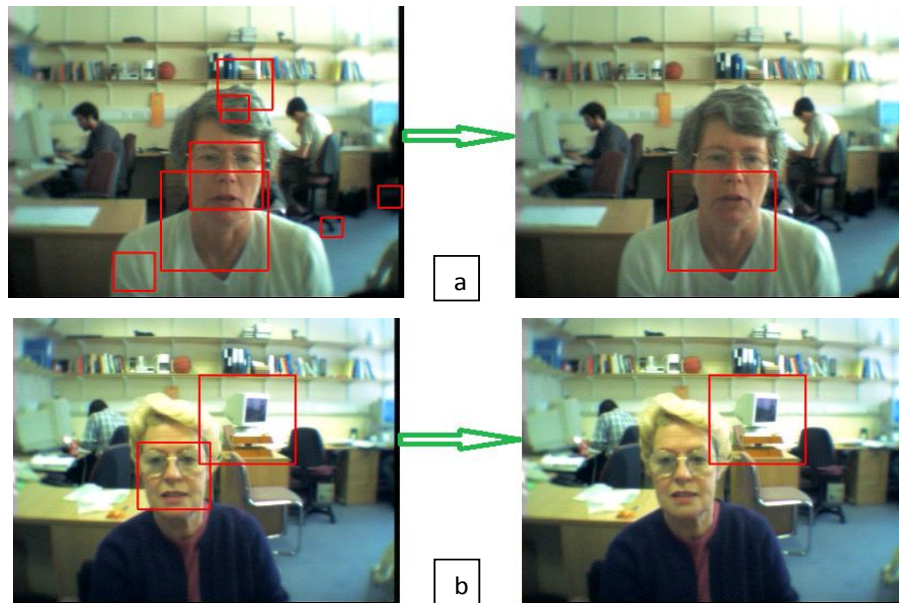


Figure 20- Résultats de détection avec la position fausse

En résumé, même s'il existe des résultats de détection de mauvaise position, cette approche qui utilise le choix la taille des boîtes englobantes des objets détectés représente des avantages par rapport à d'autres méthodes en termes de la performance générale de détection, surtout de la diminution du taux de fausse alarme.

Afin d'examiner la participation de chaque catégorie de la base BANCA dans le résultat final comme nous avons fait dans l'évaluation du paramètre CV_HAAR_DO_CANNY_PRUNING, nous avons réalisé la détection de visage de façon séparée sur trois catégories de la base : contrôlée, dégradée et non-contrôlée (partie [2.3.1](#)). Chaque catégorie possède 2080 images de 52 personnes avec des conditions différentes de l'environnement. Le tableau 8 montre le résultat obtenu pour chaque catégorie en utilisant des paramètres d'**expt_2**.

Images	Hits	Missed	False Alarm	Hits Rate	False Alarm Rate
Contrôlées	2062	18	0	99.13 %	0.00 %
Dégradées	2047	33	30	98.41 %	1.44 %
Non-Contrôlées	2028	52	31	97.50 %	1.49 %

Table 8 - Résultat de chaque catégorie de la base BANCA avec le paramètre CV_HAAR_BIGGEST_OBJECT

En comparaison des valeurs dans le tableau 5 du paramètre CV_HAAR_DO_CANNY_PRUNING, le taux de détection correcte dans le tableau 8 du paramètre CV_HAAR_BIGGEST_OBJECT est également diminué pour des images dégradées ou des images dont l'arrière-plan et la lumière de scène sont complexes. Par contre, il y a des points très particuliers pour le paramètre CV_HAAR_BIGGEST_OBJECT. Il n'existe aucune fausse alarme dans le cas des images contrôlées grâce au rideau bleu qui est très différent par rapport au visage humain et à la source lumineuse claire de la scène. A l'inverse du résultat du paramètre CV_HAAR_DO_CANNY_PRUNING, le taux de fausse alarme des images dégradées et celles non-contrôlées de ce paramètre reste presque le même. Ça montre que pour cette approche, la qualité des images n'influence pas fortement le taux de fausse alarme par rapport à d'autres facteurs comme la lumière ou la complexité de l'environnement.

