Détection par corrélation croisée

3.1 Introduction

La corrélation croisée est une mesure de similarité qui a de nombreux avantages. Cette méthode est facile à implémenter, aisément adaptable à une grande variété de formes et ne requiert pas d'extraction de descripteurs complexes ou une large base d'apprentissage. De plus, la corrélation croisée est l'opération linéaire optimale d'un point de vue rapport signal sur bruit pour détecter la position spatiale ou temporelle d'un signal connu dans un bruit blanc stationnaire [96]. Le maximum du signal résultant correspondant alors à la position la plus probable du signal recherché. L'énergie du bruit dans une image (en détection, le bruit correspond à tout ce qui n'est pas l'objet à détecter) n'est pas stationnaire, ce qui rend la simple corrélation croisée peu performante pour effectuer une détection d'objets. De plus, cette mesure de similarité n'est pas très bien adaptée à la détection d'objets complexes tels que des visages. En effet, la corrélation n'est que peu robuste aux variations d'illumination, d'échelle ou de rotation qui ne peuvent être considérées comme un simple bruit blanc. La corrélation croisée normée est une mesure de similarité basée sur la corrélation mais qui permet grâce à une normalisation des signaux, de rendre plus robuste la détection aux variations d'énergie et de luminosité de l'image. Excepté pour des formes simples [68], la corrélation croisée normée n'a été que peu employée pour la détection d'objets car elle ne permet pas de tenir compte des variations de forme, de couleur, de prise de vue, ou d'échelle. Dans ce chapitre, nous proposons d'associer la corrélation normée croisée à la méthode des plus proches voisins afin de pouvoir représenter les différentes formes que peut prendre l'objet à détecter. Ainsi, un objet sera détecté à une position et échelle donnée si la mesure de similarité maximum entre l'image test et l'ensemble des images exemples est supérieure à un seuil donné. Plus la base d'exemples est grande, mieux cet objet sera modélisé. Cependant, les temps de calculs sont directement proportionnels au nombre d'images exemples disponibles, ce qui limite la taille de la base d'exemples. Nous effectuerons nos expérimentations sur la détection de visages qui bénéficie d'une littérature abondante, les visages étant considérés comme

l'objet complexe par excellence (un visage peut revêtir différentes formes, différentes couleurs, de nombreuses expressions différentes, ainsi que des éclairages très divers). Nous commencerons par décrire les principes de la corrélation croisée ainsi que la corrélation croisée normée et la corrélation croisée normée centrée. Nous étudierons ensuite la corrélation directement appliquée aux images en Niveaux de Gris afin de déterminer l'influence des différentes variations de forme, de position et d'échelle. Ensuite, nous étudierons la corrélation sur des images préalablement traitées par la méthode de Sobel afin d'extraire les contours des images et de diminuer la sensibilité de la mesure de similarité aux variations de luminosité. Finalement, nous introduirons une méthode dérivée de la PCA permettant d'extraire les formes revenant le plus souvent dans une base d'images exemples et de calculer ainsi des filtres adaptés à l'objet que nous souhaitons détecter.

3.2 Principes de la corrélation

La corrélation croisée est aussi connue en statistique pour désigner la covariance de vecteurs aléatoires $\mathbf{x}$ et $\mathbf{y}$. En traitement du signal, la corrélation permet de mesurer la similarité entre deux signaux $x(\mathbf{t})$ et $y(\mathbf{t})$. En traitement d'image, $\mathbf{t}$ est un vecteur à deux dimensions représentant les coordonnées (i, j) des pixels des images. La fonction $s(\boldsymbol{\tau})$ résultante s'écrit :

$$s(\boldsymbol{\tau}) = x(\mathbf{t}) * y(-\mathbf{t}) = \int x(\mathbf{t})\, y(\mathbf{t}+\boldsymbol{\tau})\, d\mathbf{t} \tag{3.1}$$

En pratique, si $x(i,j)$ représente l'image exemple de dimension $h_x \times l_x$ et $y(i,j)$ représente l'image test de dimension $h_y \times l_y$ dans laquelle nous souhaitons effectuer la détection. Alors l'image résultante $s(u,v)$ de dimension $(h_s = h_y - h_x) \times (l_s = l_y - l_x)$ s'ecrie :

$$s(u,v) = \sum_{i=j=0}^{i<h_x, j<l_x} y(u+i, v+j)\, x(i,j) \tag{3.2}$$

La corrélation croisée normée est basée sur la norme $L2$ des images et se calcule ainsi :

$$\sigma_x = \left(\sum_{i=j=0}^{i<h_x, j<l_x} x(i,j)^2 \right)^{\frac{1}{2}} \qquad \sigma_y(u,v) = \left(\sum_{i=j=0}^{i<h_x, j<l_x} y(u+i, v+j)^2 \right)^{\frac{1}{2}} \tag{3.3}$$

$$s_n(u,v) = \frac{1}{\sigma_x \sigma_y(u,v)} \left(\sum_{i=j=0}^{i<h_x, j<l_x} y(u+i, v+j)\, x(i,j) \right) \tag{3.4}$$

Enfin, la corrélation croisée normée centrée correspond à la corrélation croisée normée des images auxquelles on soustrait leur valeur moyenne :

$$m_x = \frac{1}{h_x l_x} \sum_{i=j=0}^{i<h_x, j<l_x} x(i,j) \qquad m_y(u,v) = \frac{1}{h_x l_x} \sum_{i=j=0}^{i<h_x, j<l_x} y(u+i, v+j) \tag{3.5}$$

$$\sigma_x = \left(\sum_{i=j=0}^{i<h_x, j<l_x} [x(i,j) - m_x]^2 \right)^{\frac{1}{2}} \qquad \sigma_y(u,v) = \left(\sum_{i=j=0}^{i<h_x, j<l_x} [y(u+i, v+j) - m_y(u,v)]^2 \right)^{\frac{1}{2}}$$

$$s_{nc}(u,v) = \frac{1}{\sigma_x \sigma_y(u,v)} \left(\sum_{i=j=0}^{i<h_x, j<l_x} [y(u+i, v+j) - m_y(u,v)] [x(i,j) - m_x] \right) \tag{3.6}$$

3.3 Détection par Niveaux de Gris

Dans cette section, nous commençons par mettre en œuvre un système de détection basé sur la corrélation des images en Niveaux de Gris, associé à la méthode des plus proches voisins. Nous utiliserons pour tester ce système la base de données 'Face 1999 (Front)' (Annexe : A.1). Cette base comporte 450 visages de 27 personnes distinctes. Elle est normalement destinée à la reconnaissance de visages plutôt qu'à la détection ; elle est cependant bien adaptée pour comparer les résultats d'un système de détection aussi basique qu'une simple corrélation. Nous utiliserons une base d'exemples composée de 80 visages (figure : 3.1).

Chaque visage exemple est redimensionné à la dimension minimum des visages que nous souhaitons détecter. Les images de test sont successivement sous-échantillonnées avec un facteur 1.2. Nous formons ainsi une pyramide d'images permettant d'effectuer une détection multi-échelle. Si un visage est détecté sur la première image de la pyramide (l'image non sous-échantillonnée), alors la dimension du visage est la même que celle de l'image de référence. Dans le cas général, si un visage est détecté dans une image de la pyramide sous échantillonnée d'un facteur α, alors la dimension du visage dans l'image de test sera celle de l'image de référence multipliée par α. Chaque visage de la base de référence est corrélé avec l'ensemble des images de la pyramide résultant ainsi en un ensemble de cartes de scores de corrélation. Un visage est détecté s'il correspond à un maximum local sur la carte de score et s'il n'est pas superposé à un autre visage avec un score de détection supérieur. Nous considérerons que deux visages sont superposés si l'aire $A1$ formée par l'intersection des deux régions correspondantes divisée par l'aire $A2$ de la superposition des deux régions est supérieure à 0.1. De même, nous considérerons que nous avons effectué une bonne détection si pour les deux régions correspondant respectivement au visage à détecter et à la détection effectuée par le système, $A1/A2$ est supérieur 0.5 (figure : 3.2).

Les résultats de la détection sont influencés par différents paramètres. Le premier d'entre eux est l'utilisation comme mesure de similarité, de la corrélation simple,

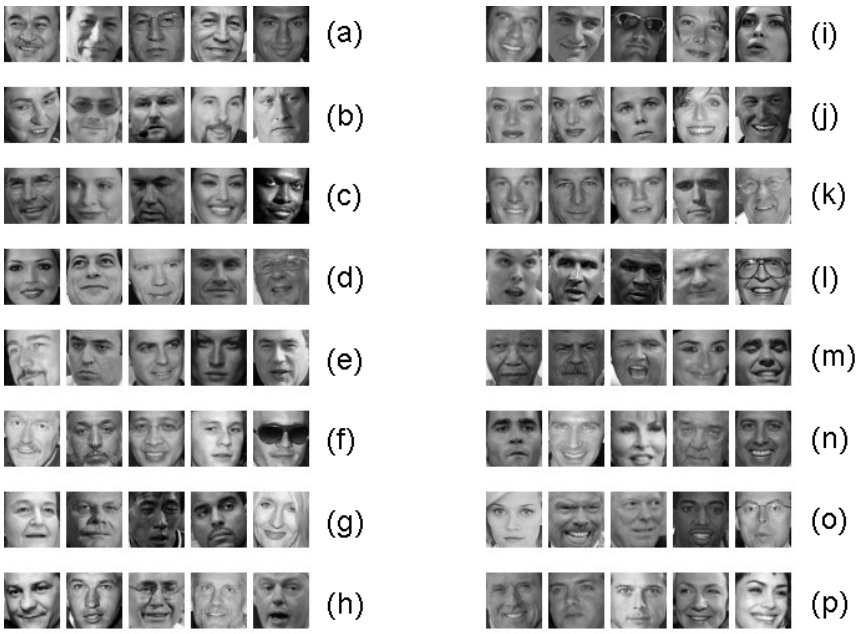


FIGURE 3.1 – Base d'images de référence pour le système de détection par corrélation. Elle est divisée en 16 bases de données de 5 visages respectivement nommées (a) à (p).

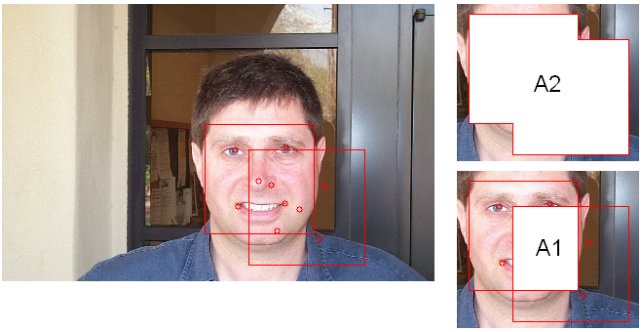


FIGURE 3.2 – Exemple de deux régions superposées. L'aire $A1$ correspond à l'intersection des deux régions, l'aire $A2$ correspond à l'union des deux régions. Si $\frac{A1}{A2} > 0.1$ alors on considère que les deux régions sont superposées. Si $\frac{A1}{A2} > 0.5$ et que une des deux régions correspond à l'objet à détecter manuellement annoté, alors nous effectuons une bonne détection.

la corrélation normée ou bien la corrélation normée centrée. Mais aussi le nombre d'exemples de la base de référence ou la dimension des images exemples. Nous commencerons par nous attacher à déterminer l'influence de la mesure de similarité utilisée.

3.3.1 Influence de la normalisation et du centrage sur la corrélation

La corrélation simple est peu utilisée en détection dans des images, on lui préfère souvent la corrélation normée, ou normée centrée [68]. Nous nous proposons dans cette section d'étudier l'influence de la normalisation et du centrage sur un problème de détection de visages. Comme nous travaillons avec des bases d'exemples de taille réduite, nous utiliserons, afin de tenir compte de l'influence de la base utilisée, quatre bases de données de référence de cinq visages (figure : 3.1) respectivement nommées base (a), (b),(c) et (d). Les images de références utilisées pour l'expérience sont de dimension hauteur fois largeur égal $44 \times 44 = 1936$ pixels. Nous présentons les résultats sous forme de courbes Rappel Précision. Ces courbes rapportent la Précision de la détection (nombre de bonnes détections divisé par le nombre total de détections) par rapport au Rappel (nombre de bonnes détections divisé par le nombre total d'objets à détecter). Ainsi, plus une courbe se rapproche du coin supérieur droit, meilleur est le système de détection.

Nous pouvons constater que la normalisation est indispensable au système de détection. En effet, la courbe correspondant à la corrélation croisée n'est visible sur aucune des quatre figures car le système n'a pu effectuer aucune bonne détection. Ensuite, on remarque que ce système est très dépendant de la base de référence utilisée. Les résultats montrés par les figures 3.3a et 3.3d correspondant aux bases d'exemples (a) et (d) sont très nettement supérieurs aux résultats obtenus avec les bases de référence (b) et (c) (figure : 3.3b,3.3c). Il est par contre difficile ici de départager la corrélation croisée normée de la corrélation croisée normée centrée, c'est pourquoi nous continuerons par la suite à utiliser ces deux mesures.

Enfin, même si un tel système est capable d'effectuer quelques bonnes détections en utilisant la corrélation croisée normée et une base de référence adaptée, les taux de détection restent insuffisants pour une application pratique.

3.3.2 Influence du nombre d'échantillons exemples

Nous avons vu dans la section précédente que les résultats de la détection sont très dépendants de la base d'exemples utilisée. Les mauvais résultats des bases d'exemples (b) et (c) peuvent s'expliquer de deux manières. Soit les exemples de référence sont trop différents des visages de la base de test, soit certains visages de la base exemple peuvent avoir une très forte corrélation avec des images n'étant pas des visages. L'augmentation du nombre d'exemples de la base de référence maximise les chances d'avoir une forte corrélation entre un visage exemple et un visage de la base de test mais augmente aussi le score de corrélation des fausses détections. Afin de déterminer la sensibilité de notre système de détection au nombre d'exemples de la base de référence, nous avons appliqué notre système pour une base de données de cinq, dix, vingt, quarante puis quatre-vingt exemples avec la corrélation normée et la corrélation normée centrée (figure : 3.4).

