Détection des yeux, du nez et de la bouche par filtres de Haar adaptatifs

N.Pyun[1,2] , M. Marmouget[2] et N. Vincent[1]

[1]Université de Paris Descartes, Paris, France
[2]Konbini, Paris, France

Résumé
L'extraction des yeux, du nez et de la bouche du visage humain sont des tâches largement étudiées dans le domaine de la reconnaissance de formes. Localiser ces régions anatomiques pertinentes du visage est souvent la première étape de nombreuses approches de la vision par ordinateur, comme la segmentation, la reconnaissance ou l'identification de personne, la reconnaissance de l'expression ou de l'émotion du visage, la localisation de points d'intérêts, l'estimation de pose ou encore le suivi du visage. La télésurveillance, l'indexation automatique ou semi-automatique d'images ou de vidéos, la robotique sont autant de domaines applicatifs. Dans cet article, nous proposons une méthode basée sur l'analyse des lignes horizontales. Elles sont extraites d'une carte d'énergie calculée sur des filtres de Haar adaptatifs. L'introduction de connaissances, notamment sur les positions des différentes régions anatomiques pertinentes, ainsi que sur leurs relations spatiales nous permet de les séparer. Une des difficultés majeures de la détection des éléments anatomiques pertinents du visage réside dans la variabilité de l'illumination d'un visage à l'autre, mais aussi des conditions d'illumination inégale sur un visage donné. Afin de rendre la méthode robuste à ces variations d'illumination, nous proposons une analyse multi-seuils capable de choisir, pour chaque région anatomique, un seuil adéquat sur la carte d'énergie horizontale. Notre approche est testée sur les bases BioID, Color FERET et LFW et montre des résultats prometteurs.

Extracting human eyes, nose and mouth are widely studied tasks in pattern recognition. Finding localization of these relevant face anatomic regions is often the first step in many approaches in computer vision, such as segmentation, person recognition or identification, facial expression or emotion recognition, landmarks localization, head pose estimation or face tracking. Such methods are used in many applications such as telemonitoring, automatic and semi-automatic indexation of images and videos, or in robotics. In this paper, a method based on horizontal lines analysis is proposed. Lines are extracted from a energy map computed on adaptive Haar-like features. Briging knowledge, in particular, related to positions of these relevant facial anatomic regions, as well as their relative positions enable to separate them. One of the difficulties of detecting these components lies in illumination variability of a face compared to another, as well as inegal illumination on a single given face. To overcome these illumination variations which often occur, we propose a multi-thresholds analysis able to choose, for each anatomic region, a suitable threshold of the energy map. Our approach is tested on BioID, Color FERET and LFW databases and shows promising results.

Mots clé : Œil, yeux, nez, bouche, Haar, carte d'énergie, analyse multi-seuils, relations spatiales, connaissance

1. Introduction

Extraire les caractéristiques du visage est une étape nécessaire dans de nombreuses applications, comme la reconnaissance de visage [JP09], l'estimation de pose [MCT09], le suivi du visage [MZW10] ou encore la reconnaissance de l'expression faciale [LZM12]. En particulier, les positions des yeux, du nez et de la bouche sont des informations souvent recherchées. Par exemple, dans le suivi du visage, de nombreuses méthodes recourent aux AAM (Active Appareance Models) qui font preuve d'efficacité et de précision. La première étape des AAM [TCT01] implique l'apprentissage de la déformation des visages en appliquant une analyse en composante principale sur les positions et intensités de points saillants (du contours des yeux, du nez et de la bouche) du visage. La seconde étape consiste à la mise en correspondance d'un jeu de points avec le modèle issu de l'apprentissage. Le principal inconvénient des AAM réside dans la nécessité de placer manuellement ces points lors de la phase d'apprentissage. Bien que trouver les boîtes englobantes des yeux, du nez et de la bouche soit insuffisant pour

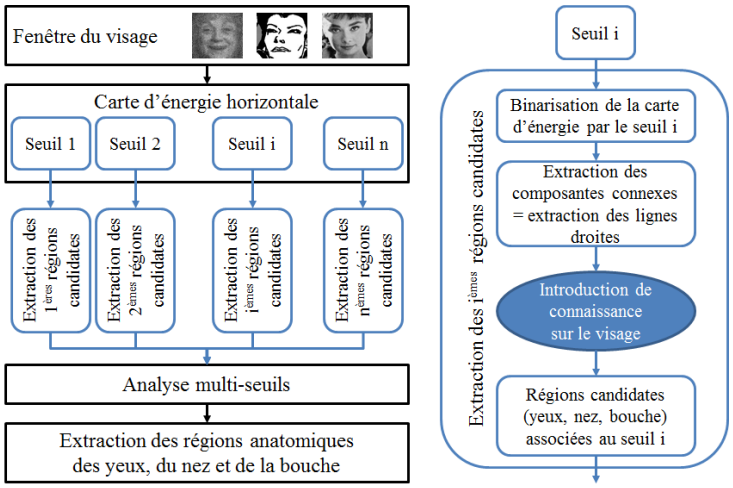

Figure 1: *Schéma global de la méthode.*

déterminer les points saillants, cela permettrait d'améliorer la précision, puisque ces boîtes englobantes réduisent la zone de recherche.

Il existe de nombreuses méthodes capable d'extraire des caractéristiques à partir des visages. Elles peuvent être regrouper en trois catégories. La première est constituée des approches basées sur l'apparence [IC13], [LR12], [qZhC12]. Dans cet ensemble, le but est d'extraire des caractéristiques locales dans un sous-espace approprié en faisant appel aux techniques d'apprentissage. Par exemple, dans [VP05], GentleBoost est utilisé sur des filtres de Gabor locaux ou encore dans [AB10], l'analyse en composantes principales est appliquée sur les LTP [TT10]. Les AAM font partie de cette catégorie. Ces méthodes nécessitent une étape d'apprentissage, et donc une base conséquente d'images annotées. Cependant, la littérature montre qu'elles parviennent à obtenir de bons résultats.