3.3. CONCLUSION

Dans l'étape de post-traitement, nous avons appliqué de différents traitements pour diminuer le nombre de fausses alarmes. En testant de différents paramètres du détecteur de Viola & Jones sauf les paramètres CV_HAAR_DO_CANNY_PRUNING, CV_HAAR_DO_ROUGH_SEARCH et CV_HAAR_FIND_BIGGEST_OBJECT par deux expérimentations **expt_1** et **expt_2**, on constate que l'expérimentation **expt_2** représente une bonne performance en comparaison de celle d'**expt_1**. La différence principale de ces deux expérimentations est le nombre d'images négatives utilisées dans l'étape d'entraînement et la taille d'échantillon. La comparaison de la performance de trois paramètres pour l'expérimentation **expt_2** est illustrée dans la figure 21. La performance générale en examinant à la fois le taux de détection correcte et le taux de fausse alarme est la plus élevée pour l'utilisation du paramètre

CV_HAAR_FIND_BIGGEST_OBJECT. Le résultat de deux autres paramètres est presque semblable même si le paramètre CV_HAAR_DO_CANNY_PRUNING représente une performance un peu plus élevée que celle du paramètre CV_HAAR_DO_ROUGH_SEARCH.

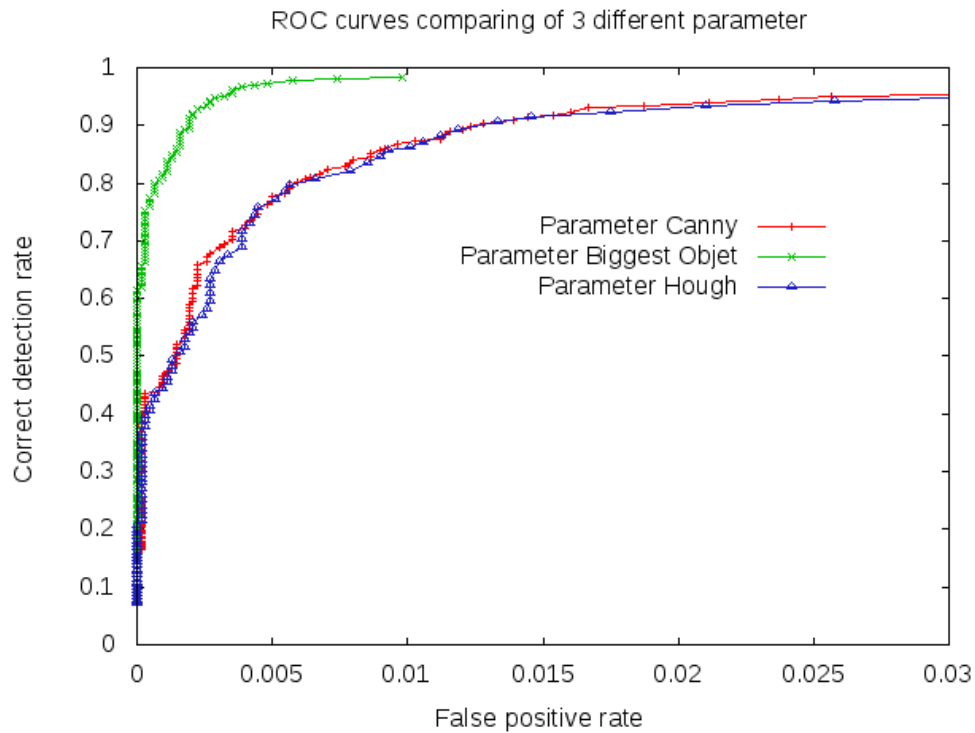


Figure 21 - Comparaison de la performance de trois paramètres

4. EVALUATION DU SYSTEME DE DETECTION DU VISAGE DE PROFILE

Dans cette partie, nous allons présenter la performance du détecteur de visage de profile en utilisant l'algorithme *AdaBoost* et des caractéristiques de *Haar* dans la partie 2.1.2. Pour pouvoir détecter le visage dans des images, cette approche doit se baser sur la base de données d'entraînement FPI. Cette base se compose de deux parties différentes : les images de visage de profile et les images de non-visage. Les images originales de cette base viennent de la base d'images FERET, PIE et sur l'Internet.