La base de cinq visages que nous avons utilisée est celle qui a donné les meilleurs résultats de détection, à la fois avec la corrélation normée et avec la corrélation

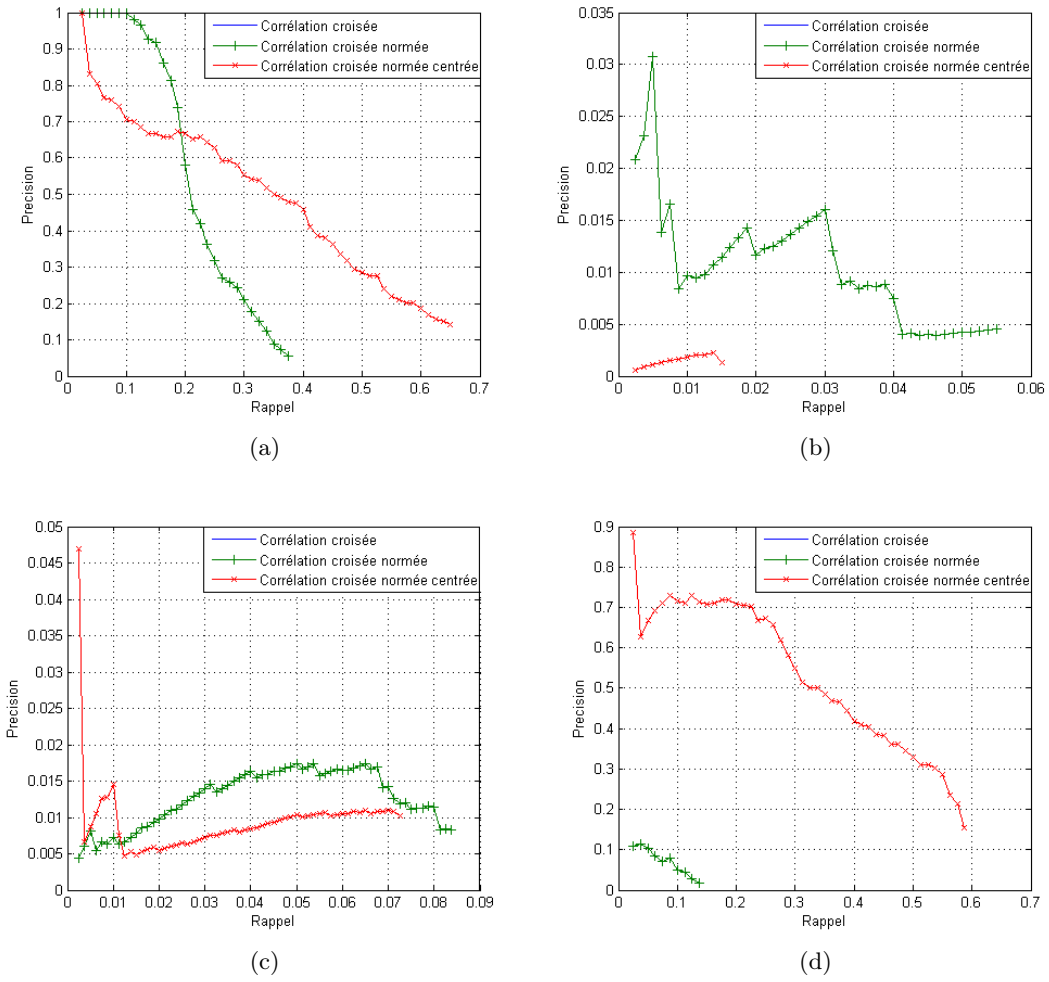
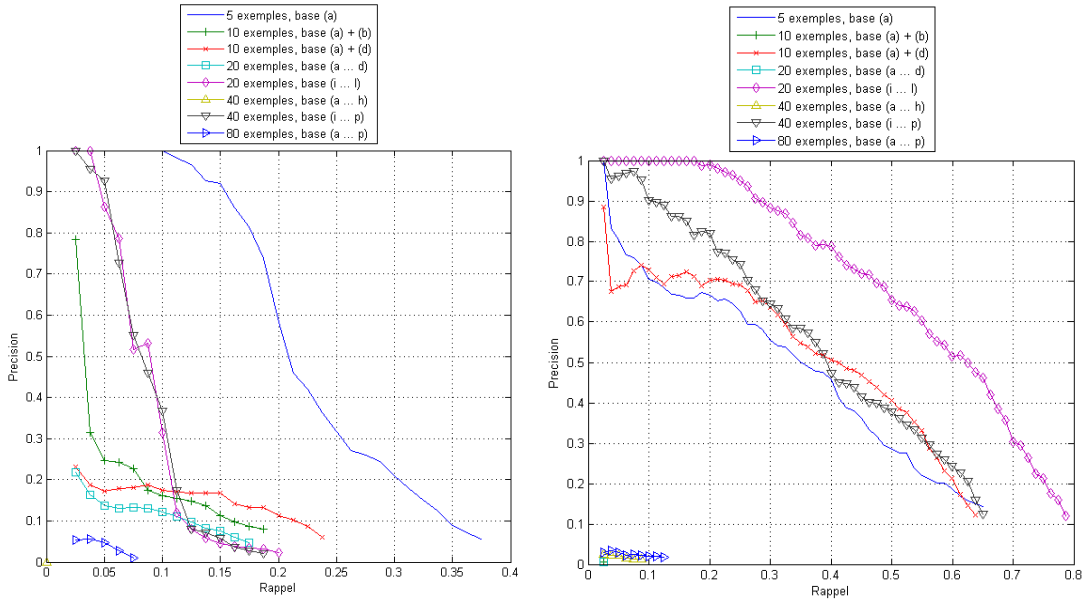


FIGURE 3.3 – Courbes Rappel Précision du système de détection par corrélation croisée des images en Niveaux de Gris. Les figures (a) (b) (c) et (d) ont été obtenues avec quatre bases de référence distinctes de cinq visages. On remarque donc que les résultats sont très dépendants de la base utilisée. La courbe correspondant à la corrélation croisée simple n'est pas visible car un tel système n'a pu effectuer aucune bonne détection.

normée centrée, *i.e*, la base (a). Nous avons ensuite utilisé une base de dix visages constituée des bases d'exemples (a) et (b), puis une seconde base de dix visages constituée des bases (a) et (d). On remarque ainsi que la combinaison de deux bases d'exemples donnant des résultats équivalents (a) et (d) entraîne une amélioration du taux de détection. Cependant, combiner deux bases d'exemples (a) et (b), l'une donnant des résultats nettement supérieur à l'autre, entraîne des taux de détection proches de ceux obtenus avec la base d'exemples la moins adaptée. Ainsi, bien que l'augmentation du nombre d'exemples a tendance à entraîner une amélioration des



(a) Corrélation croisée normée (b) Corrélation croisée normée centrée

FIGURE 3.4 – Courbes Rappel Précision du système de détection par corrélation croisée des images en Niveaux de Gris pour différents nombres d'images de référence.

résultats (les meilleurs résultats étant obtenus pour des bases de 20 et 40 images exemples), les taux de détection sont toujours très dépendants de la base de référence. Ainsi, les taux de détection obtenus avec la base de référence de 80 images sont parmi les plus faibles. Enfin, on peut remarquer que la corrélation croisée normée centrée obtient de meilleurs résultats que la corrélation croisée normée (figure : 3.5).

3.3.3 Influence de la dimension des images exemples

Nous avons vu que la corrélation croisée normée des images en Niveaux de Gris permet, dans une certaine mesure, d'obtenir un détecteur de visages capable de fonctionner avec une base d'exemples très réduite. Cependant, les résultats sont très liés à la base d'exemples utilisée et l'augmentation du nombre d'images exemples n'entraîne pas systématiquement une amélioration du taux de détection et peut même entraîner une nette baisse. Jusqu'à présent, nous avons toujours utilisé des images exemples d'une aire de 1936 Pixels. Nous proposons dans cette section de déterminer l'influence de la dimension des images exemples et en particulier, de vérifier s'il est possible de rendre le système moins sensible à la base de référence utilisée en augmentant la dimension des images de référence. Il faut noter que la dimension des images exemples correspond à la dimension minimale des visages détectables. Dans notre base de données d'images test, les visages sont toujours de dimension supérieure à la dimension des visages de références. La figure 3.6b compare les résultats obtenus pour la base (a) de cinq visages avec des images de référence d'une aire de 484, 961, 1936, 2916 et

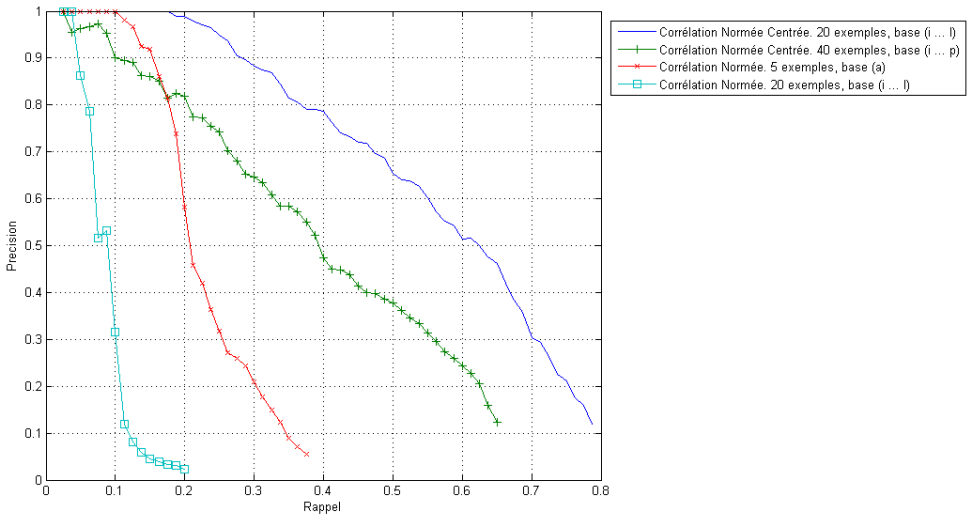
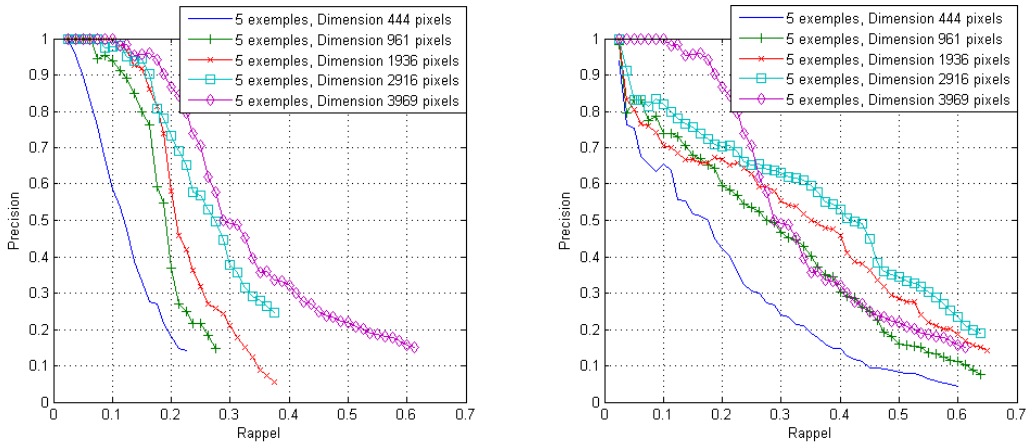


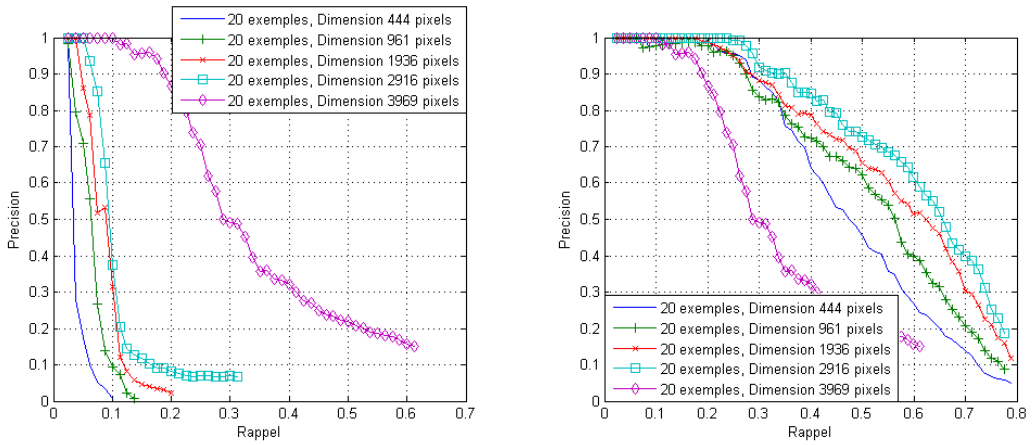
FIGURE 3.5 – Comparaisons des meilleurs résultats pour le système de détection par corrélation croisée normée et croisée normée centrée des images en Niveaux de Gris. La corrélation croisée normée centrée obtient des taux de détection nettement supérieurs à la corrélation normée.

3969 pixels. On remarque une amélioration sensible des résultats avec l'augmentation de la dimension des images de référence, aussi bien pour la corrélation croisée normée centrée que pour la corrélation croisée normée. Le même constat peut être fait pour la base de 20 visages (i ... l) (figure : 3.7b). Nous avons ensuite effectué (figure : 3.8b), les mêmes expérimentations pour le base de 10 visages formée de l'association des bases (a) et (b) qui donnaient des taux de détection très faibles. L'augmentation de la dimension des images exemples entraîne une amélioration des résultats, sans pour autant permettre que cette base d'exemples obtienne des résultats comparables aux deux précédentes.



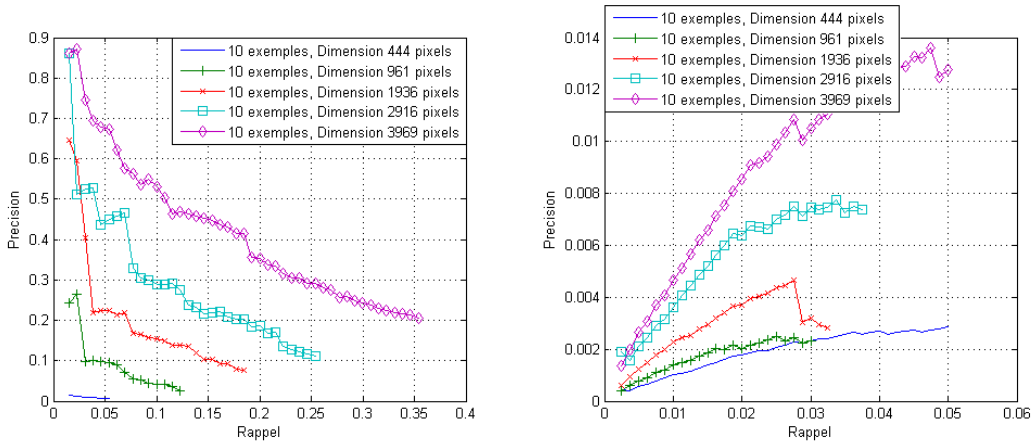
(a) Corrélation croisée normée (b) Corrélation croisée normée centrée

FIGURE 3.6 – Courbes Rappel Précision du système de détection par corrélation croisée des images en Niveaux de Gris avec 5 images exemples. Influence de la dimension des exemples.