La seconde catégorie est constituée d'approches où un template est utilisé. Dans ces méthodes [DZZ14], la première partie consiste à définir un template spécifique d'une partie du visage (par exemple les yeux). Généralement, ce template contient les contours et cherche à modéliser les possibles déformations. Puis, différents candidats sont extraits avant de les comparer au template. Dans [JH09], les auteurs extraient les yeux en utilisant un template multi-angle. Les candidats sont extraits à l'aide d'opérations morphologiques et des informations sur la symétrie des yeux. Dans [AYH89], un template déformable est utilisé. Celui-ci se déforme en minimisant un coût afin de trouver la meilleure correspondance. Le principal inconvénient de ces méthodes réside dans la difficulté à les généraliser lorsque les conditions de vue, d'illumination ou d'échelle changent. Par exemple, un template définissant des yeux ne parvient à en trouver que lorsqu'ils sont ouverts. Par contre, ces approches ont l'avantage d'être capables de trouver ce qu'elles cherchent dans des images où n'apparaitraient que l'élément recherché.

Enfin, la dernière catégorie regroupe des approches qui incluent de la connaissance et des informations spatiales sur le visage [DPM11], [ZZ12] . Dans [GS07], Les points d'intérêt sont détectés automatiquement sur des visages aux expressions variées. Des informations spatiales sont introduites pour améliorer la précision de la localisation de ces points. Dans [KP97], les auteurs proposent une méthode de détection de visage qui requiert des règles relatives aux informations spatiales du visage.

La méthode que nous proposons appartient à cette dernière catégorie. Le but est d'extraire les rectangles englobants des yeux, du nez et de la bouche. Elle inclut de la connaissance sur le visage, notamment sur la distribution spatiale de ces caractéristiques sur le visage. Par exemple, les yeux se situent sur la partie supérieure du visage, le nez et la bouche sont alignés sur l'axe de symétrie du visage, etc. Afin de parvenir à de bonnes détections et malgré la variation d'échelle, une carte d'énergie s'adaptant à l'échelle du visage est proposée. Les connaissances anatomiques sur les tailles des différents éléments recherchés, ainsi que leurs positions relatives nous permettent de fixer certaines limites dans nos calculs.

La section suivante présente la méthode de façon générale. la section 3 est consacrée à l'extraction des régions anatomiques candidates. La section 4 décrit l'extraction des régions anatomiques finales grâce à une analyse multi-échelle. La section 5 est consacrée à l'évaluation de la méthode.

2. Architecture générale

2.1. Vue générale de la méthode

Dans la littérature, les éléments anatomiques du visages sont souvent déterminés individuellement ; certains recherchent les yeux, d'autres la bouche, ou plus rarement le nez. Pourtant, il est certain qu'ils peuvent être utiles ensemble, par exemple pour établir un modèle 3D ou encore pour faire de la reconnaissance. C'est donc l'objectif de cet article. Toutefois, nous nous plaçons ici au niveau des boites englobantes et non de points anatomiques. L'avantage de se placer à ce niveau est la possibilité de détecter l'élément anatomique en question, malgré la présence visible de points

anatomiques due à des occlusions ou à des conditions d'illumination inégales. On pourrait rechercher indépendamment chaque élément anatomique, car ils ne sont pas toujours illuminés de la même manière. Toutefois, en les recherchant ensemble, nous pouvons utiliser leurs positions relatives qui constituent une source d'information.

La figure 1 présente une vue globale de la méthode. D'un point de vue général, la méthode proposée est constituée de trois étapes principales. Tout d'abord, à partir des fenêtres de visage, une carte d'énergie horizontale est calculée. Dans un second temps, nous cherchons à extraire des régions anatomiques candidates. En effet, les conditions d'illuminations peuvent varier d'un élément anatomique à l'autre, un seuil global, appliqué à la carte, serait alors insuffisant. A ce stade, nous avons donc quatre ensembles distincts regroupant respectivement les candidats de l'œil droit et gauche, du bout du nez et de la bouche. La dernière étape consiste à choisir, pour une région anatomique donnée, un candidat par une analyse multi-seuil. Comme nous venons de voir, la deuxième étape consiste à extraire les régions anatomiques candidates. Pour un seuil de binarisation donné de la carte d'énergie horizontale, on extrait les composantes connexes (CC), ce qui revient à extraire les lignes horizontales. Puis en introduisant de la connaissance sur le visage, notamment sur les positions et tailles relatives des différents éléments anatomiques, nous parvenons à extraire les boites englobantes candidates de l'œil droit, gauche, du nez et de la bouche associées à un seuil de binarisation.

2.2. Carte d'énergie horizontale

Tout d'abord, les visages sont extraits dans des fenêtres rectangulaires, par exemple en utilisant la méthode de Viola et Jones [VJ01] ou encore les LBP [TOM02]. Nous supposons que la détection du visage réussit ; le visage est détecté dans son intégralité. Ainsi, l'échelle du visage correspond approximativement à la taille de la fenêtre englobante du visage. Les lignes du visage qui correspondent aux régions saillantes sont essentiellement horizontales (voir figure 2). Cette section décrit les étapes qui nous permettent d'extraire les lignes horizontales.

Figure 2: *Détection de lignes verticales et horizontales du visage par convolution.*

De nombreuses approches permettent d'extraire les lignes horizontales. Dans une transformée de Fourier, nous nous intéresserions aux fréquences verticales. Plus récentes, les transformées en ondelettes apportent beaucoup plus d'information tout en conservant une certaine localité à l'information. Les ondelettes de Haar mettent en œuvre des échelles différentes. La difficulté est alors de choisir le bon niveau d'observation et donc le critère qui permet de le déterminer en fonction des coefficients calculés, et aussi de sélectionner les meilleurs coefficients. Grâce à la connaissance de la fenêtre englobant le visage, nous avons une idée de l'échelle des éléments recherchés. Les largeurs et hauteurs de la bouche et des yeux sont du même ordre de grandeur, la taille de la matrice de convolution utilisée peut donc être fixée et ne dépend que de la taille de la fenêtre du visage. La méthode de Viola et Jones recherche la nature et la taille des bons motifs à appliquer lors de la détection. Puisque la nature (lignes horizontales) et l'échelle des éléments recherchés sont connues, la matrice de convolution de notre méthode correspond au motif horizontal décrit par Viola et Jones (Figure 3).