- ✓ La base de données du visage de profile : elle contient 5000 images de visage de profile de détournage et le détournage de ces images est réalisé de manière manuelle à partir d'une vérité de terrain.
- ✓ La base de données de non-visage : cette base contient 3800 images qui ne possèdent aucun visage.

4.1. EN UTILISANT LES CARACTERISTIQUES DE HAAR

Dans cette partie, nous présentons la performance du système de détection sur la base de données de test CMU (Carnegie Mellon University) qui se trouve sur la ligne http://vasc.ri.cmu.edu/idb/html/face/frontal_images/ pour le test de détection de visage de profile. Cette base se compose de 208 images et elle contient totalement 400 visages de profile. Le résultat de détection est comme le suivant :

Hits	Missed	False Alarm	Hits Rate	False Alarm Rate
116	284	57	29.00 %	14.25 %

Table 9 - Résultat de la détection de visage de profile

Le nombre de visages bien détectés est 116, ce qui correspond au taux de détection correcte $116/400 = 30\%$. Le nombre de fausses alarmes est 57 et il nous semble que ce résultat est acceptable même si le taux de détection correcte n'est pas vraiment élevé. Quelques exemples des visages bien détectés sont illustrés dans la figure 22. La raison principale est que des visages dans cette figure ont des caractéristiques

visuelles très différentes et marquantes par rapport à celles de l'arrière-plan de ces images. En plus, des caractéristiques visuelles du corps des personnes et de l'arrière-plan dans ces images sont assez simples. Par contre, les visages dans des images de la figure 23 ont des caractéristiques semblables avec les visages de la figure 22 mais l'arrière-plan de ces images est un peu plus complexe. C'est la raison pour laquelle dans le résultat il apparaît quelques fausses alarmes.