(a) Corrélation croisée normée (b) Corrélation croisée normée centrée

FIGURE 3.7 – Courbes Rappel Précision du système de détection par corrélation croisée des images en Niveaux de Gris avec 20 images exemples. Influence de la dimension des exemples.



(a) Corrélation croisée normée (b) Corrélation croisée normée centrée

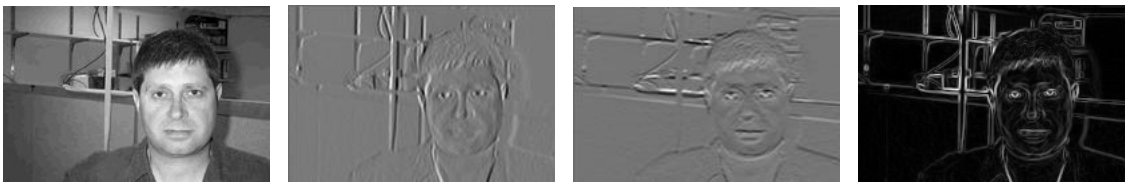
FIGURE 3.8 – Courbes Rappel Précision du système de détection par corrélation croisée des images en Niveaux de Gris avec 10 images exemples. Influence de la dimension des exemples.

3.4 Utilisation de filtres détecteurs de contours

Nous avons vu dans la section précédente que la mesure de similarité basée sur la corrélation des images en Niveaux de Gris permet d'obtenir, sous certaines conditions, des résultats intéressants compte tenue du très faible nombre d'images exemples sur des objets aussi complexes que des visages. Ces résultats sont cependant très dépendants de la base de référence utilisée et ceci indépendamment du nombre ou de la dimension des images exemples. Ainsi, la corrélation croisée des images en Niveaux de Gris semble difficilement utilisable en pratique. Dans cette section, nous utilisons la corrélation croisée sur les contours des images plutôt que sur les images en Niveaux de Gris. Le but est de diminuer la sensibilité de la mesure de similarité aux variations de luminosité et ainsi améliorer les résultats de notre système de détection tout en minimisant l'influence des bases d'exemples utilisées. Dans un premiers temps, nous utiliserons l'algorithme de Sobel pour détecter les contours.

3.4.1 Algorithme de Sobel

L'algorithme de Sobel [113] est connu pour être l'une des méthodes les plus simples et efficaces pour effectuer la détection de contours. Cette méthode consiste à extraire en chaque point d'une image donnée une approximation de la norme du gradient de l'image I afin de faire ressortir les fortes variations de Niveaux de Gris de l'image. Cette méthode consiste à combiner deux filtres (F_x) et (F_y) généralement de dimension 3×3 permettant d'approximer la dérivée horizontale (G_x) et verticale (G_y) de l'image. L'image finale (G) est la norme $(\sqrt{G_x^2 + G_y^2})$ du gradient de l'image traitée. (figure : 3.9).



(a) Image Originale (I) (b) G_x (c) G_y (d) G

FIGURE 3.9 – Exemple d'utilisation de l'algorithme détecteur de contours de Sobel. L'image (a) est l'image originale en Niveaux de Gris, l'image (b) est une approximation du gradient horizontal de l'image, l'image (c) est une approximation du gradient vertical de l'image et enfin l'image (d) représente une approximation de la norme du gradient de l'image et permet d'en faire ressortir les contours.

Les filtres F_x et F_y combinent un lissage gaussien et un opérateur différentiel afin de minimiser la sensibilité au bruit des dérivées horizontales et verticales. Les filtres étant constitués de nombres entiers, le lissage n'est qu'une approximation d'un lissage Gaussien. Les filtres F_x et F_y (3×3) se calculent ainsi :

$$F_x = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \times \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \tag{3.7}$$

$$F_y = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \times \begin{pmatrix} 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \tag{3.8}$$

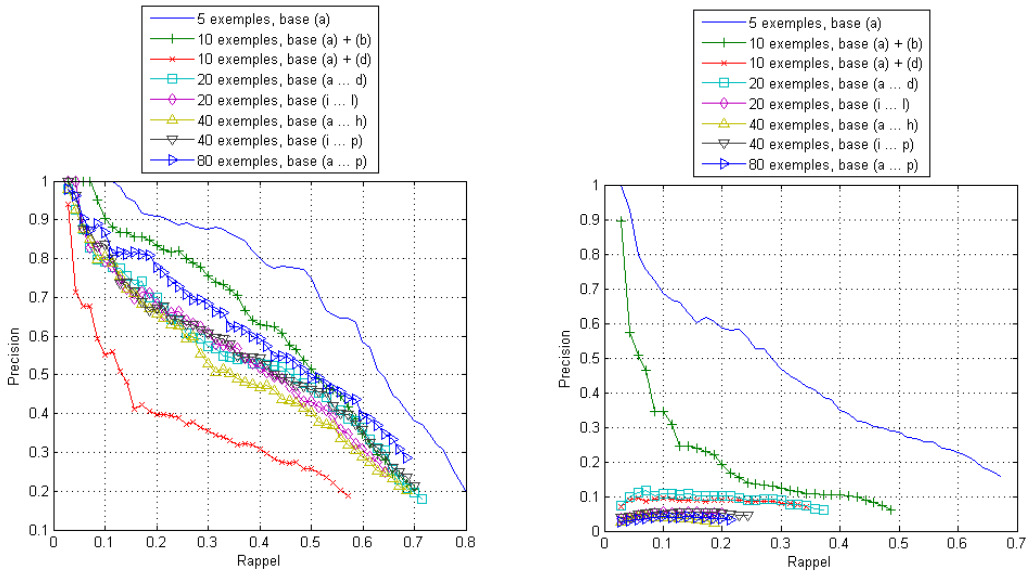
Les images filtrées correspondantes G_x et G_y sont alors calculées par convolution :

$$G_x = F_x * I \tag{3.9}$$

$$G_y = F_y * I \tag{3.10}$$

3.4.2 Corrélation avec utilisation d'un algorithme détecteur de contours

Dans cette section, nous proposons de remplacer l'image en Niveaux de Gris par une image représentant les contours. Pour se faire nous utilisons l'algorithme de Sobel. Le but est d'améliorer les taux de détection et de minimiser la sensibilité du système à la base d'exemples utilisée.



(a) Corrélation croisée normée

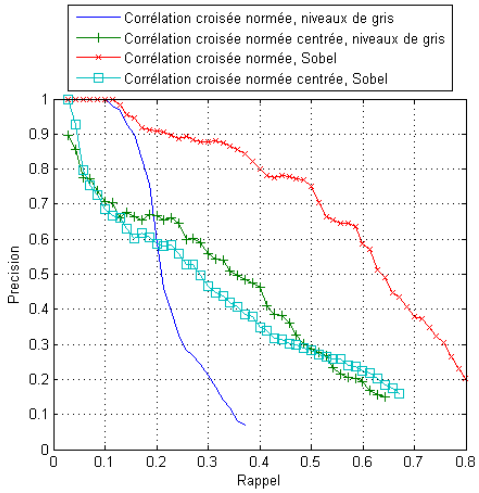
(b) Corrélation croisée normée centrée

FIGURE 3.10 – Courbes Rappel Précision du système de détection par corrélation croisée des contours des images pour différents nombres d'images de référence.

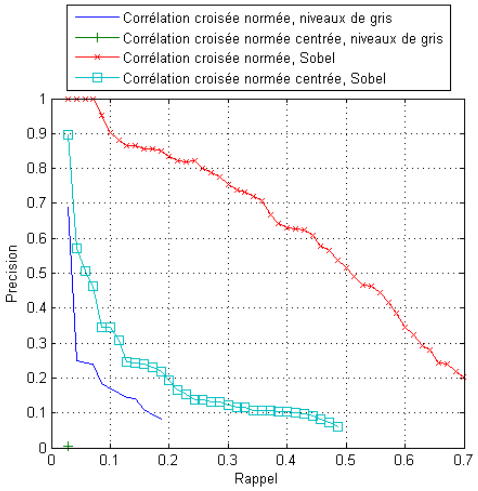
L'utilisation de la corrélation croisée normée comme mesure de similarité devient alors une mesure de la superposition des contours des images. La figure 3.10 montre

que les résultats restent très sensibles à la base de référence utilisée. On remarque notamment que les meilleurs résultats sont obtenus avec 5 et 10 exemples. Cependant, la corrélation croisée normée semble moins sensible à la base utilisée que la corrélation croisée normée centrée. De plus, l'utilisation de l'image représentant les contours en lieu et place de l'image en Niveaux de Gris pour le système de détection par corrélation croisée normée améliore nettement les résultats. En effet, nous obtenons des taux de détection toujours au moins équivalents à l'utilisation de la corrélation croisée normée centrée avec les images en Niveaux de Gris (figure : 3.11). De plus, on remarque que bien que les écarts de résultats en fonction de la base d'exemples utilisée restent importants, ils sont bien inférieurs aux écarts obtenus en utilisant les images en Niveaux de Gris.

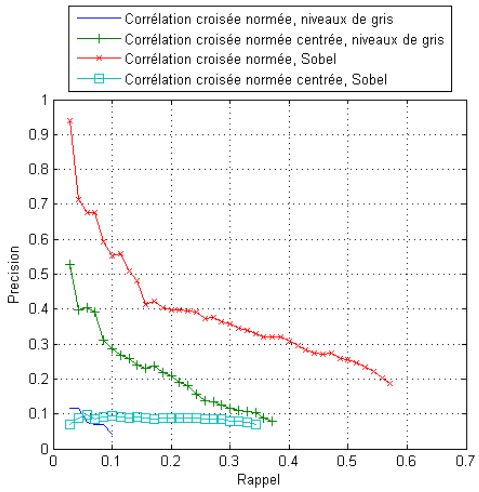
En ce qui concerne la sensibilité de la corrélation croisée normée des contours à la dimension des images exemples (figure : 3.12), on remarque que l'on améliore nettement les taux de détection en augmentant les dimensions des images de référence jusqu'à une aire d'environ 2000 pixels. Les gains pour des dimensions supérieures sont ensuite minimes.



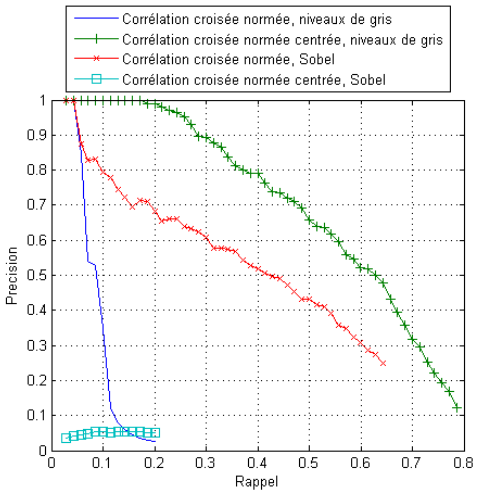
(a) 5 exemples, base (a)



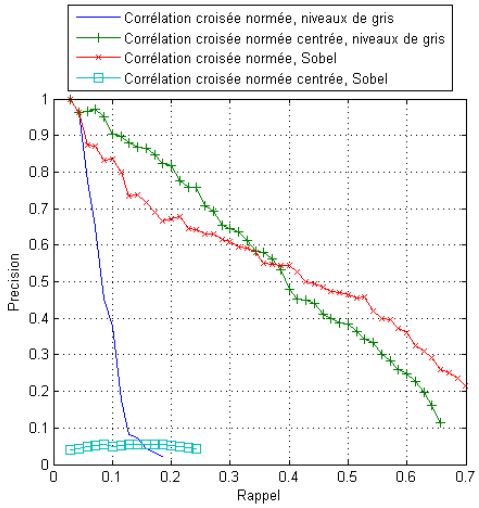
(b) 10 exemples, base (a)+(b)



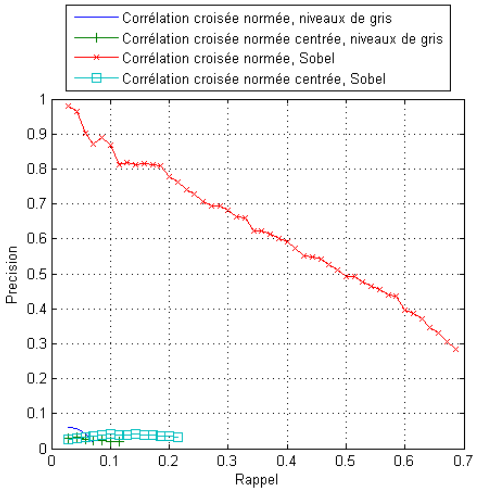
(c) 10 exemples, base (a)+(d)



(d) 20 exemples, base (i ... l)

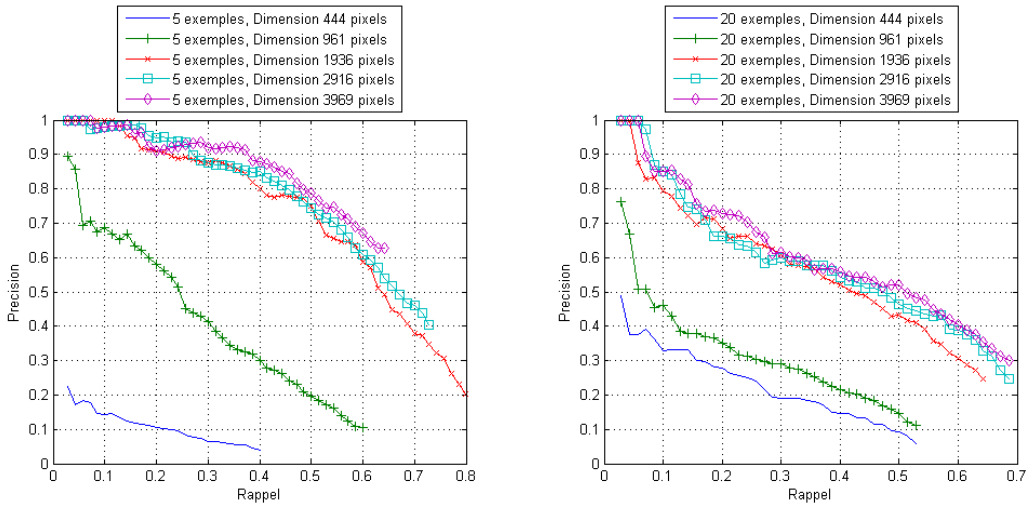


(e) 40 exemples, base (i ... p)



(f) 80 exemples, base (a ... p)

FIGURE 3.11 – Courbes Rappel Précision du système de détection par corrélation croisée. Comparaison entre l'utilisation des images en Niveaux de Gris et des images de contours. La corrélation croisée normée des contours des images semble donner les meilleurs résultats.