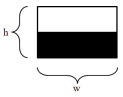

Figure 3: *Filtre de Haar horizontal de hauteur h et de largeur w.*

D'un côté, si le détecteur est trop local, les résultats sont trop bruités. De l'autre, s'il ne l'est pas assez, le détecteur donne des résultats où manquent des informations. Nous utilisons une matrice de convolution horizontale dont la taille dépend seulement de celle de la fenêtre de visage. Soit H et L, la hauteur et largeur respectives de la fenêtre de visage, nous définissons h et l, la hauteur et largeur de la matrice de convolution par la formule (1).

Figure 4: *Les fenêtres de visages sont sur la première ligne, les cartes d'énergie horizontales normalisées E_{norm} sont sur la seconde. plus la valeur est faible, plus le pixel est noir.*

$$\begin{cases} h = max(2; 2 \cdot \alpha) \; et \; l = max(2; 3 \cdot \alpha) \\ avec \; \alpha = min(H/40; L/40) \end{cases} \quad (1)$$

Toutes ces remarques nous conduisent à appliquer à la fenêtre de visage I, un filtre de convolution de noyau H_{lh} (équation (2)).

$$J = I * H_{lh} \quad (2)$$

La carte d'énergie horizontale est alors donnée par l'équation (3).

$$E(X,Y) = |J(X,Y)| \quad (3)$$
$$E(X,Y) = \left| \sum_{blanc} p(x_1,y_1) - \sum_{noir} p(x_2,y_2) \right|$$

Puis, la carte d'énergie est normalisée en E_{norm} par la valeur maximale M de la partie centrale ($a = 1/3 \cdot X$ et $b = 2/3 \cdot X$) de E par l'équation (4) et (5). En effet cette partie centrale ne contient que des pixels issus du visage tandis que les tiers droit ou gauche contiennent souvent des éléments de l'arrière-plan.

$$M = \max_{a<X<b} E(X,Y) \quad (4)$$

$$E_{norm}(X,Y) = 1 - min\left(1, \frac{E(X,Y)}{M}\right) \quad (5)$$

Cette normalisation nous donne l'ordre de grandeur de la variation de la carte d'énergie au niveau du visage. Après cette normalisation, et contrairement à E, plus une valeur dans E_{norm} est faible, plus nous sommes confiant en la présence d'une ligne horizontale dans le voisinage. La Figure 4 montre quelques exemples de cartes d'énergie horizontales normalisées.

3. Extraction des régions anatomiques candidates

Une fois E_{norm} obtenue, nous devons extraire les lignes à direction horizontale. Différents seuils sont appliqués sur E_{norm}. La Figure 5 montre des cartes d'énergie binarisées à différents seuils d'un même visage. Malgré la normalisation de la carte d'énergie, un seuil adéquat global pour tous les visages n'existe pas. De plus, comme l'illumination sur un visage donné peut être irrégulière, un seuil fixe ne peut être appliqué à tous les éléments de ce visage. Pour toutes ces raisons, les régions anatomiques candidates (RAC) des yeux, du nez et de la bouche sont détectés pour chaque seuil. Cette section concerne l'extraction de RAC pour un seuil fixe donné.

Figure 5: *Seuillage de la carte d'énergie. L'image de gauche est la carte d'énergie normalisée, les autres sont les images binaires de cette carte d'énergie (de gauche à droite, les seuils sont de 0,4 ; 0,6 ; 0,8 et 0,99).*

Après le seuillage, les composantes connexes (CC) sont extraites ainsi que leur boîtes englobantes. Le but est alors de regrouper les boîtes représentant respectivement chacun des yeux droit et gauche, le nez et la bouche. Tout d'abord, nous calculons le nombre de pixels appartenant aux CC sur chaque ligne. Toutes les CC qui ont une projection commune sur l'axe des ordonnées sont fusionnées pour former une RAC. Deux RAC consécutives sont fusionnées en une si la distance maximum entre elles est plus petite que $max(1; H/40)$, comme le montre la figure 6.

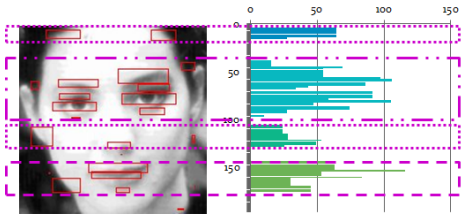

Figure 6: *Histogramme des largeurs des CC sur l'axe des ordonnées. 4 RAC sont extraites de cet histogramme.*

Après cette étape, une liste de RAC rangée est construite. Chaque RAC est représentée par les CC inclues, le point supérieur gauche S et le point inférieur droit T. Cependant, lorsque H ou W est inférieur à 60 pixels, les projections des CCs sur l'axe des ordonnées ne sont plus séparées ; chaque CC devient alors un RAC. Puis, les RAC sont rangées par rapport à l'ordonnée y_S.

3.1. Extraction des RAC des yeux

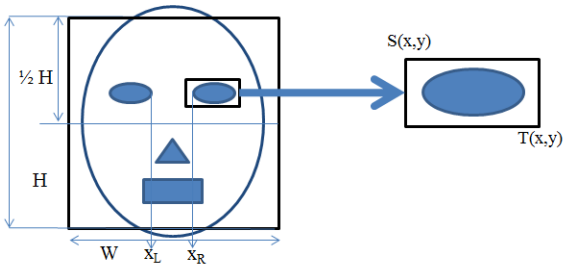

Figure 7: ***Gauche*** *: Connaissances basiques utilisées dans notre méthode.* ***Droite*** *: Exemple d'une RAC définie par le point supérieur gauche S et le point inférieur droit T.*

Afin d'extraire les régions saillantes du visage, des informations sur la distribution spatiale du visage sont utilisées (Figure 7). Puisque les yeux sont situés sur la partie supérieure du visage. Ils sont contenus dans la RAC dont l'ordonnée ${y_S}^*$ respecte les conditions (6).

$$\begin{cases} {y_S}^* < 1/2 \cdot H \\ {y_S}^* = min(|y_S - 1/2 \cdot H|) \end{cases} \quad (6)$$

Les RAC au-dessus de la RAC des yeux sélectionnée et proches du bord supérieur de la fenêtre du visage sont supprimées. Sinon, elles sont fusionnées avec la RAC des yeux. Par exemple dans l'exemple de la Figure 6, seules les deux RAC supérieures sont, dans un premier temps prises en compte. Comme la RAC la plus proche du bord supérieur

est trop éloignée de l'autre, seule la RAC proche du centre de l'image est prise en compte.