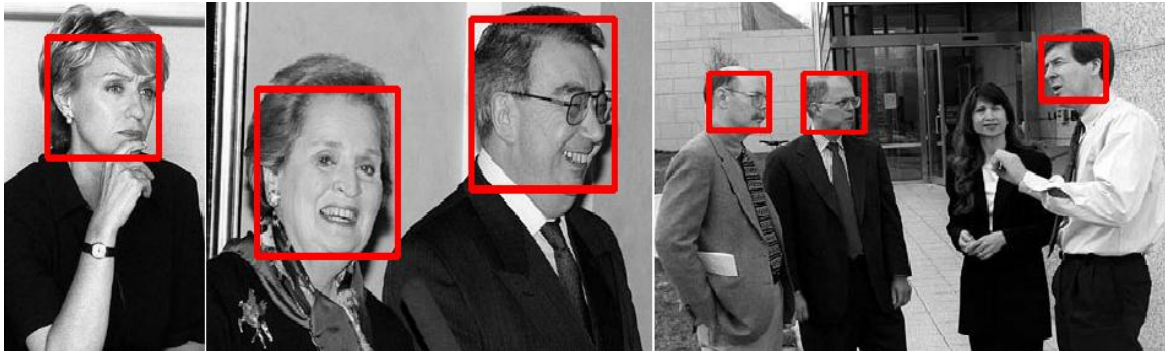


Figure 22 - Résultats de détection correcte avec l'image de profile

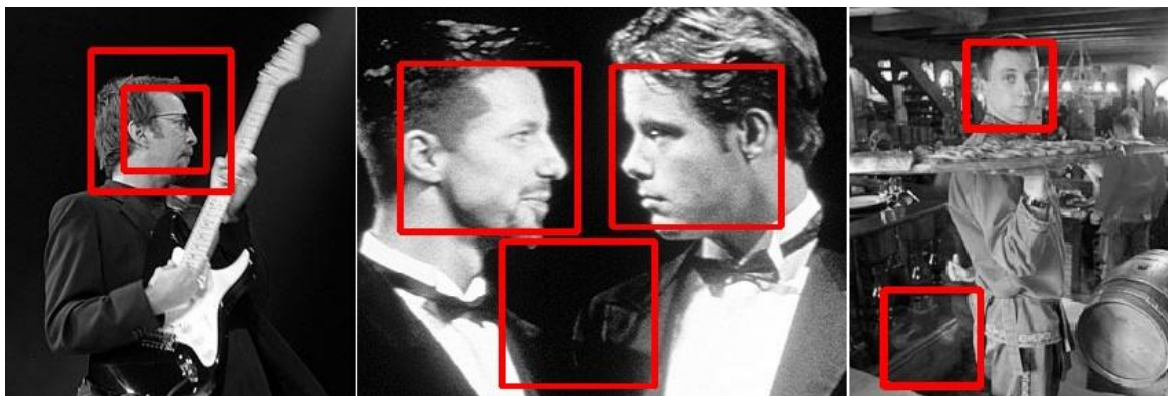


Figure 23 - Résultats de détection avec des fausses alarmes

La figure 24 montre des exemples qui sont totalement différents avec ceux de deux figures précédentes. Premièrement, des caractéristiques visuelles comme la couleur, la texture des visages sont pareilles avec celles de l'arrière-plan de ces images. Puis, la lumière dans ces images est brillante. À cause de la lumière, une partie des visages de profile est sombre et l'autre est éblouie. On peut voir l'effet de la lumière sur des visages dans la troisième image de figure 24.



Figure 24- Exemples de non-détection

4.2. CONCLUSION

La performance de détection de visage de profile en utilisant les caractéristiques de Haar est montrée dans la figure 24. A partir de ce graphique, on peut dire que les caractéristiques de Haar dans la partie [2.1.2](#) ne sont pas convenables pour la détection de visage de profile. La nature des caractéristiques de Haar est symétrique. C'est pourquoi la détection de visage frontal obtient de bons résultats avec les caractéristiques de Haar. Néanmoins, des visages de profile n'ont pas de caractéristiques symétriques et ça donne de mauvaises détections dans le résultat.