(a) (b)

FIGURE 3.12 – Courbes Rappel Précision du système de détection par corrélation croisée normée des contours des images avec 5 puis 20 images exemples. Influence de la dimension des exemples. On remarque une amélioration des taux de détection avec l'augmentation des dimensions des images exemples puis une stagnation lorsque l'aire des images de référence atteint 2000 pixels.

3.5 Association de la corrélation des contours et des Niveaux de Gris

L'image en Niveaux de Gris et les contours de cette dernière apportent des informations différentes. Afin d'améliorer les taux de détection du système basé sur la corrélation, nous avons associé les résultats obtenus pour les images en Niveaux de Gris et les images de contours.

3.5.1 Principe de l'association

Afin d'associer les résultats de la corrélation croisée des contours et de l'image en Niveaux de Gris, nous avons mis au point un système fonctionnant en deux phases (figure : 3.13). Une première étape que nous appellerons phase de prédétection et une seconde que nous nommerons phase de décision. Pour chaque image de référence, nous effectuons une prédétection en utilisant la corrélation croisée normée des contours des images. Un objet est prédétecté s'il correspond à un maximum local sur la carte de score et si le score correspondant est supérieur à un seuil s_1. Ainsi, le système de prédétection est équivalent au système décrit dans la section précédente sans l'utilisation de la fonction permettant d'éliminer les détections superposées. Si un objet est prédétecté, le système effectue une corrélation normée centrée entre l'image de référence et l'image en Niveaux de Gris de l'objet prédétecté. Le résultat de cette corrélation correspondra alors au score s de détection. La même opération est effectuée pour l'ensemble des images de référence. Enfin, seul sont gardées les détections non superposées avec une autre détection ayant obtenu un score s supérieur.

Un tel système présente l'avantage de ne consommer que peu de ressources supplémentaires par rapport au système présenté dans les sections précédentes. En effet, la première phase de prédétection effectue, comme pour le système précédent une corrélation normée sur l'ensemble des positions possibles de l'objet à détecter dans l'image. Cependant, la seconde phase n'effectue une corrélation que pour les objets prédétectés, ce qui ne constitue qu'une infime partie de l'ensemble des positions possibles. Les temps de calculs de la phase de décisions sont ainsi négligeables comparés à l'étape de prédétection.

Afin de tester ce système, nous avons utilisé les mêmes bases de données que précédemment. Les images de référence utilisées ont une aire de 1936 pixels, aussi bien pour la phase de prédétection que pour la phase de décision. La première question que l'on peut se poser est la valeur du seuil de prédétection s_1 que nous devons utiliser. Une valeur faible entraîne un grand nombre de prédétections et risque d'augmenter le nombre de fausses détections mais permettra de minimiser les chances de ne pas détecter un objet. Une valeur élevé de s_1 minimisera le nombre de prédétections diminuant ainsi les chances de fausses détections mais augmentera la proportions d'objets non détectés. Comme on peut le voir sur la figure 3.14, le Rappel du système de prédétection est très sensible à la valeur du seuil s_1. Une valeur de 0.70 permet de prédétecter plus de 95% des objets, mais pour une probabilité de bonne détection de seulement 1%. Une valeur de 0.77 permet d'obtenir une Précision de 1, mais ne peut détecter qu'un objet sur dix. Dans la mesure où la première phase n'est qu'une simple

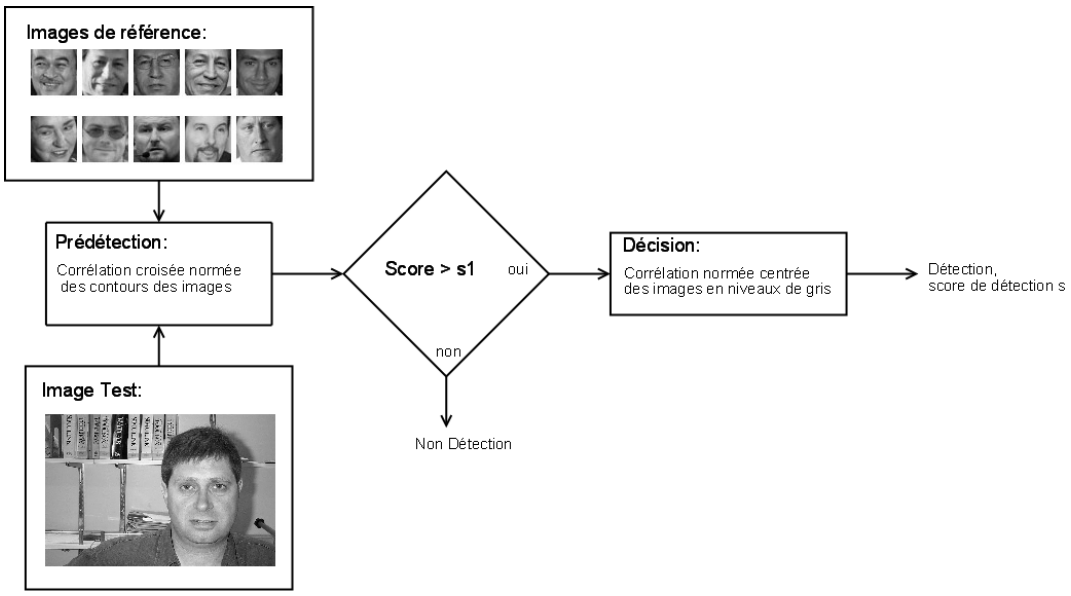


FIGURE 3.13 – Système de détection par association de la corrélation des contours et des images en Niveaux de Gris. Le système de détection par corrélation croisée normée des contours est utilisé comme système de prédétection. Les prédétections associées à un score supérieur à un seuil s_1 sont gardées et vérifiées grâce au système de décision basé sur la corrélation croisée normée centrée des images en Niveaux de Gris.

phase de prédétection, nous devons favoriser le Rappel par rapport à la Précision.

De plus, un seuil de 0.70, qui entraîne une Précision très faible pour le système de prédétection permet de ne tester qu'une centaine de positions possibles de l'objet dans la phase suivante au lieu des 327150 positions possibles au départ, minimisant ainsi grandement les possibilités de fausses détections.

Nous avons testé notre système pour quatre seuils différents (figure : 3.15). Une première fois avec une base d'exemples qui donnait de bons résultats $i.e$, la base (a). Une seconde fois avec une base d'exemples bien moins adaptée, entraînant des taux de détection très faible, $i.e$, la base (b). Une valeur du seuil s_1 de 0.72 semble donner le meilleurs compromis entre la Précision et le Rappel.

On remarque que notre système combinant deux corrélations améliore nettement les performances du système de détection (figure : 3.16), particulièrement lorsque le Rappel se rapproche de 1. Cette méthode permet, avec très peu d'exemples, d'atteindre des taux de détection permettant une utilisation pratique d'un tel système. On remarque en particulier qu'une base de seulement cinq exemples permet de détecter plus de 80% des 450 visages pour un total de seulement 32 fausses détections, c'est a dire, moins d'une fausse détection toutes les dix images. Ce système reste malgré tout très sensible à la base d'exemples utilisée. En particulier, on remarque que la base de 80 exemples obtient des résultats inférieurs aux bases de 5, 10 ou 20 exemples. Cela est dû à l'utilisation des images en Niveaux de Gris pour la phase de

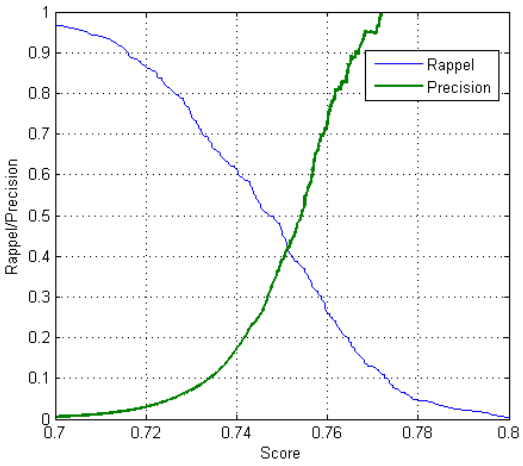
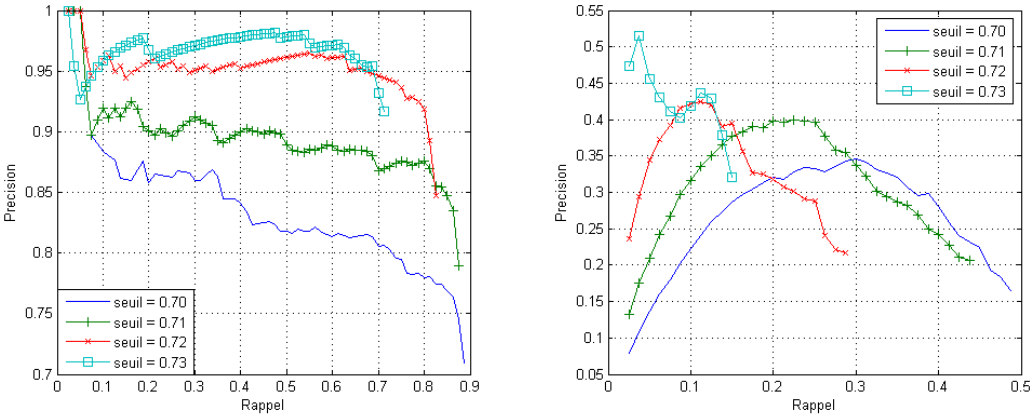


FIGURE 3.14 – Influence du seuil de prédétection sur la Précision et le Rappel du système de prédétection avec une base de 5 exemples (base (a)).

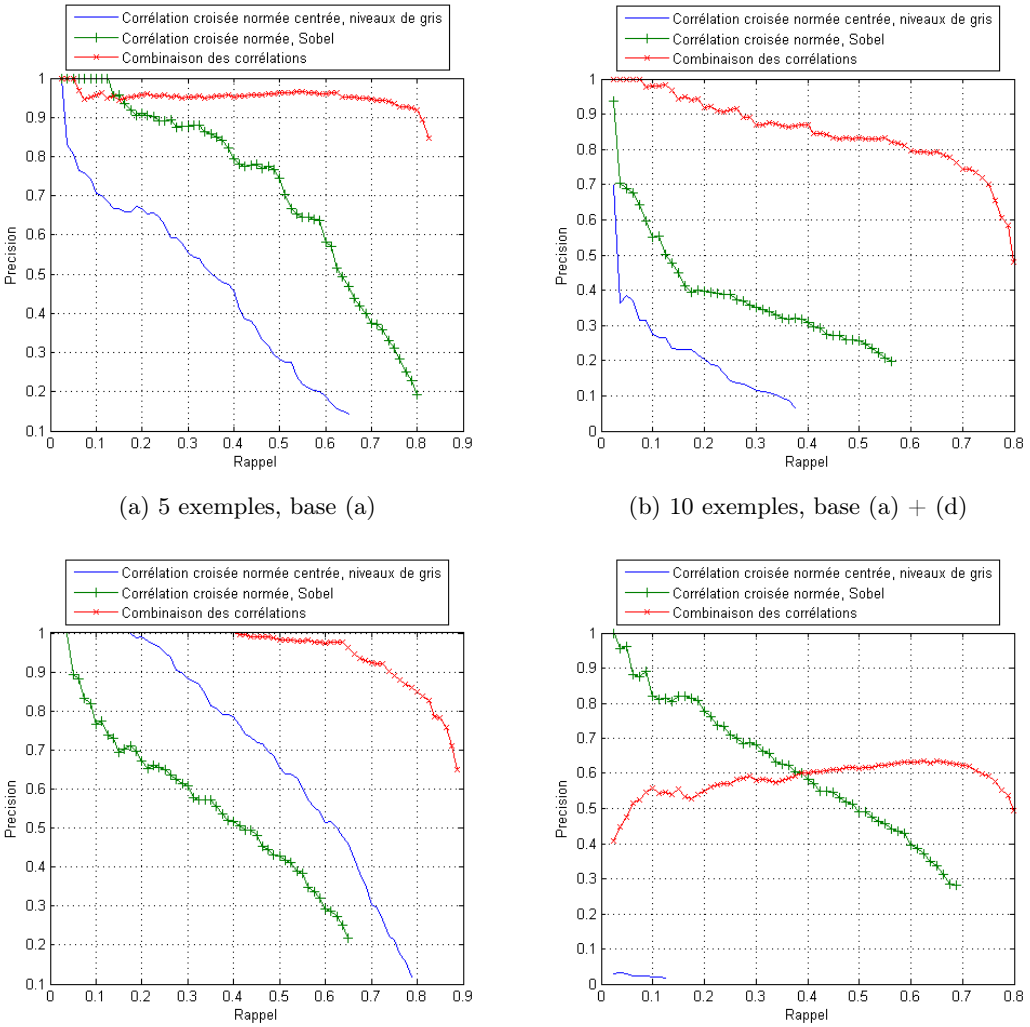


(a) 5 exemples, base (a)

(b) 5 exemples, base (b)

FIGURE 3.15 – Courbes Rappel Précision du système de détection par corrélation croisée normée des contours puis par corrélation normée centrée des images en Niveaux de Gris. Influence du seuil de prédétection.

décision qui pour cette base de 80 exemples donne des résultats très inférieurs à ceux obtenus avec les contours des images. Ainsi, de par l'utilisation des images en Niveaux de Gris, ce système reste très sensible aux variations de luminosité des images. Afin de minimiser ce problème, nous proposons dans la section suivante d'appliquer des traitements d'images permettant de minimiser les effets des variations d'éclairage sur les images en Niveaux de Gris



(a) 5 exemples, base (a)

(b) 10 exemples, base (a) + (d)

(c) 20 exemples, base (i ... l)

(d) 80 exemples, base (a ... p)

FIGURE 3.16 – Courbes Rappel Précision du système de détection par corrélation croisée normée des contours, par corrélation normée centrée des images en Niveaux de Gris, puis en combinant ces deux méthodes. Nous constatons une nette amélioration des résultats, particulièrement pour les valeurs du Rappel proches de 1.