Notons qu'à ce moment, seule l'ordonnée et la hauteur des deux sont identifiés dans la RAC des deux yeux.

Puis, les RAC de l'œil droit et gauche sont extraites à partir de la RAC des deux yeux. Les occurrences des boîtes englobantes des CC contenues dans la RAC des deux yeux sont projetées sur l'axe des abscisses. Les rectangles ayant une projection commune sur l'axe des abscisses sont fusionnés pour former une RAC d'un seul œil., comme le montre la figure 8.

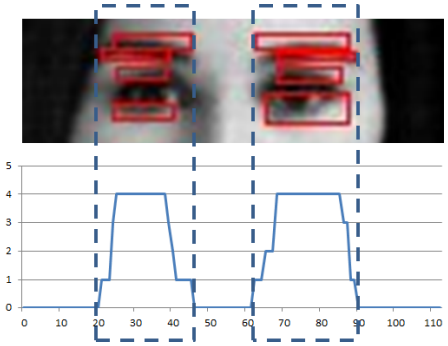

Figure 8: *Projection des occurrences des CC sur l'axe des abscisses de la RAC des deux yeux. les CC avec une projection commune sont fusionnées pour former la RAC de l'œil droit et gauche.*

Deux cas se présentent alors à nous : soit les projections des RAC de l'œil droit et gauche sont séparées, soit elles ne le sont pas. La région des yeux n'est plus une seule RAC, mais une liste de RAC. Puisque les yeux droit et gauche sont les plus significatifs dans la RAC des yeux. Seules les deux RACS ayant les aires les plus élevées sont conservées.

Si la seconde plus grande RAC a une aire ne dépassant pas 10% de celle ayant l'aire la plus élevée, alors elle n'est pas prise en compte. Sinon, la RAC ayant la plus petite abscisse devient la RAC de l'œil droit et l'autre devient la RAC de l'œil gauche.

Souvent, les projections des yeux droit et gauche ne sont pas séparées. Cela arrive lorsque L et H sont basses ou lorsque le seuillage de la carte d'énergie est élevé ou encore lorsque le sujet porte des lunettes. Si, nous n'avons toujours qu'une seule RAC au lieu de deux, afin de séparer la RAC des deux yeux en deux, nous calculons l'histogramme des occurrences des pixels de la carte d'énergie binarisée sur l'axe des abscisses. Bien que les yeux droit et gauche ne soient pas séparés sur cet histogramme, les valeurs entre les deux yeux sont nettement inférieures à celles se trouvant au niveau des yeux. Nous utilisons alors un simple level set supérieur dont la valeur de séparation est de la moitié de la valeur maximale de l'histogramme (Figure 9). Les deux RAC les plus larges sont alors prises en compte et forment les RAC des yeux droit et gauche. Afin de maintenir la cohérence des RAC, les points S, T ainsi que la taille des CC contenues dans chaque RAC détectée, sont modifiées en conséquence.

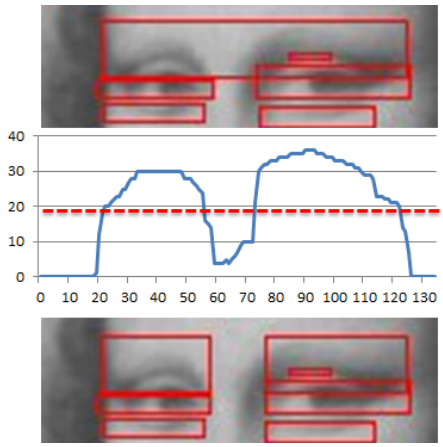

Figure 9: *Projection des pixels se trouvant sur les CC de la RAC des deux yeux. Le level set supérieur permet de séparer la RAC des yeux en RAC de l'œil droit et gauche.*

On note que les RAC de l'œil droit et gauche détectées sont parfois trop larges, car elles incluent les sourcils ou l'arcade sourcilière. D'un autre côté, le level set supérieur permet d'exclure les sourcils ou l'arcade sourcilière par rapport à l'axe des abscisses. C'est pourquoi un level set supérieur est encore une fois appliqué sur chacune des RAC des yeux. Au final, nous obtenons les deux RACS correspondant chacun à l'œil droit et gauche.

3.2. Extraction des RAC du nez et de la bouche

Ici, nous introduisons une nouvelle connaissance commune du visage humain : le nez et la bouche sont situés sur l'axe de symétrie du visage. Cet axe passe par l'espace qui sépare les deux yeux. Ainsi, le nez et la bouche ont une partie située entre les yeux.

A partir des RAC de l'œil droit et gauche, les abscisses x_R et x_L de l'intervalle IGD entre les yeux sont déterminées. La première étape consiste à conserver seulement les boîtes englobantes dont la projection verticale intersecte IGD. Les CC qui ne sont pas inclues dans une des RAC des yeux et dont l'abscisse d'au moins un point de celles-ci est comprise dans $[x_L, x_R]$ sont conservées. En d'autres termes, seules les CC des yeux et celles situées sur l'axe de symétrie verticale du visage sont conservées. Cette étape permet de supprimer les CC situées sur le bord du visage comme le montre la Figure 10.

Puis, afin de séparer la RAC du nez de la RAC de la bouche, une autre connaissance basique sur le visage humain est utilisée. La bouche a une largeur plus importante que celle de la base du nez. Parmi les CC conservées celle qui possède la plus grande largeur (CC_{bouche}) appartient à la bouche. D'un autre côté, la CC qui est la plus proche des yeux (CC_{nez}) appartient au nez.