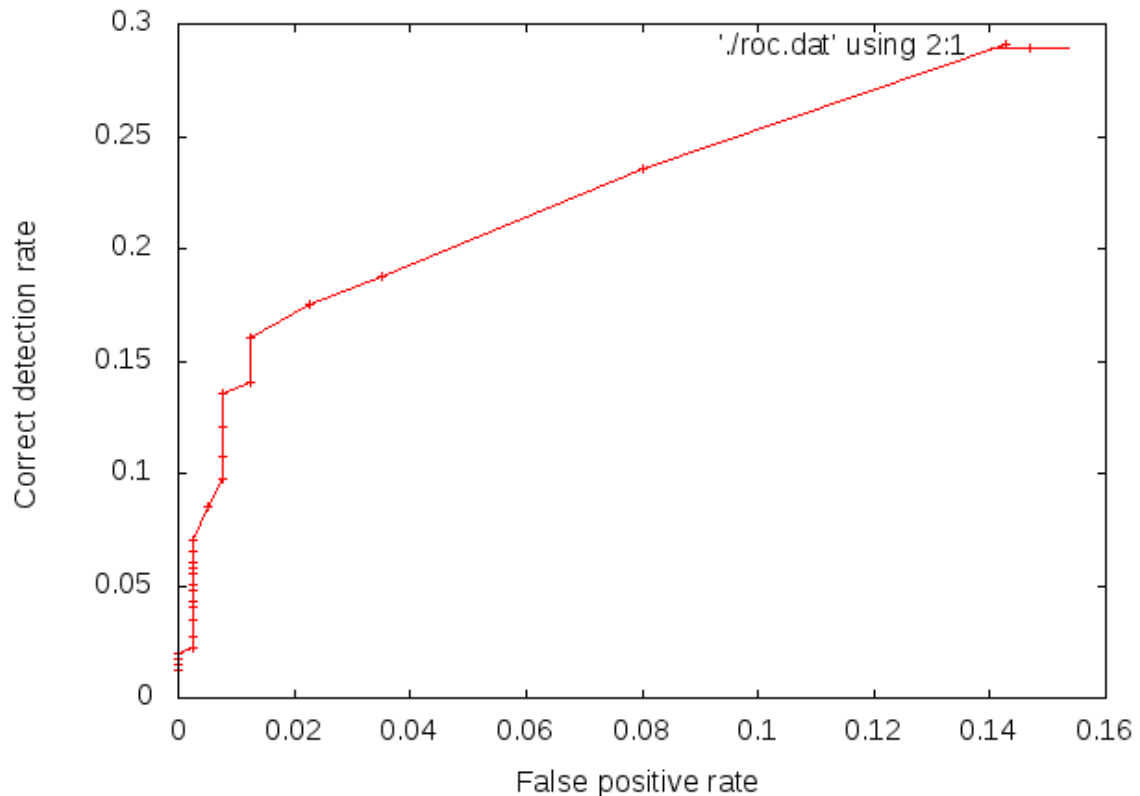


Figure 25 - Courbe ROC de détection du visage de profile en utilisant les caractéristiques de Haar

5. CONCLUSION ET PERSPECTIVES

Mon stage se situe dans le domaine de détection de visage pour les images fixes. La détection de visage pour des images fixes est une recherche très importante et aussi très difficile dans le domaine de la vision par ordinateur. Jusqu'à maintenant, des approches pour résoudre le problème de détection de visage sont développées en se basant sur l'extension de l'étape de boosté ou des caractéristiques utilisé par Viola & Jones. Le détecteur de Viola & Jones qui se base sur l'algorithme AdaBoost nous apporte le meilleur résultat en termes de la précision de détection et du temps réel. Par contre, les approches en se basant sur le détecteur de Viola & Jones doivent prendre beaucoup de temps pour l'étape d'entraînement. En plus, il est difficile de choisir les paramètres optimaux pour la cascade.

Dans mon stage, nous avons étudié le détecteur de Viola & Jones, un algorithme très populaire utilisé dans la détection de visage de face. Nous avons également évalué la performance de ce détecteur sur la base de test BANCA. Après avoir évalué tous les résultats obtenus, nous pouvons dire que la performance de détection de visage est affectée par quelque paramètres comme : la base de données pour l'entraînement, le paramètre utilisé pour balayer l'image dans l'étape de détection, le critère de fusion la détection couverte. Par ailleurs, une partie importante de mon stage est d'utiliser des caractéristiques de Haar et l'algorithme AdaBoost pour la détection de visage de profile. Nous avons évalué cette approche sur la base d'images de profile de test CMU. Le résultat obtenu nous montre que des caractéristiques de Haar ne sont pas convenable pour la détection de visage de profile.

Pour les perspectives, il nous reste encore plusieurs problèmes à améliorer la détection de visage de profile. Pour avoir un système de détection de visage efficace, on doit construire un système qui peut détecter de différentes poses du visage. On pourrait construire ce système grâce à un arbre de détection. Plus précisément, le nœud racine de cet arbre couvre le plus grand espace de face. Puis, il est divisé en sous-espaces. A la deuxième couche de l'arbre, elle se compose des

visages de profile à gauche, des visages de face et des visages de profile à droite. A la troisième couche, on l'utilise pour représenter des visages de ROP (Rotation Off Plane). En fin, à la dernière couche, on précise chaque catégorie des visages en correspondant des visages de RIP (Rotation In Plane).