3.5.2 Correction des variations d'illumination

La phase de décision n'est utilisée dans ce système que pour un nombre très limité de prédétections. il devient alors possible, sans allonger significativement les temps de calculs, d'utiliser des traitements d'images complexes avant d'effectuer la corrélation des Niveaux de Gris. Nous retrouvons dans la littérature deux traitements d'images très souvent utilisés pour minimiser les variations d'illumination [77, 42, 122, 112]. Le premier est la correction du gradient d'illumination par la soustraction de la fonction linéaire la mieux adaptée à l'image en Niveaux de Gris. Le second est l'égalisation d'histogramme afin de corriger les variations d'intensité de l'éclairage.

3.5.2.1 Correction du gradient d'illumination

La correction du gradient d'illumination s'effectue en soustrayant la fonction linéaire la mieux adaptée à l'image en Niveaux de Gris. Cette fonction est calculée de façon à minimiser la distance $L2$ entre l'image à traiter et la fonction linéaire bidimensionnelle. Une telle fonction s'écrit sous la forme :

$$f(x,y) = a\,x + b\,y + c \tag{3.11}$$

$$= \mathbf{p}\cdot\mathbf{v} \tag{3.12}$$

$$\mathbf{p} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \qquad \mathbf{v} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{3.13}$$

Déterminer la fonction linéaire la mieux adaptée correspond donc pour une image $I(x,y)$ de dimension $h \times l$ à calculer les paramètres $\mathbf{p}$ tels que l'erreur quadratique E soit minimum, *i.e*, trouver $\mathbf{p}$ tel que $\nabla_{\mathbf{p}} E = 0$.

$$E = \sum_{x,y} (f - I)^2 \tag{3.14}$$

$$\nabla_{\mathbf{p}} E = \underbrace{\left[\sum_{x,y} \mathbf{v}\mathbf{v}^T\right]}_{A} \mathbf{p} - \underbrace{\sum_{x,y} (I\mathbf{v})}_{\mathbf{s}} = 0 \tag{3.15}$$

$$\mathbf{p} = A^{-1}\mathbf{s} \tag{3.16}$$

Les images de références ayant toutes la même dimension $(h \times l)$ et la matrice A ne dépendant que des dimensions des images à traiter. A^{-1} peut être calculé une seule fois pour l'ensemble des images. Afin d'illustrer les résultats de ce traitement d'image, nous avons appliqué cette méthode à des images de visages (figure : 3.17).

3.5.2.2 Egalisation d'histogramme

Cette méthode permet d'augmenter le contraste des images en modifiant la distribution des Niveaux de Gris de manière à ce que l'histogramme des Niveaux de

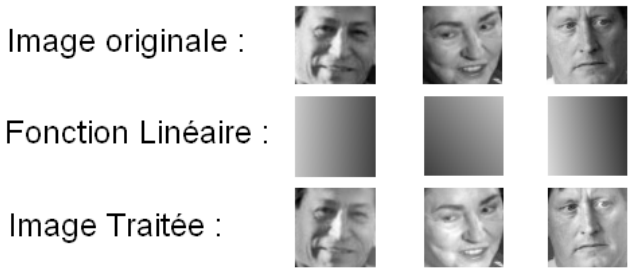


FIGURE 3.17 – Exemple d'application de la correction du gradient d'illumination sur trois images de visages. Nous pouvons constater que cette méthode permet de minimiser les effets dues aux éclairages latéraux.

Gris $H_I(g)$ corresponde à une distribution uniforme. Pour ce faire, on définit l'histogramme cumulatif de l'image I de dimension $n = h \times l$ pixels comme la fonction C_I définie sur l'intervalle de Niveaux de Gris $\{0, ..., M\}$, associant à tout niveau de gris g, le nombre de points ayant un niveau de gris inférieur ou égal à g dans l'image I. Obtenir une répartition uniforme des Niveaux de Gris dans l'image revient à associer à chaque niveau de gris g, le niveau de gris défini par la fonction de rehaussement $f(g)$ tel que $C_I(f(g))$ soit une fonction linéaire. Ceci peut être réalisé en définissant la fonction $f(g)$ comme suit :

$$f(g) = \frac{M}{2n} \left[C_I(g) + C_I(g-1) \right] \tag{3.17}$$

L'effet de cette fonction de rehaussement est d'espacer chaque Niveaux de Gris d'un écart proportionnel au nombre de points ayant ce niveau de gris dans l'image I. Ainsi, tout niveau de gris fortement représenté dans l'image se verra écarté des Niveaux de Gris adjacents. En pratique, ce traitement aura tendance à équilibrer la proportion de zones claires, moyennes et sombres dans l'image. La figure 3.18 montre l'application de cette méthode à un visage et un paysage. On constate sur ces images que l'égalisation d'histogramme permet une meilleure visualisation des détails.

3.5.2.3 Résultats expérimentaux

Afin de vérifier l'intérêt de traiter les images en Niveaux de Gris lors de la phase de décision, nous avons comparé le système de détection sans traitement d'image, avec l'utilisation de la correction du gradient d'illumination, puis l'égalisation d'histogramme et enfin la combinaison des deux traitements d'image (figure : 3.19). On remarque que les deux traitements d'image ont généralement tendance à améliorer les taux de détection. Cependant, cette amélioration est très variable en fonction de la base d'exemples utilisée. Il peut notamment arriver qu'un des deux traitements d'image entraîne une diminution des taux de détection (figure : 3.19b et 3.19d). Cependant, la combinaison des deux traitements d'image entraîne systématiquement une nette amélioration des résultats et permet, avec une base de seulement cinq

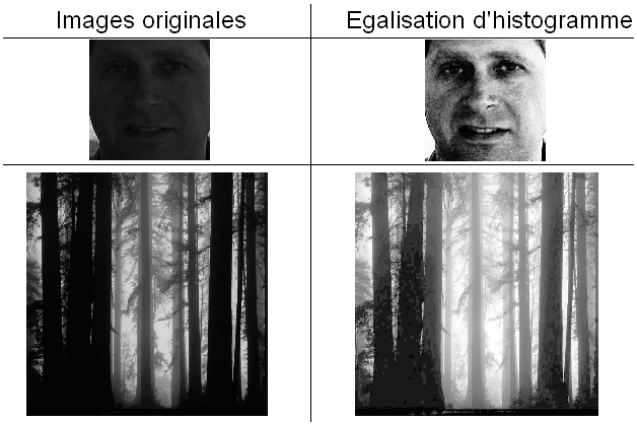
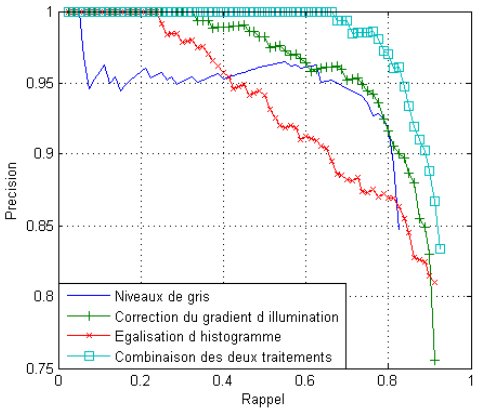
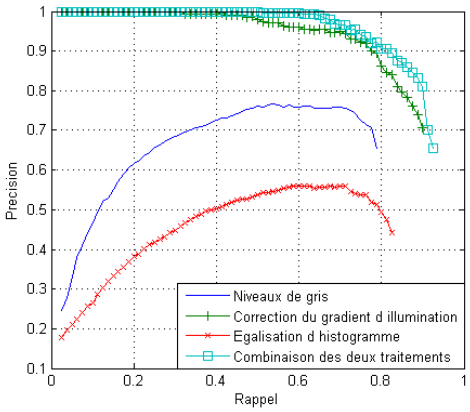


FIGURE 3.18 – Exemple d'application de l'égalisation d'histogramme sur un visage et une image de paysage. Nous pouvons constater que ce traitement permet une amélioration du contraste ainsi qu'une nette amélioration de la visualisation des détails des images.

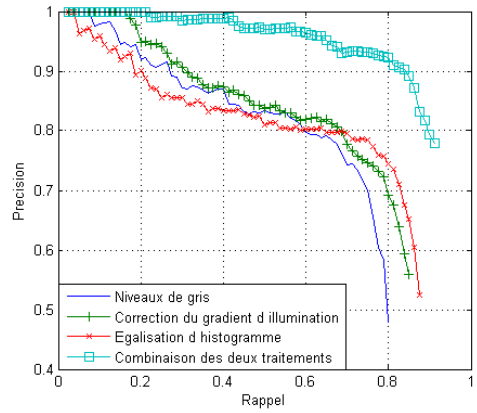
exemples, de détecter près de 90% des 450 visages pour seulement 45 fausses détections (figure : 3.19a). Nous pouvons aussi noter que la correction du gradient d'illumination se montre d'autant plus efficace que le système sans traitement d'image donne de mauvais résultats (figure : 3.19b, 3.19f). Ainsi, cette méthode de correction d'illumination permet de rendre le système de détection par corrélation beaucoup moins sensible à la base d'exemples.



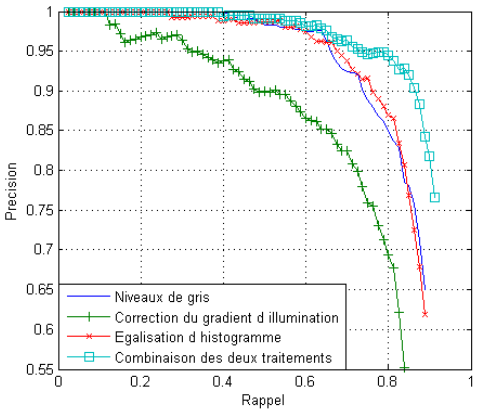
(a) 5 exemples, base (a)



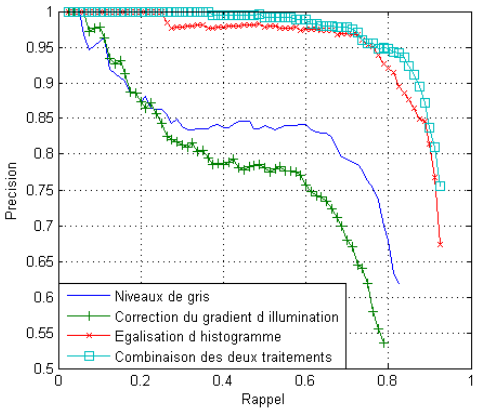
(b) 10 exemples, base (a)+(b)



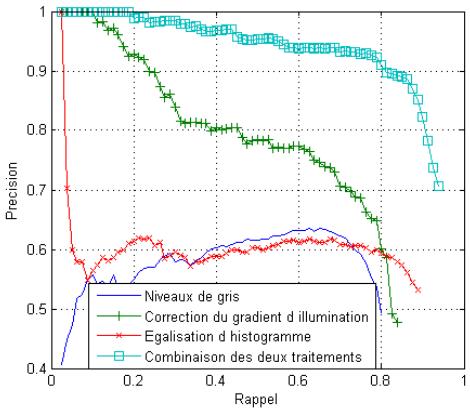
(c) 10 exemples, base (a)+(d)



(d) 20 exemples, base (i ... l)



(e) 40 exemples, base (i ... p)



(f) 80 exemples, base (a ... p)

FIGURE 3.19 – Courbes Rappel Précision du système de détection par corrélation croisée. Comparaison des résultats pour différents traitements de l'image en Niveaux de Gris. La combinaison de la correction du gradient d'illumination et de l'égalisation d'histogramme donne systématiquement les meilleurs résultats.

3.5.3 Correction des variations de forme par la méthode de déformation affine

Nous avons constaté dans la section précédente l'utilité des méthodes de correction des variations d'illumination appliquées au système de décision. Le score donné par le système de décision étant basé sur la corrélation, il est aussi sensible aux variations de forme que peut subir un objet complexe tel qu'un visage et notamment, à la rotation ou aux variations d'échelle. Afin de minimiser les effets des variations de forme sur le score du système de décision et donc sur les taux de détection, nous avons introduit une méthode permettant de déformer les images des 'objet' prédétectés de façons à maximiser la correspondance entre ces derniers et l'image de référence ayant conduit à la prédétection. Ainsi, le système de décision applique une déformation affine à l'image de l'objet prédétecté afin de corriger les variations de forme entre les deux images et de maximiser le score de la corrélation normée centrée (figure : 3.20). Par la suite, nous parlerons généralement de déformation de l'image de test par rapport à l'image de référence car, dans notre algorithme, la déformation est appliquée en modifiant la fonction bidimensionnelle représentative de l'image test. Cependant, il n'y a mathématiquement aucune différence entre déformer l'image de référence par rapport à l'image test ou l'inverse, ceci ne dépendant que du référentiel dans lequel nous nous plaçons.

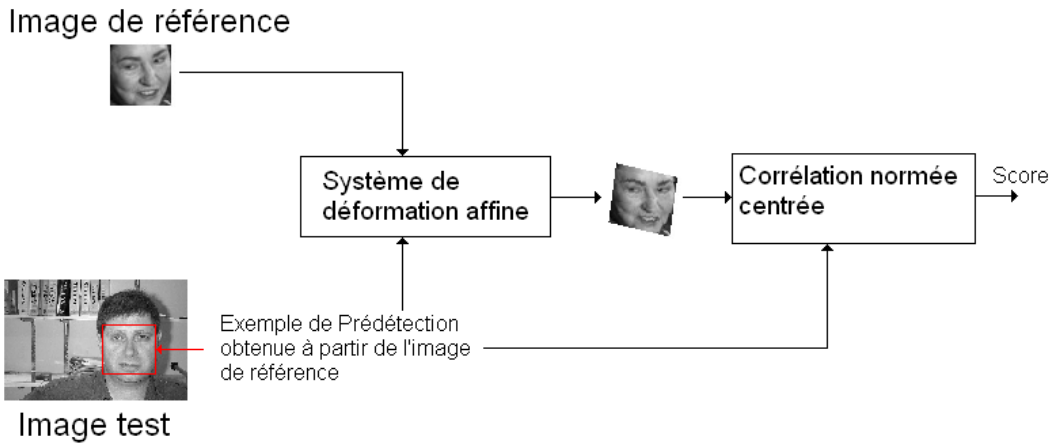


FIGURE 3.20 – Utilisation de la méthode de déformation affine dans la phase de décision du système de détection. Lorsqu'une image de référence conduit à une prédétection, on applique la méthode de déformation affine afin d'appareiller au mieux l'image de référence et l'image de test.

La méthode de compensation de mouvement par déformation affine consiste à déformer une image de test selon, le plus généralement, six paramètres afin de minimiser la distance entre cette dernière et une image de référence. Elle est le plus souvent utilisée dans l'estimation de mouvements pour des séquences d'images [28]. Le but de la compensation de mouvement est d'identifier les mouvements d'objets dans une scène ou plus précisément, la projection de ce mouvement dans le plan de l'image.