Toutes les CC en-dessous de CC_{bouche} appartiennent à la bouche. Quant aux CC restantes, elles sont comparées et classifiées en fonction de leur largeur et celles de CC_{bouche} et CC_{nez}. A partir de ces deux ensembles de CC, celui qui se

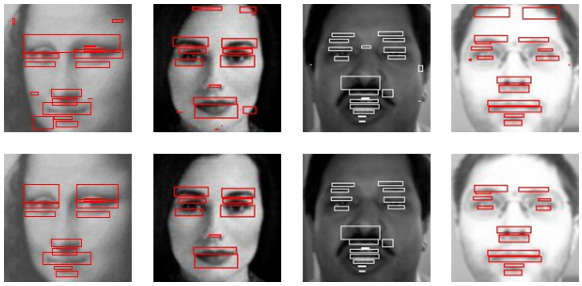

Figure 10: *Les images de la première ligne montrent les boîtes englobantes des CC détectées initialement tandis que celles de la seconde montrent les boîtes des CC conservées.*

trouve au-dessus devient la RAC du nez et l'autre devient la RAC de la bouche. La Figure 11 montre quelques exemples de RAC obtenues en fonction du seuillage de la carte d'énergie.

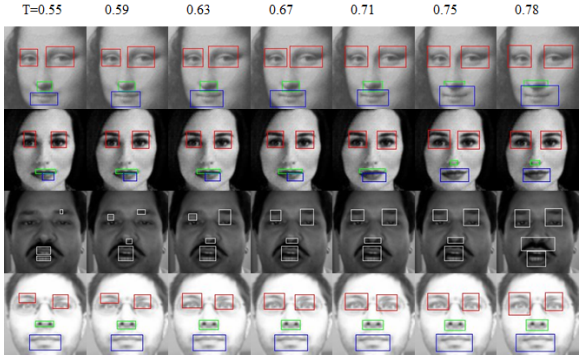

Figure 11: *Détection des RAC en fonction du seuillage de la carte d'énergie.*

3.3. Analyse multi-seuils de la carte d'énergie

Nous avons présenté jusque là comment nous recherchions des régions anatomiques candidates de chacun des yeux, du nez et de la bouche. Toutefois, l'illumination sur un visage donné peut varier, à cause de conditions d'éclairage inégales ou encore d'ombres provenant d'autres objets (cheveux, main...). Un seuillage de la carte d'énergie adéquat pour extraire les yeux n'est pas toujours celui qui permet d'extraire le nez ou la bouche. Ainsi, un seuil doit être choisi pour chaque élément anatomique spécifique du visage. Pour cette tâche, une analyse multi-seuils est proposée. A ce moment précis de notre approche, plusieurs seuils ont été utilisés, chacun permettant d'extraire 4 RAC qui correspondent respectivement aux yeux droit et gauche, au nez et à la bouche. Nous avons généré ainsi 4 ensembles. Le premier contient tous les RAC de l'œil droit, le second, tous les RAC de l'œil gauche et ainsi de suite. Le but de l'analyse multi-seuils est de trouver un seuil adéquat et donc une RAC adéquate pour chaque région spécifique R du visage. Ici, nous supposons que la RAC recherchée soit celle dont la position et la taille varie peu en fonction du seuil t de la carte d'énergie. En effet, si la position et la taille d'une RAC varie peu, cela signifie que l'énergie de la zone définie par la RAC est stable par rapport à t. Ainsi, pour une région donnée R, nous définissons 4 fonctions : $x_R(t)$, fonction des abscisses du point S des RAC, $y_R(t)$, fonction des ordonnées du point S des RAC, $w_R(t)$, fonction des largeurs des RAC et $h_R(t)$, fonction des hauteurs des RAC de la région R. Un seuil adéquat t^* d'une région spécifique R est donné par l'équation (7).

$$\begin{aligned} D(t) &= \left|\frac{\delta}{\delta t}x_R\right| + \left|\frac{\delta}{\delta t}y_R\right| + \left|\frac{\delta}{\delta t}w_R\right| + \left|\frac{\delta}{\delta t}h_R\right| \qquad (7)\\ A(t) &= \beta \cdot W \cdot H - w_R(t) \cdot h_R(t)\\ t^* &= \max_{A(t)>0 \text{ and } D(t)<\varepsilon_R} t \end{aligned}$$

β est le ratio maximum entre l'aire de la région spécifique R et celle de la fenêtre du visage. β dépend de la connaissance que nous avons sur les proportions maximales de chaque région par rapport à l'ensemble du visage. Par exemple, pour un œil, il est de 0,1. Notons que β est une borne supérieure, elle n'est en réalité dans notre méthode que très rarement atteinte, mais nous permet d'exclure quelques valeurs absurdes.

Quant à ε_R, il s'agit de la moyenne des $D(t)$ où les valeurs nulles de $D(t)$ ne sont pas prises en compte. Au final, 4 seuils sont calculés indépendamment et permettent de choisir une RAC pour une région R spécifique comme le montre la Figure 12.

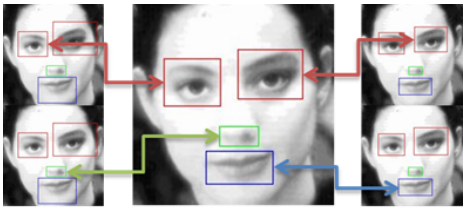

Figure 12: *Sélection des RAC pour chaque région R par l'analyse multi-échelle.*

4. Evaluation

Dans la littérature, rares sont les méthodes qui cherchent à la fois les yeux, le nez et la bouche, ainsi une évaluation générale comparative de l'ensemble est difficile. Par contre, de nombreuses méthodes cherchent à extraire un de ces éléments spécifiques du visage. Pour cette évaluation, nous utilisons les bases BioID et Color FERET. La base d'images BioID est composée de 1520 images. chacune d'entre elles contient un visage. Sur cette base, seuls les yeux, plus précisément l'iris des yeux sont annotés. Les conditions d'illumination ne sont pas contrôlées et varient. De nombreux visages subissent des occlusions (main, lunettes). Cette évaluation se basera essentiellement sur la détection des yeux. En effet, en ce qui concerne la détection des yeux, il est possible de donner non seulement le taux de bonne détection,

mais aussi la précision rattachée à cette détection. Malheureusement, ce n'est ni le cas pour le nez, ni pour la bouche où le seul indicateur sera le taux de bonne détection. Nous méthode est comparée à celles de Li et al. [YLM08] et celle de Asteriadis et al. [SAP09]. Dans [YLM08] et [SAP09], la mesure de Jesorsky [OJF01] est utilisée pour évaluer la précision de la détection des yeux sur la base BioID.