6. BIBLIOGRAPHIE

- [1] M.H. Yang, D. Kriegman et N. Ahuja, « Detecting faces in images: A survey », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 24, pages 34–58, 2003.
- [2] E. Hjelm et B. K. Low, « Face Detection: A Survey », *Computer Vision and Image Understanding*, volume 83, pages 236-274, 2001.
- [3] P. Viola et M. Jones, « Rapid object detection using a boosted cascade of simple features », *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511-518, 2001.
- [4] Z. Lei, C. Wang, Q. Wang et Y. Huang, « Real-Time Face Detection and Recognition for Video Surveillance Applications », *WRI World Congress on Computer Science and Information Engineering*, volume 5, pages 168 – 172, 2009.
- [5] Y. Freund et R. E. Schapire, « A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting », *In European Conference on Computational Learning Theory*, volume 904, pages 23-37, 1995.
- [6] R. Lienhart, A. Kuranov et V. Pisarevsky, « Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection », *Pattern Recognition*, volume 2791, pages 297-304, 2003.
- [7] T. Sakai, M. Nagao et S. Fujibayashi, « Line extraction and pattern detection in a photograph », *Pattern Recognition*, volume 1, pages 233–236, 1969.
- [8] Y. Gizatdinova et V. Surakka, « Feature-Based Detection of Facial Landmarks from Neutral and Expressive Facial Images », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 28, pages 135-139, 2006.
- [9] A. J. Colmenarez et T. S. Huang, « Face detection with information based maximum discrimination », *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 782 – 787, 1997.

- [10] K-K. Sung et T. Poggio, « Example-based learning for view-based human face detection », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 20, pages 39-51, 1998.
- [11] H. Schneiderman et T. Kanade, « A statistical method for 3D object detection applied to faces and car », *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 746 – 751, 2000.
- [12] C. Garcia, G. Zikos et G. Tziritas, « Wavelet Packet Analysis for Face Recognition », *Image and Vision Computing*, volume 18, pages 289 – 297, 2000.
- [13] M.-H. Yang, D. Kriegman et N. Ahuja, « Face Detection Using Multimodal Density Models », *Computer Vision and Image Understanding*, volume 84, pages 264-284, 2001.
- [14] C. Liu, « A Bayesian Discriminating Features Method for Face Detection », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 25, pages 725 – 740, 2003.
- [15] E. Osuna, R. Freund et F. Girosit, « Training support vector machines: an application to face detection », *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 130-136, 1997.
- [16] H.A. Rowley, S. Baluja et T. Kanade, « Neural Network-Based Face Detection », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 20, pages 23-38, 1998.
- [17] R. Feraud, O.J. Bernier, J.-E. Viallet et M. Collobert, « A Fast and Accurate Face Detection Based on Neural Network », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 23, pages 42-53, 2001.
- [18] Y. Rodriguez, « Face Detection and Verification using Local Binary Patterns », *Thèse N° 3681 du laboratoire de l'IDIAP de l'École Polytechnique Fédérale de Lausanne*, 2006.
- [19] J. Wu, S.C. Brubaker, M.D. Mullin et J.M. Rehg, « Fast Asymmetric Learning for Cascade Face Detection », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 30, pages 369-382, 2008.

- [20] R. Lienhart et J. Maydt, « An extended set of haar-like features for rapid object detection », *2002 International Conference on Image processing*, volume 1, pages 900 – 903, 2002.
- [21] Y. Freund et R.E. Shapire, « Experiments with a new boosting algorithm », *In Proceedings of the Thirteenth International Conference on Machine Learning*, pages 148-156, 1996.
- [22] L.K. Hansen et P. Salamon, « Neural Network Ensembles », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 12, pages 993-1001, 1990.
- [23] R. Belaroussi, « Localisation du Visage dans des Images et Séquences Vidéo Couleur », *Thèse de Doctorat de l'Université Pierre et Marie Curie*, 2006.
- [24] S.Z. Li et A.K. Jain, « Handbook of Face Recognition », Springer-Verlag New York Inc., 2005.