On retrouve de très nombreuses applications comme pour les systèmes de caméras stéréoscopiques [33], la compression vidéo [80], ou encore l'interpolation temporelle d'images [40, 111]. La déformation affine constitue une des méthodes d'estimation de déformation les plus simples puisque seul six paramètres sont nécessaires pour caractériser la déformation. Cependant, nous allons voir que cette méthode permet de compenser des déformations importantes appliquées à des visages.

3.5.3.1 Principe de la déformation affine

La déformation affine est l'approximation au premier ordre de la déformation subie par l'image d'un objet plan rigide soumis à un déplacement et une rotation. Ainsi, la déformation affine consiste à translater, incliner et modifier l'échelle d'une image test G afin de maximiser la correspondance avec l'image de référence F.

Si nous notons $G^* = \{g^*(\mathbf{r})\}$ le résultat de la déformation affine d'une image $G = \{g(\mathbf{r})\}$, nous pouvons écrire :

$$g^*(\mathbf{r}) = g(\mathbf{r} + \mathbf{d_r}) \tag{3.18}$$

$$\mathbf{r} = \begin{pmatrix} u \\ v \end{pmatrix} \tag{3.19}$$

$$\mathbf{d_r} = \begin{pmatrix} d_u \\ d_v \end{pmatrix} = \begin{pmatrix} a_0 u + a_1 v + a_2 \\ a_3 u + a_4 v + a_5 \end{pmatrix} \tag{3.20}$$

Les six paramètres $(a_0, ..., a_5)$ définissent la déformation affine. a_2 et a_5 correspondent à la translation, a_0, a_1, a_2 et a_3 déterminent l'inclinaison et l'échelle horizontale et verticale de l'image. Ainsi, déterminer la déformation affine qui maximise la correspondance entre les image G^* et F revient à trouver les paramètres $(a_0, ..., a_5)$ qui maximisent le critère de similarité Ψ.

3.5.3.2 Maximisation de la corrélation normée centrée

Le critère le plus couramment utilisé pour déterminer la meilleure déformation affine est la minimisation de la distance L2 entre les images que nous souhaitons appareiller. Afin de maximiser la robustesse du système de déformation aux variations d'illumination et d'être plus cohérent avec la mesure de similarité que nous utilisons déjà, nous avons utilisé la corrélation normée centrée comme critère de similarité. Ainsi, nous cherchons les paramètres $(a_0, ..., a_5)$ qui maximisent la fonction Ψ :

$$\Psi = \sum_{\mathbf{r} \in F} \overbrace{\left(\frac{f(\mathbf{r}) - m_f}{\sigma_f} \right)}^{f_n} \underbrace{\left(\frac{g(\mathbf{p} + \mathbf{r} + \mathbf{d_r}) - m_g}{\sigma_g} \right)}_{g_n} \tag{3.21}$$

$F = \{f(\mathbf{r})\}$ et $G = \{g(\mathbf{r})\}$ sont respectivement l'image de référence de dimension $n = l \times h$ et l'image de test, $\mathbf{p}$ les coordonnées initiales de G ou l'objet a été prédétecté.

m_f et m_g sont les moyennes des fonctions $f(\mathbf{r})$ et $g^*(\mathbf{p} + \mathbf{r})$, $\mathbf{r} \in F$:

$$m_f = \frac{1}{n} \sum_{\mathbf{r} \in F} f(\mathbf{r}) \tag{3.22}$$

$$m_g = \frac{1}{n} \sum_{\mathbf{r} \in F} g(\mathbf{p} + \mathbf{r} + \mathbf{d_r}) \tag{3.23}$$

σ_f et σ_g représentent l'écart type des fonctions $f(\mathbf{r})$ et $g^*(\mathbf{p} + \mathbf{r})$, $\mathbf{r} \in F$:

$$\sigma_f = \sqrt{\sum_{\mathbf{r} \in F} \left(f(\mathbf{r}) - m_f\right)^2} \tag{3.24}$$

$$\sigma_g = \sqrt{\sum_{\mathbf{r} \in F} \left(g(\mathbf{p} + \mathbf{r} + \mathbf{d_r}) - m_g\right)^2} \tag{3.25}$$

Maximiser la fonction Ψ conduit donc à résoudre le système de six équations suivant :

$$\frac{\partial \Psi}{\partial a_i} = 0 \qquad i \in [0, 5] \tag{3.26}$$

Ce système d'équation ne peut pas être résolu analytiquement, cependant il existe différentes méthodes permettant d'optimiser la fonction Ψ. Ce type de problèmes est le plus souvent résolu par des méthodes de descente par gradient. Cependant, le problème d'optimisation n'étant que de dimension six, il semble approprié d'utiliser une méthode d'optimisation non linéaire. Dans [28] Dugelay et Sanson montrent que la méthode d'optimisation itérative de Gauss Newton permet une convergence rapide et robuste vers une solution au problème de la déformation affine.

Cette méthode est basée sur deux approximations pour effectuer l'optimisation :

- La fonction Ψ à optimiser est localement une fonction polynomiale du second ordre.
- La dérivée seconde de la fonction g est nulle (La matrice Hessien de $g(\mathbf{r})$, $H_g = \mathbf{0}$). Autrement dit, la variation de la luminance de l'image G est localement linéaire.

On note $A_k = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 \end{pmatrix}^T$ la valeur des paramètres de la déformation affine à la k^{ieme} itération.

De par l'approximation de la forme localement polynomiale à l'ordre deux de la fonction Ψ, les paramètres de la déformation affine à la $(k+1)^{ieme}$ itération se déterminent par l'équation suivante.

$$A_{k+1} = A_k - H_A^{-1} G_A \tag{3.27}$$

Ou H_A est la matrice Hessien de la fonction Ψ et G_A en est le gradient.

$$G_A = \begin{pmatrix} \frac{\partial \Psi}{\partial a_i} \\ \vdots \end{pmatrix} \quad H_A = \begin{pmatrix} \frac{\partial^2 \Psi}{\partial a_i \partial a_j} & \cdots \\ \vdots & \ddots \end{pmatrix} \tag{3.28}$$

Afin de rendre plus claires les équations, nous utiliserons les notations suivantes :

$$\begin{array}{ll} g_p \Leftrightarrow g(\mathbf{p}+\mathbf{r}+\mathbf{d_r}) & \\ \mathbf{g_r} \Leftrightarrow \nabla_r(g_p) & \mathbf{d_r^i} \Leftrightarrow \frac{\partial \mathbf{d_r}}{\partial a_i} \\ m_g^i \Leftrightarrow \frac{\partial m_g}{\partial a_i} & \sigma_g^i \Leftrightarrow \frac{\partial \sigma_g}{\partial a_i} \end{array}$$

- g_p la valeur de g au point $(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$.
- $\mathbf{g_r}$ la valeur du gradient de g au point $(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$.
- $\mathbf{d_r^i}$ la fonction dérivée de $\mathbf{d_r}$ par rapport au paramètre a_i.
- m_g^i la fonction dérivée par rapport au i^{ieme} paramètre de la moyenne de la fonction $g(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$, $\mathbf{r} \in F$.
- σ_g^i la fonction dérivée de l'écart type de $g(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$, $\mathbf{r} \in F$ par rapport à a_i.

Afin de déterminer A_{k+1}, nous devons calculer à chaque itération les matrices G_A et H_A. L'approximation de linéarité locale de la fonction $g(\mathbf{r})$ permet de déterminer G_A et H_A simplement à partir des fonctions $f(\mathbf{r})$ et $g(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$ ainsi que le gradient de $g(\mathbf{r})$ numériquement calculable par des méthodes d'approximation bilinéaires. G_A se détermine ainsi :

$$\begin{aligned} \frac{\partial \Psi}{\partial a_i} &= \sum_{\mathbf{r} \in F} \left(\frac{f(\mathbf{r}) - m_f}{\sigma_f} \right) \frac{\partial}{\partial a_i} \left(\frac{g_p - m_g}{\sigma_g} \right) \\ &= \sum_{\mathbf{r} \in F} f_n \frac{\partial g_n}{\partial a_i} \\ &= \sum_{\mathbf{r} \in F} f_n \frac{\left({\mathbf{d_r^i}}^T \mathbf{g_r} - m_g^i \right) \sigma_g - (g_p - m_g)\, \sigma_g^i}{\sigma_g^2} \end{aligned} \tag{3.29}$$

Avec :

$$m_g^i = \frac{1}{n} \sum_{\mathbf{r} \in F} {\mathbf{d_r^i}}^T \mathbf{g_r} \tag{3.30}$$

Si on note $V_g^i = \frac{\partial V_g}{\partial a_i}$ la dérivée de la variance V_g de $g(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$:

$$\begin{aligned} \sigma_g^2 = \quad V_g &= \sum_{\mathbf{r} \in F} (g_p - m_g)^2 \\ V_g^i &= \sum_{\mathbf{r} \in F} 2 \left({\mathbf{d_r^i}}^T \mathbf{g_r} - m_g^i \right) (g_p - m_g) \\ \sigma_g^i &= \frac{V_g^i}{2 V_g^{\frac{1}{2}}} \end{aligned} \tag{3.31}$$

De même, si on remarque que $\forall(i,j)$, $\frac{\partial \mathbf{d_r^i}}{\partial a_j} = \mathbf{0}$, La matrice Hessien H_A est déterminée comme suit :

$$\frac{\partial^2 \Psi}{\partial a_i \partial a_j} = \sum_{\mathbf{r}\in F} f_n \frac{\partial^2 g_n}{\partial a_i \partial a_j} \tag{3.32}$$

$$\begin{aligned}\sigma_g^4 \frac{\partial^2 g_n}{\partial a_i \partial a_j} = & \\ & 2\left(g_p - m_g\right)\sigma_g \sigma_g^i \sigma_g^j - \left(\mathbf{d_r^j}^T \mathbf{g_r} - m_g^j\right)\sigma_g^i \sigma_g^2 - \\ & \left(g_p - m_g\right)\sigma_g^{ij}\sigma_g^2 - \left(\mathbf{d_r^i}^T \mathbf{g_r} - m_g^i\right)\sigma_g^j \sigma_g^2\end{aligned} \tag{3.33}$$

$\sigma_g^{ij} = \frac{\partial^2 \sigma_g}{\partial a_i \partial a_j}$ étant la dérivée seconde de σ_g par rapport aux paramètres a_i et a_j.

$$\begin{aligned}\sigma_g^{ij} &= \frac{\partial}{\partial a_j}\frac{1}{2}\left(V_g^i V_g^{\frac{-1}{2}}\right) \\ &= \frac{1}{2V_g}\left(V_g^{ij}\sigma_g - \frac{V_g^i V_g^j}{2\sigma_g}\right)\end{aligned} \tag{3.34}$$

Avec $V_g^{ij} = \frac{\partial V_g^i}{\partial a_j}$ La dérivée seconde de la variance V_g :

$$V_g^{ij} = \sum_{\mathbf{r}\in F} 2\left(\mathbf{d_r^i}^T \mathbf{g_r} - m_g^i\right)\left(\mathbf{d_r^j}^T \mathbf{g_r} - m_g^j\right) \tag{3.35}$$

Une fois les matrices G_A et H_A déterminées pour la k^{ieme} itération, il suffit d'inverser la matrice H_A et d'utiliser l'équation 3.27 afin de déterminer une nouvelle valeur des paramètres de la déformation affine. Nous considérons qu'il y a convergence si la distance $L2$ entre les paramètres calculés pour deux itérations successives est inférieure à un seuil $d = 10^{-5}$, le nombre maximum d'itérations du système étant fixé à 40 afin d'éviter des temps de calculs prohibitifs et des solutions trop éloignées des conditions initiales. Autrement dit, la solution à la détermination des paramètres optimums de la déformation affine est donnée par A^k, la valeur des six paramètres à la k^{ieme} itération, avec $k \leq 40$ et $\left\|A^k - A^{k-1}\right\| < d$.

3.5.3.3 Résultats expérimentaux sur la corrélation d'images de visages

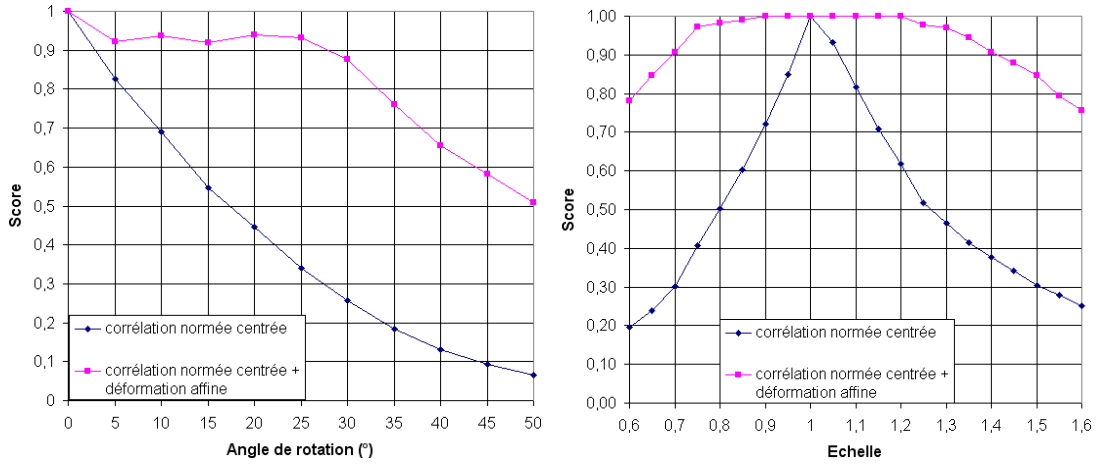
Dans cette section, nous présentons deux expérimentations distinctes. La première partie consiste à mesurer la capacité de la méthode de déformation affine à compenser les déformations et à améliorer la robustesse de notre mesure de similarité basée sur la corrélation normée centrée. La seconde partie aura pour but de mesurer l'efficacité d'une mesure de similarité basée sur la déformation affine et la corrélation normée centrée pour notre système de détection.