Soient d_r, la distance entre la position réelle du centre de l'iris droit et le centre de la région correspondant à l'œil droit et d_l, l'équivalent pour l'œil gauche. Soit d_{rl} la distance réelles entre les centres de l'iris droit et gauche. Jesorsky definit l'erreur *err* par l'équation (8).

$$err = \frac{max(d_r, d_l)}{d_r l} \tag{8}$$

Pour la tâche qui consiste à détecter les yeux, une erreur de 0,25 est acceptée. En d'autres termes, quand $err < 0,25$, on considère que le détection a réussi. Notons que cet article traite du sujet de la détection des yeux et non celui de la localisation. La différence principale entre détection et localisation est que la première cherche une région alors que la seconde essaie de localiser un point saillant précis. Pour les problèmes de localisation de l'œil, une erreur inférieure à 0,05 ou 0,1 est demandée [XTC09]. Le standard de Jesorsky utilise les positions du centre des iris. Ainsi, nous devons tout d'abord estimer la position de l'iris.

Notre approche qui consiste à détecter la région des yeux n'a pas été exclusivement conçue pour les images de visages frontales. Au fur et à mesure que le visage tourne, seul un des deux yeux devient visible. Notre approche est donc conçue pour détecter au moins un œil. Par contre, les visages dans BioID sont frontales. Ainsi, les deux yeux sont visibles. Nous avons donc calculé le taux d'images dans BioID où un seul œil est détecté. Dans 0,0034% des visages de BioID, notre méthode détecte un œil alors que les deux sont présents et comme le standard de Jesorsky requiert les deux yeux, notre évaluation se portera donc sur 99,9966% des visages.

Dans notre approche, les lignes horizontales sont extraites. Puisque les sourcils ou encore l'arcade sourcilière sont à direction horizontale, ils sont systématiquement inclus dans la région des yeux. Puisque le but est d'extraire les boîtes englobantes des régions anatomiques du visage, nous supposons que le sourcil peut être être pertinent et ainsi peut être incorporé dans la région des yeux. Cependant, comme le centre de l'œil dans BioID est l'iris et comme notre approche inclut les sourcils ou l'arcade sourcilière, nous avons choisi de réduire les RAC des yeux du tiers de leur hauteur. Ainsi, l'abscisse et la largeur des RAC ne changent pas, contrairement à l'ordonnée et la hauteur. Soient y_E et h_E l'ordonnée et la hauteur de la région détectée de l'œil, nous définissons la nouvelle ordonnée y_{eye} et la nouvelle hauteur h_{eye} par l'équation 9.

$$\begin{aligned} y_{eye} &= y_E - \frac{1}{3} \cdot h_E \\ h_{eye} &= \frac{2}{3} \cdot h_E \end{aligned} \tag{9}$$

Alors, les coordonnées du centre *C* de l'œil sont estimées par l'équation 10.

$$\begin{aligned} x_C &= x_E + \frac{1}{2} \cdot w_E \\ y_C &= y_{eye} + \frac{1}{2} \cdot h_{eye} \end{aligned} \tag{10}$$

Méthode	Détection (%)	Erreur moyenne
Li et al.	96	0,1004
Asteridias et al.	96	non indiqué
Notre méthode	97,23	0.1130

Table 1: *Comparaison entre les méthodes de Li et al. Asteriadis et al. et de la nôtre sur BioID.*

Le tableau 1 compare notre méthode avec celles de Li et al. et de Asteriadis et al. Le taux de détection correspond au pourcentage de visages où $err < 0,25$. Notre méthode possède un meilleur taux de détection pour une précision moindre par rapport à Li et al. Malgré l'approximation sur le centre des yeux que nous avons utilisée, les résultats sur BioID sont similaires.

Seuil	Détection (%)	Erreur moyenne
0.39	58.98	0.3872
0.43	62.47	0.3495
0.47	67.95	0.3081
0.47	67.95	0.3081
0.51	72.90	0.2768
0.55	77.37	0.2540
0.59	82.79	0.2127
0.63	88.48	0.1773
0.67	91.67	0.1549
0.71	93.90	0.1417
0.74	**95.39**	**0.1309**
0.78	93.43	0.1434
0.82	87.80	0.1792
0.86	72.22	0.2786
0.90	45.60	0.4971
multi-threshold	**97.23**	**0.1130**

Table 2: *Taux de détection et erreur moyenne relative à la détection des yeux en fonction du seuil appliqué sur la carte d'énergie.*

Notre méthode est basée sur la sélection de candidats adéquats parmi les RAC d'une région saillante du visage. La première question que l'on peut se poser et comment serait la détection sans l'analyse multi-seuils pour évaluer l'intérêt de celle-ci. C'est pourquoi le tableau 2 donne le taux de détection et l'erreur moyenne de la détection des yeux en fonction du seuil utilisé sur la carte d'énergie horizontale.

Comme nous pouvons le voir sur le tableau 2, le pourcentage de bonne détection ($err < 0,25$) n'est pas distribué de manière égale. On observe un maximum du pourcentage de détection et une erreur minimale pour un seuil proche de

0,74. C'est en partie dû à la normalisation de la carte d'énergie horizontale. Cette normalisation a aussi pour effet de réduire l'espace de recherche. Notons que pour un seuil supérieur à 0,86, le taux de détection décroît et l'erreur augmente rapidement. En effet, lorsque le seuil de la carte d'énergie est trop élevé, les CC issues de différents éléments saillants du visage ont tendance à fusionner. Ce phénomène est brutal, puisqu'une toute petite augmentation de ce seuil peut fusionner deux RAC auparavant distincts. Sans l'analyse multi-seuils, pour un seuil fixé à 0,74, la détection de l'œil réussit avec toutefois de bons rappel et précision. Néanmoins, clairement, l'approche multi-seuils montre de meilleurs résultats à la fois en terme de taux de détection, mais aussi de précision que l'utilisation d'un seuil fixe.

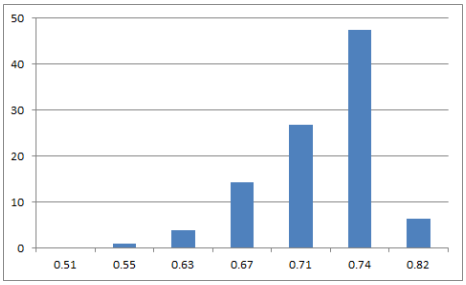

Figure 13: *Pourcentage des seuils choisis par l'analyse multi-seuils.*

La Figure 13 montre le pourcentage des sept seuils choisis par l'analyse multi-seuils pour la détection des yeux. On remarque tout d'abord que le seuil le plus sélectionné est celui qui correspond à la meilleure détection en terme de taux de détection et d'erreur moyenne. Cependant, seul 47% des RAC choisis des yeux sont issues de ce seuil, bien que le taux de bonne détection à ce seuil soit de 95,39%. L'analyse multi-échelle permet de choisir d'autres seuils permettant d'obtenir des RAC de l'œil plus précise pour la détection des yeux. Ceci confirme l'hypothèse que nous avons formulée, à savoir que les RAC adéquats sont celles qui présentent une certaine stabilité de position et de taille malgré la variation du seuil. Enfin, la Figure 13 montre aussi que l'analyse multi-seuils préfère sélectionner les seuils inférieurs à 0,74 plutôt que ceux qui lui sont supérieurs. Comme nous l'avons dit plus tôt, lorsque le seuil atteint une certaine valeur, les CC des RAC sélectionnées ont tendance à fusionner entre elles. Ceci a pour conséquence d'introduire une discontinuité dans la variation de la position et de la taille des éléments détectés.

Nous avons aussi évalué notre méthode sur la base Color FERET. Chaque image contient un visage dans différentes conditions de pose, d'illumination... De nombreuses personnes ont une barbe, une moustache ou encore des lunettes. Les positions de l'iris, du bout du nez, du centre de la bouche sont données sur la majorité des visages frontaux et sur quelques images de visages non frontaux.

Nous avons aussi évalué notre méthode sur la base LFW. Les visages sont au centre de l'image et ont toutes été détectées par la méthode de Viola et Jones. Cette base est particulièrement diversifiée en termes de conditions de prise de vue, d'illumination, de qualité (flou). Il s'agit de l'une des bases les plus difficiles et exigeantes qui existent actuellement. Malheureusement, les positions des différents éléments anatomiques n'y sont pas annotées. La base LFW contient un nombre très important d'images. Il nous était impossible de les annoter toutes. Nous avons choisi d'annoter la position des iris de toutes les images dont le prénom de la personne centrale à l'image commence par un "C" (plus de 900 images).

Base (%)	Détection(%)	erreur moyenne
BioID	97,23	0,1130
Color Feret	97.60	0,1110
LFW	93.74	0,1107

Table 3: *Taux de détection et erreur moyenne sur différentes bases de visages*

Comme le montre le tableau 3, notre méthode parvient à détecter les yeux avec une erreur moyenne semblable pour les trois bases. On remarque que le taux de détection des yeux est plus faible sur la base LFW, ce qui confirme la difficulté de cette base. Les figures 14,15 et 16 montre respectivement des exemples de résultats visuels sur les base BioID, Color Feret et LFW.

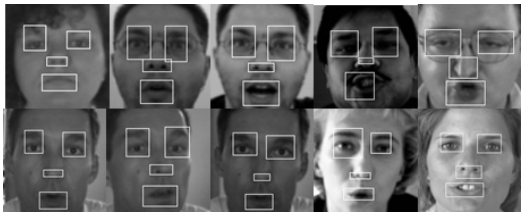

Figure 14: *Résultats visuels sur BioID.*

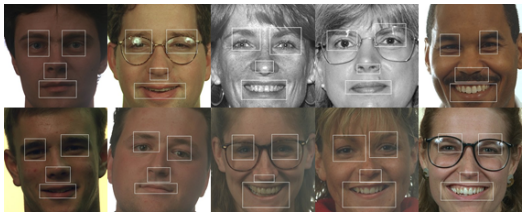

Figure 15: *Résultats visuels sur Color FERET.*

La figure 17 montre des exemples de détections erronées ou incomplètes.

En ce qui concerne la détection du nez et de la bouche, seule le taux de détection a été testé. Notre approche est conçue pour donner au moins une RAC du nez et de la bouche, puisque au moins une partie de ces éléments est supposée visible. Le tableau 4 donne le résultat du taux de détection du nez et de la bouche dans la base Color FERET et MIT/CMU. Ces bases ont l'avantage d'avoir les positions du bout du nez et de la bouche annotées. La détection du nez et celle de la bouche n'ont pas été évaluées sur BioID, car ces éléments ne sont pas annotés dans cette base. Pour ces tests, nous supposons qu'une région est correctement détectée si

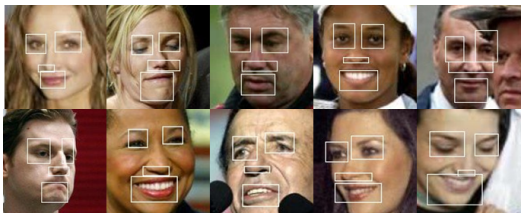

Figure 16: *Résultats visuels sur LFW.*

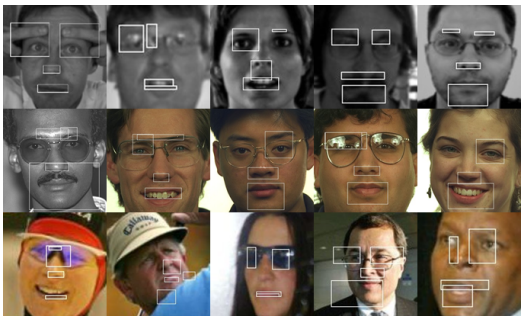

Figure 17: *Exemples de détections erronées ou incomplètes.*

elle contient le point annoté correspondant tout en respectant une contrainte sur l'aire de la région. Pour le nez, l'aire de la région détectée doit être inférieure à 5% de l'aire de la fenêtre du visage et pour la bouche, elle doit être inférieure à 8%. Le tableau 4 montre le taux de détection du nez et de la bouche. Globalement, la bouche semble correctement détectée. Le taux de détection du nez est nettement inférieur, d'une part parce que la détection du nez est difficile lorsque le sujet porte la moustache et d'autre part parce que notre méthode a tendance à détecter la base du nez et non le bout du nez.