Déformation affine appliquée à la compensation de déformations : Afin de mesurer l'efficacité de la déformation affine pour améliorer la robustesse de la corrélation normée centrée aux rotations et aux changements d'échelles, nous avons commencé par comparer le score de corrélation normée centrée avec et sans l'utilisation de la méthode de déformation affine. Pour ce faire, nous avons utilisé 80 images contenant chacune un visage de dimension 44×44 soit environ 2000 pixels. Pour chaque image, nous en avons extrait le visage et appliqué une rotation ou un changement d'échelle afin d'obtenir une image de référence F que nous comparons à l'image originale G. Ainsi, plus la méthode de déformation affine sera capable de compenser les déformations dues à la rotation ou au changement d'échelle, plus le score de la mesure de similarité associée se rapprochera de 1. Nous avons reporté sur la figure 3.21 le score moyen de la corrélation normée centrée en fonction des déformations (rotation ou changement d'échelle) avec et sans l'utilisation de la méthode de déformation affine. La première constatation que nous pouvons faire est que la corrélation normée centrée est assez sensible aux déformations puisqu'une rotation de 10° ou une variation de 10% de l'échelle entraîne une chute du score moyen de 30%. La seconde constatation que nous faisons est que la méthode de déformation affine permet de compenser efficacement les déformations que nous avons appliquées. Ainsi, le système de déformation affine est capable d'appareiller des images de visages ayant subi des rotations de plus de 40° ou de fortes variations d'échelle.

Si les déformations dues aux variations d'échelle relativement faibles sont parfaitement compensées par la méthode de déformation affine, ce n'est pas le cas de la rotation qui n'obtient pas un score moyen de corrélation de 1 après compensation même pour des rotations faibles. Ceci s'explique par les interpolations effectuées lors du calcul des rotations des images. Cependant, on constate que la déformation affine permet tout de même de bien corriger cette déformation puisque nous obtenons un score moyen de 0.9 pour une rotation des visages de 30° (une correction parfaite donnant un score de corrélation de 1).

Enfin, afin de mesurer l'efficacité de l'algorithme d'optimisation de Newton Gauss appliqué à l'optimisation de la corrélation normée centrée, nous avons mesuré le nombre moyen d'itérations nécessaire à notre méthode pour converger vers une solution (figure : 3.22). Si pour de fortes déformations le nombre moyen d'itérations de l'algorithme tend vers les 40 itérations qui correspondent au nombre maximum autorisées, nous pouvons constater que pour des déformations modérées, l'algorithme converge en seulement une dizaine d'itérations, démontrant ainsi l'efficacité de cette méthode. Une approche multirésolution permettrait certainement de corriger des variations d'échelle et des rotations plus importantes, cependant nous n'en avons pas l'utilité pour notre système de détection.

Application au système de détection : Nous avons pu constater que la méthode de déformation affine permet de compenser efficacement les déformations simples appliquées à des visages. Cependant, cela ne permet pas d'assurer l'efficacité d'une mesure de similarité basée sur la déformation affine et la corrélation normée centrée dans le cadre d'un système de détection. En effet, si la compensation de mouvement

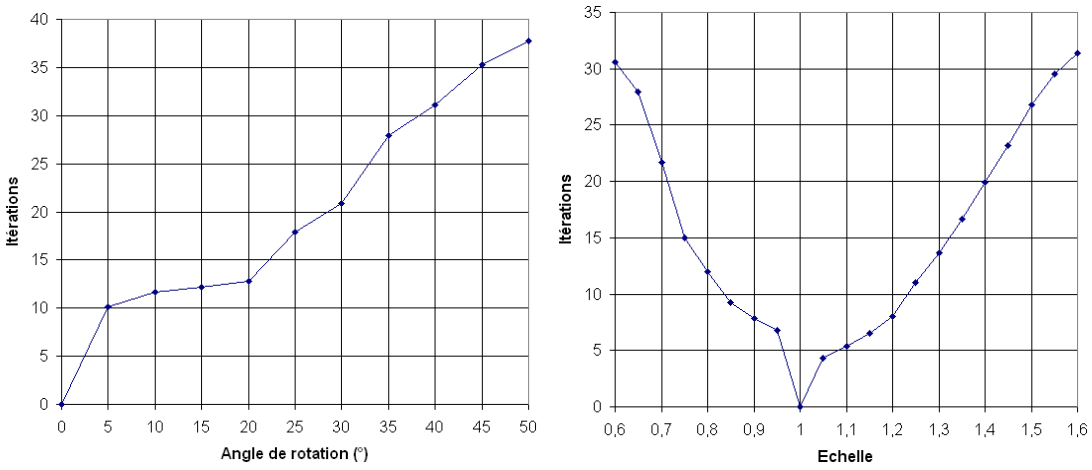


(a) (b)

FIGURE 3.21 – Score moyen sur 80 visages de la corrélation normée centrée avec et sans l'application de la méthode de déformation affine pour des visages ayant subi une rotation ou un changement d'échelle. On constate que la corrélation normée centrée est très sensible à la rotation ou aux variations d'échelle, cependant le système de déformation affine permet de compenser ces déformations et d'obtenir une mesure de similarité presque insensible aux variations modérées d'échelle ou d'angle d'inclinaison des visages.

par déformation affine permet de maximiser le score de corrélation entre deux visages en compensant les variations de forme entre ces derniers, elle permet aussi de maximiser le score de corrélation entre un visage de référence et un objet qui ne serait pas un visage. Autrement dit, la méthode de déformation peut entraîner un score de détection plus important pour les fausses alarmes et provoquer une baisse des performances du système de détection. Afin de tester l'utilité d'un tel système, nous avons appliqué le système de détection à différentes bases de référence. Nous avons comparé les résultats obtenus avec et sans l'utilisation de la déformation affine (figure : 3.23). Nous constatons que l'efficacité du système de déformation affine est relativement dépendante de la base de détection utilisée, si dans certains cas nous constatons une amélioration sensible des résultats (figure : 3.25b,3.23f), le système de déformation peut aussi conduire à une baisse des taux de détection (figure : 3.25b, 3.23d). De façon générale, le système de déformation affine tend à augmenter la Précision au détriment du Rappel. En effet, la déformation affine permet d'augmenter la proportion de détection avec un score de similarité élevé qui garantit une très faible probabilité d'effectuer une fausse détection, ce qui augmente la Précision du système lorsque cette dernière est déjà proche de 1. Cependant, la déformation affine augmente aussi le score de similarité de nombreuses fausses détections, entraînant une baisse de la Précision lorsque cette dernière diminue.

Si nous combinons la méthode de déformation affine avec la correction d'illumination, c'est à dire, que nous appliquons la correction du gradient d'illumination

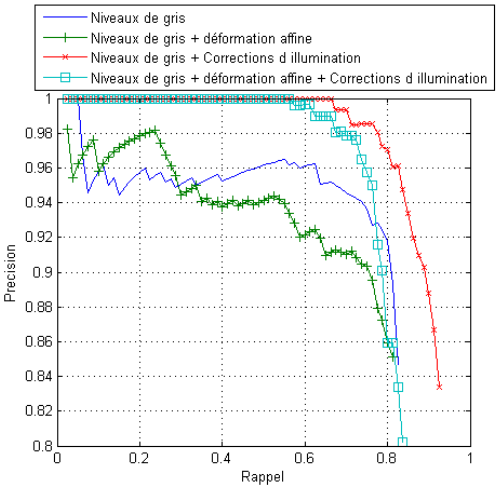


(a) (b)

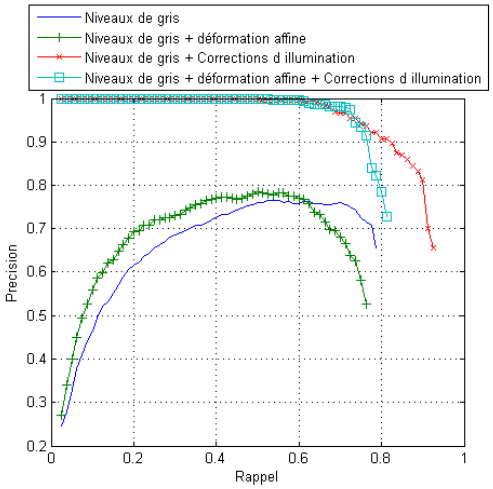
FIGURE 3.22 – Nombre moyen d'itérations nécessaire au système de déformation affine pour converger vers une solution pour 80 visages ayant subi une rotation ou un changement d'échelle. Plus la déformation à compenser est importante plus le nombre d'itérations nécessaire pour obtenir une solution augmente. Nous pouvons cependant remarquer que pour des rotations et des variations d'échelle modérées, le système de déformation affine converge en une dizaine d'itérations.

et l'égalisation d'histogramme aux images déformées avant d'effectuer la corrélation normée centrée donnant le score de décision, nous ne constatons pas une amélioration générale des résultats par rapport au même système sans l'utilisation de la déformation affine. Cependant, on remarque que le système de détection est moins sensible à la base de données utilisée, en particuliers, les meilleurs taux de détection sont atteints pour le système utilisant le maximum d'exemples ce qui n'était pas le cas sans l'utilisation de la méthode déformation affine.

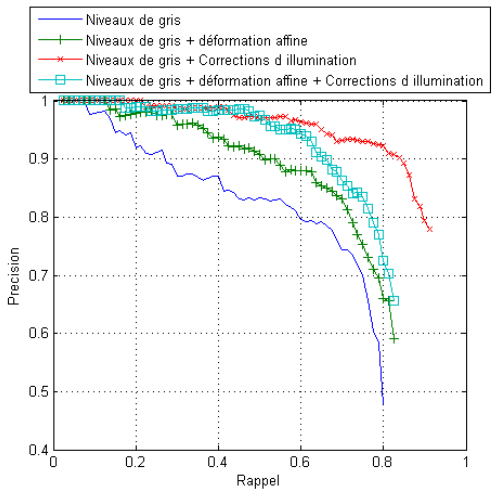
Si la correction des variations d'illumination et, dans une moindre mesure, des déformations affines des images en Niveaux de Gris a permis une amélioration sensible des résultats, nous pouvons constater que le gain de performances le plus important est apporté par l'association des deux mesures de similarité basées sur la corrélation d'images dont nous avons extrait des informations différentes (Contours et images en Niveaux de Gris).



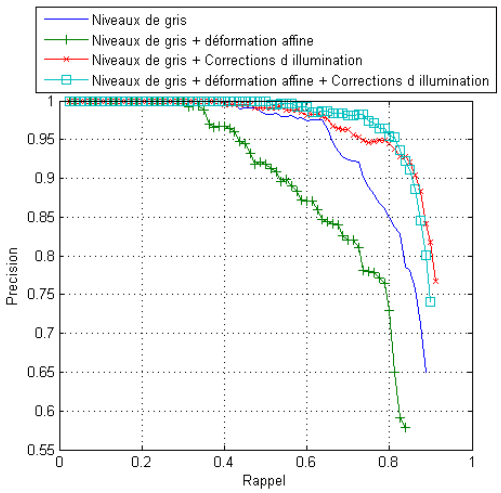
(a) 5 exemples, base (a)



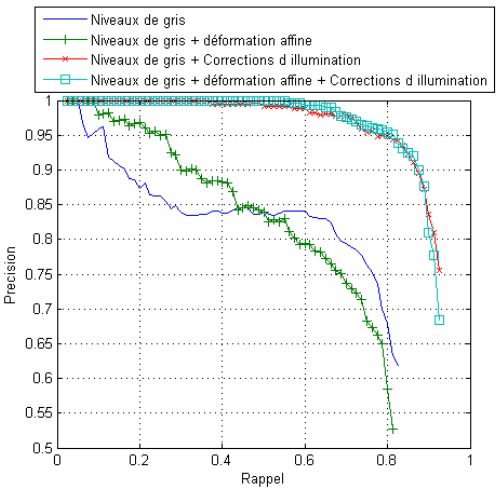
(b) 10 exemples, base (a)+(b)



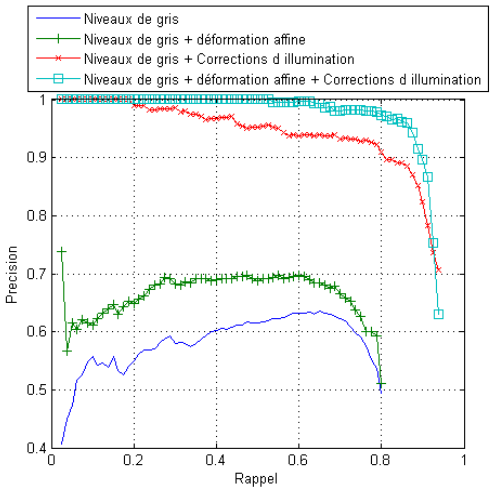
(c) 10 exemples, base (a)+(d)



(d) 20 exemples, base (i ... l)



(e) 40 exemples, base (i ... p)



(f) 80 exemples, base (a ... p)

FIGURE 3.23 – Courbes Rappel Précision du système de détection par corrélation croisée. Comparaison des résultats pour différents traitements des images en Niveaux de Gris, avec et sans l'utilisation de la méthode de déformation affine. On observe que la méthode de déformation affine a tendance à améliorer la Précision au détriment du Rappel.

3.6 Corrélation croisée avec utilisation de filtres de contours orientés

Dans les sections précédentes, nous avons pu constater que si une simple corrélation croisée ne permet pas d'effectuer une détection efficace d'objets complexes, l'utilisation de combinaisons de corrélations des contours et des images en Niveaux de Gris permet une très nette amélioration des résultats. Ainsi, l'utilisation de plusieurs corrélations apportant chacune une information différente permet l'utilisation d'une mesure de similarité aussi simple qu'une corrélation pour détecter des objets aussi complexes que des visages. Dans cette section, nous proposons d'utiliser les filtres de contours orientés afin de combiner non plus seulement une image des contours et l'image en Niveaux de Gris, mais plusieurs images de contours.

3.6.1 Méthode d'association des mesures de similarité

Le système précédent associait deux mesures de similarité (corrélation normée des contours des images et corrélation normée centrée des images en Niveaux de Gris) en divisant le système en deux parties. La première mesure de similarité étant utilisée pour effectuer une prédétection, *i.e*, donner les positions et échelles probables de l'ensemble des objets que nous souhaitons détecter. La seconde mesure de similarité consistant à donner un score nous permettant de décider si la prédétection est correcte. Autrement dit, le système de détection détecte un objet si, à une certaine position et échelle, le score de similarité donné par le système de prédétection est supérieur à un seuil s_1 et le score de similarité donné par le système de décision est supérieur à un seuil final s_f ; la valeur du seuil s_f permettant de choisir un compromis entre le Rappel et la Précision du système de détection. Afin de généraliser cette méthode à un nombre quelconque de mesures de similarité, nous avons modifié le fonctionnement du système de détection comme indiqué sur la figure 3.24. Ainsi, chacune des n mesures de similarité entraîne un score de détection s_i. Le score final du système de détection est alors le score de la mesure de similarité ayant donnée la valeur la plus basse. Autrement dit, le score final du système de détection est donné par la mesure de similarité j telle que, quel que soit i, $s_j \leq s_i$. Ainsi, si l'on fixe le seuil de détection final à une valeur fixe s_f, un objet sera détecté si l'ensemble des mesures de similarité donne un score supérieur au seuil s_f. De cette façon, comme pour le système basé sur une phase de prédétection et une phase de décision, les temps de calculs ne seront que peu impactés par le nombre de mesures utilisées. En effet, un tel système constitue en fait une cascade classifieurs, à l'image du système de détection d'objets de Viola-Jones [122]. La seconde mesure de similarité n'est effectuée que si la première est supérieure au seuil s_f, la troisième si les deux premières sont supérieures à ce seuil et de même jusqu'à la dernière mesure de similarité.