Base (%)	Nez(%)	Bouche(%)
Color Feret	75,23	97,50
MIT/CMU	83,63	97,76

Table 4: *Taux de détection du nez et de la bouche dans les bases Color FERET et MIT/CMU.*

Sur la base BioID, Le temps moyen de calcul de la méthode pour un visage est de 16 ms sur un processeur Intel Core i7-2670 QM cadencé à 2,2 GHz. L'extraction des RAC à un seuil donné se fait en moyenne en 2 ms. Ces mesures ont été prises sur un logiciel de test qui n'est pas optimisé. De plus, toute la partie d'extraction des RAC en fonction des seuils qui est la plus longue, car implémentée de manière séquentielle peut être incorporée dans une architecture à processus parallèle. Sinon, le calcul de la carte horizontal d'énergie est relativement rapide, puisque l'image intégrale proposée par Viola et Jones est utilisée. Enfin, la partie concernant l'analyse multi-seuils est la plus rapide dans cette approche. Nous utilisons ici 7 seuils. Pour chaque seuil, 4 régions sont extraites. chaque région génère 4 valeurs (position et taille). Ainsi, l'extraction des 4 seuils adéquats demande un nombre de calculs constants sur un tableau de 112 valeurs.

5. Conclusion

Dans cet article, nous proposons une nouvelle méthode qui utilise un seul type de motif de filtre de Haar à taille adaptative permettant d'extraire les boîtes englobantes des yeux, du nez et de la bouche. Connaissant le niveau d'observation du visage, nous sommes capables de détecter les différentes parties du visage grâce à une carte d'énergie horizontale qui s'avère efficace. Cet article montre aussi comment de simples connaissances sur les visages permet d'améliorer ou de consolider la détection des régions anatomiques faciales. De plus, nous proposons une méthode d'analyse multi-seuils capable de choisir un seuil adéquat pour chaque élément du visage, malgré des conditions d'illumination difficiles. L'évaluation a montré l'efficacité de l'analyse multi-seuil. Elle a aussi montré que la méthode permet de détecter les yeux avec précision, malgré l'approximation relative au sourcil ou à l'arcade sourcilière que nous avons utilisée. La détection de la bouche est aussi bonne tandis que celle du nez reste toujours difficile, en particulier, lorsque le sujet a une moustache et une barbe.

Références

[AB10] AKHLOUFI M., BENDADA A. : Locally adaptive texture features for multispectral face recognition. *Systems, Man and Cybernetics* (octobre 2010), 3308–3314.

[AYH89] A. YUILLE D. C., HALLINAN P. : Feature extraction from faces using deformable templates. *CVPR* (juin 1989), 104–109.

[DPM11] D. PETRISOR C. FOSALAU M. A., MARIUT F. : Algorithm for face and eye detection using colour segmentation and invariant features. *TSP* (2011), 564–569.

[DZZ14] DI ZHU SIYU XIA X. Z., ZHENG J. : Hybrid method for human eye detection. *CCDC* (2014), 5368–5373.

[GS07] GIZATDINOVA Y., SURAKKA V. : Automatic detection of facial landmarks from au-coded expressive facial images. *ICIAP* (septembre 2007), 419–424.

[IC13] INHO CHOI D. K. : Generalized binary pattern for eye detection. *Signal Processing Letters* (2013), 343–346.

[JH09] JIAN W., HONGLIAN Z. : Eye detection based on mult-angle template matching. *Image Analysis and Signal Processing* (avril 2009), 241–244.

[JP09] JAIN A., PARK U. : Facial marks : Soft biometric for face recognition. *ICIP* (novembre 2009), 37–40.

[KP97] KOTROPOULOS C., PITAS I. : Rule-based face detection in frontal views. *Acoustics, Speech and Signal Processing. Vol. 4* (avril 1997), 2537–2540.

[LR12] LAXMI V., RAO P. : Eye detection using gabor filter and svm. *ISDA* (2012), 880–883.

[LZM12] LIN ZHONG QINGSHAN LIU P. Y., METAXAS D. : Leanrning active facial patches for expression analysis. *CVPR* (juin 2012), 2562–2569.

[MCT09] MURPHY-CHUTORIAN E., TRIVEDI M. : Head pose estimation in computer vision : A survey. *Pattern Analysis and Machine Intelligence. Vol. 31*, Num. 14 (avril 2009), 607–626.

[MZW10] MINGCAI ZHOU LIN LIANG J. S., WANG Y. : Aam based face tracking with temporal matching and face segmentation. *CVPR* (novembre 2010), 701–708.

[OJF01] O. JESORSKY K. J. K., FRISHOLZ R. W. : Robust face detection using the hausdorff distance. in : Audio and video based person authentication. *LNCS* (2001), 90–95.

[qZhC12] QING ZHU J., HUI CAI C. : Real-time face detection using gentle adaboost algorithm and nesting cascade structure. *ISPACS* (2012), 33–37.

[SAP09] S. ASTERIADIS N. N., PITAS I. : Facial feature detection using distance vector fields. *Pattern Recognition. Vol. 42*, Num. 7 (2009), 1388–1398.

[TCT01] T. COOTES G. E., TAYLOR C. : Active appearance models. *Pattern Analysis and Machine Intelligence. Vol. 23*, Num. 6 (juin 2001), 681–685.

[TOM02] T. OJALA M. P., MAENPAA T. : Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence* (juillet 2002), 971–987.

[TT10] TAN X., TRIGGS B. : Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Biometrics Compendium. Vol. 19*, Num. 6 (juin 2010), 1635–1650.

[VJ01] VIOLA P., JONES M. : Rapid object detection using a boosted cascade of simple features. *CVPR. Vol. 1* (juin 2001), 511–518.

[VP05] VUKAINOVIC D., PANTIC M. : Fully automatic facial feature point detection using gabor feature based boosted classifiers. *Systems, Man and Cybernetics. Vol. 2* (octobre 2005), 1692–1698.

[XTC09] XIAOYANG TAN FENGYI SONG Z.-H. Z., CHEN S. : Enhanced pictorial structures for precise eye localization under incontrolled conditions. *CVPR* (2009), 1621–1628.

[YLM08] YI LI PENG-FEI ZHAO B.-K. W., MING D. : An improved hybrid projection function for eye precision location. *MIMI. Vol. 4987* (2008), 312–321.

[ZZ12] ZHU S., ZHANG N. : Face detection based on skin color model and geometry features. *ICICEE* (2012), 991–994.