Finalement, comme précédemment, nous éliminons les détections superposées en ne gardant que les détections non superposées à une autre détection avec un score supérieur.

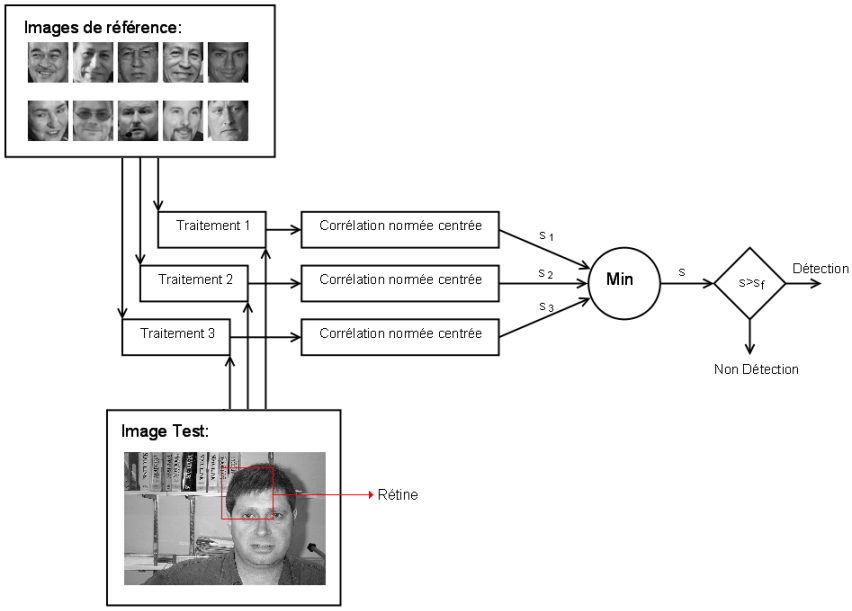


FIGURE 3.24 – Système de détection par association de corrélations (exemple avec trois corrélations). Chaque image de référence et rétines extraites de l'image de test est traitée selon trois méthodes donnant trois images distinctes ; par exemple, l'image en Niveaux de Gris, les contours verticaux et les contours horizontaux. Le traitement d'image associé à la corrélation normée centrée constitue alors une mesure de similarité. Si l'ensemble des scores de similarité s_i est supérieur au seuil s_f, alors nous détectons un objet à l'emplacement et l'échelle correspondant à la rétine. Finalement, nous éliminons les détections superposées par la même méthode que précédemment.

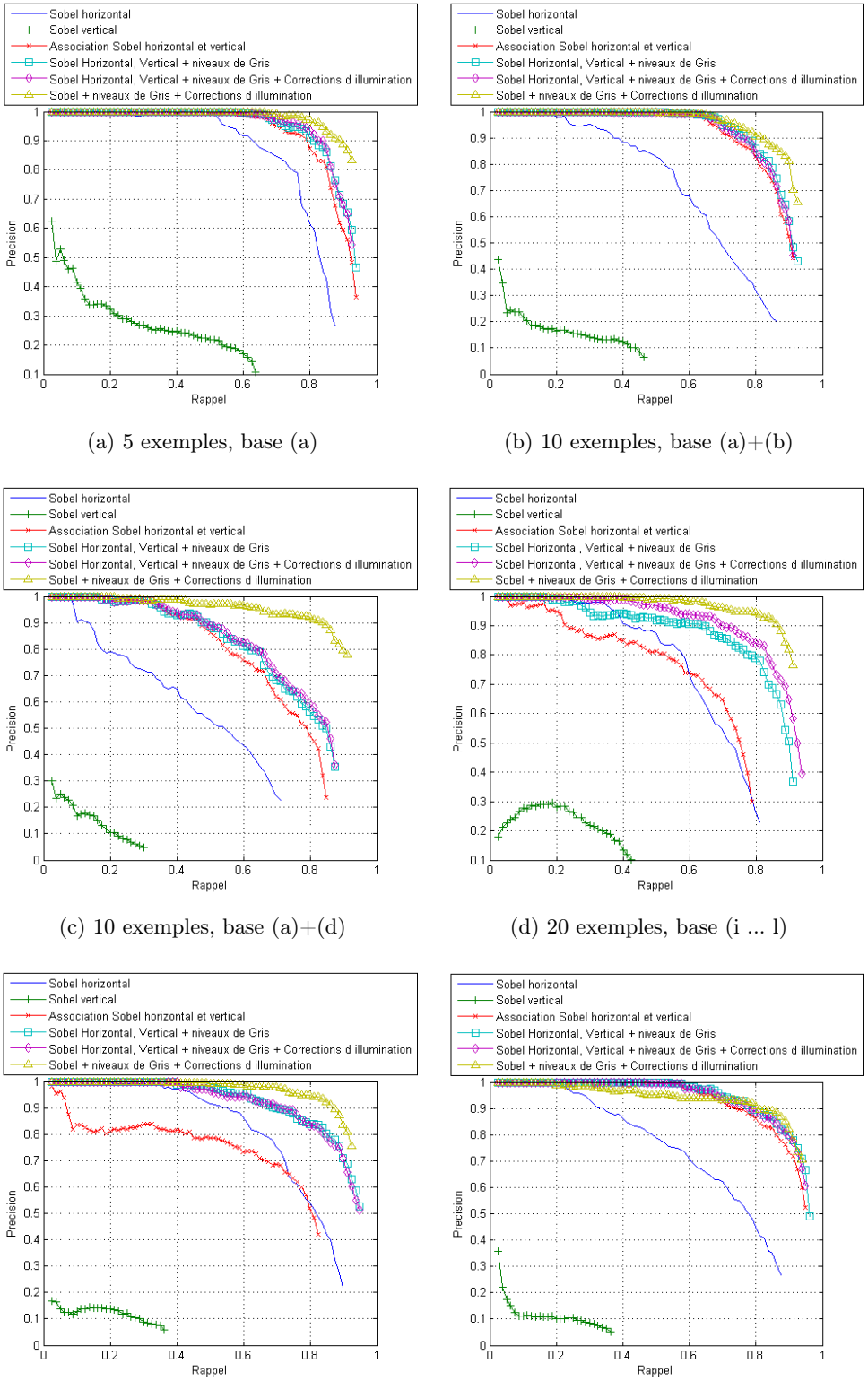
3.6.2 Utilisation des filtres de contours horizontaux et verticaux de Sobel

Nous avons remarqué que la corrélation des contours des images extraits à partir de la méthode de Sobel permet de rendre le système de détection par corrélation plus performant par rapport à l'utilisation des images en Niveaux de Gris. Nous avons ensuite associé la corrélation de ces contours avec la corrélation des images en Niveaux de Gris ce qui nous a permis une nette amélioration du système de détection. Dans cette section nous proposons, non plus d'utiliser les images de contours extraits par la méthode de Sobel, mais les filtres horizontaux et verticaux utilisés par Sobel. Ainsi, le premier et le second traitement du système de détection consisteront à extraire les contours horizontaux et verticaux des images par convolution avec les filtres utilisés par Sobel. Un troisième traitement consistant à utiliser directement l'image en Niveaux de Gris ou l'image en Niveaux de Gris avec correction des variations d'illumination sera aussi évalué. A cause de la sensibilité du système de détection à la base de référence, nous avons, comme pour les expérimentations précédentes, utilisé différentes bases de référence.

La figure 3.25 montre les résultats obtenus sous forme de courbes Rappel Précision. Afin de montrer l'intérêt de l'association des mesures de similarité, nous avons reporté les résultats obtenus avec la seule extraction des contours horizontaux puis verticaux, puis l'association des deux. Nous avons ensuite associé un troisième traitement consistant dans un premiers temps à la simple utilisation de l'image en Niveaux de Gris, puis le traitement permettant de corriger les variations d'illumination que nous avons utilisé pour le système précèdent. Enfin, afin d'avoir un point de comparaison, nous avons fourni les résultats obtenus par le système de détection précédent basée sur la phase de prédétection et la phase de décision avec la correction des variations d'illumination de l'image en Niveaux de Gris. La première constatation que l'on peut faire et sans doute la plus évidente est que, quel que soit la base d'images utilisée l'association des mesures de similarité apporte un gain de performance très important. En effet, l'association de la corrélation normée centrée des contours horizontaux et verticaux entraîne presque systématiquement une nette amélioration des résultats qui s'approchent de ceux obtenus avec le système précédent comportant un traitement complexe de correction d'illumination ainsi que la nécessité de régler un seuil de prédétection. Cependant, nous pouvons constater que si la corrélation des contours horizontaux sans associations entraîne des résultats satisfaisants, ce n'est pas le cas des contours verticaux qui donnent des résultats très variables et souvent bien inférieurs à ceux obtenus par la corrélation des contours horizontaux. Ainsi, dans certains cas ou les résultats obtenus par la corrélations des contours verticaux et horizontaux sont trop différents, alors l'association des deux corrélations peut entraîner une stagnation des résultats (figure : 3.25d), ou même une baisse (figure : 3.25e). Cependant, si l'on combine les corrélations des contours horizontaux et verticaux ainsi que l'image en Niveaux de Gris, nous obtenons une nette amélioration de l'ensemble des résultats. Ainsi, l'association des trois mesures de similarité permet de rendre le système plus robuste à la base utilisée. Dans le cas ou la combinaison des deux premiers traitements entraîne une amélioration des résultats, l'utilisation des images en Niveaux de Gris n'entraîne qu'une légère amélioration (figure : 3.25a,3.25b,3.25c,3.25f). Nous avons ensuite utilisé la correction du gradient d'illumination et l'égalisation d'histogramme sur les images en Niveaux de Gris de la troisième corrélation. Nous constatons que contrairement au système précédent, l'amélioration des résultats est très faible rendant de tels traitements presque inutiles.

En conclusion, nous pouvons dire que l'association de la corrélation des contours horizontaux et verticaux obtenus à partir d'un simple filtrage par convolution permet d'atteindre des résultats proches de ceux obtenus en utilisant les corrections d'illumination relativement complexes telles que l'égalisation d'histogramme et la correction du gradient d'illumination. De plus, on remarque que si la corrélation des contours horizontaux donne des résultats satisfaisants, généralement supérieurs à ceux de la corrélation normée centrée des contours obtenus par la méthode de Sobel, ce n'est pas le cas de la corrélation des contours verticaux. L'utilisation d'autres filtres pourrait ainsi permettre d'améliorer les résultats de ce système de détection.

Dans la prochaine section, nous proposons une méthode permettant de déterminer des filtres adaptés aux objets que nous souhaitons détecter.



(a) 5 exemples, base (a)

(b) 10 exemples, base (a)+(b)

(c) 10 exemples, base (a)+(d)

(d) 20 exemples, base (i ... l)

(e) 40 exemples, base (i ... p)

(f) 80 exemples, base (a ... p)

FIGURE 3.25 – Courbes Rappel Précision du système de détection par association de corrélations croisées des images de contours horizontaux et verticaux. Nous observons que cette association permet d'obtenir un système de détection robuste aux variations d'illumination en donnant des résultats proches de ceux obtenus par l'utilisation des traitements permettant, dans une certaine mesure, de corriger ces variations.

3.6.3 Analyse en Composantes Principales Convolutionelle

Dans cette section, nous proposons une méthode basée sur la PCA permettant de déterminer des filtres correspondant aux formes les plus présentes dans les images de l'objet recherché. Ainsi, si l'on trouve de nombreux contours verticaux cette méthode privilégiera la détermination d'un filtre détecteur de contours verticaux. Cette méthode que nous avons nommé Analyse en Composantes Principales Convolutionelle (C-PCA pour Convolutional Principal Composant Analysis) est une généralisation de la PCA classique et de la 2D-PCA (Two Dimensional Principal Component Analysis) [129]. Bien que nous utilisions la C-PCA pour déterminer des filtres, cette méthode comme la PCA classique ou la 2D-PCA peut aussi être utilisée afin de compresser l'information contenue dans une image. Dans les sections suivantes, nous proposons une description de la méthode de la 2D-PCA avant de décrire la C-PCA. Enfin, nous décrirons l'utilisation de la C-PCA pour déterminer des filtres que nous utiliserons en lieu et place de ceux utilisés par la méthode de Sobel.

3.6.3.1 Analyse en Composantes Principales 2D

L'analyse en composantes principales 2D (2D-PCA) a été introduite par Yang *et al* [129] dans le cadre de la reconnaissance de visages. Le but était d'introduire une méthode de représentation inspirée de la PCA mais tenant compte de l'aspect bidimensionnel des images. La PCA a été très largement utilisée en reconnaissance de visages [90, 41, 24] car elle permet par une simple projection linéaire de très nettement réduire la dimension des vecteurs représentant les images de visages. Cependant, la PCA appliquée à la représentation d'images pose plusieurs difficultés. La première est que les matrices représentant les images doivent être transformées en des vecteurs monodimensionnels de très grande taille. Ainsi, un très grand nombre d'images exemples est nécessaire pour obtenir une évaluation précise de la matrice de covariance et par voie de conséquence, des vecteurs propres représentant l'espace de projection des images. Contrairement à la PCA classique, la 2D-PCA est basée sur l'utilisation de matrice à la place de vecteurs monodimensionnels. Ainsi, il n'est plus nécessaire de transformer l'image originale en vecteur. De plus, les dimensions de la matrice de covariance s'en trouvent très nettement réduites, permettant outre une extraction des vecteurs propres bien plus rapide que pour la PCA, une évaluation précise de ces derniers avec un nombre bien plus réduit d'images exemples.

Pour ce faire la 2D-PCA suit le même raisonnement que la PCA classique qui détermine le sous espace orthogonal de projection de dimension k qui minimise la distance L2 moyenne entre les images originales et les images projetées. Cependant, dans le cas de la 2D-PCA, c'est la matrice A de dimension $m \times n$ représentant l'image qui est projetée dans le sous-espace. Ainsi si $\mathbf{v}$ est un vecteur de la base de projection, alors $\mathbf{s}$ le résultat de la projection s'écrit :

$$\mathbf{s} = A\mathbf{v} \tag{3.36}$$

Ainsi le vecteur $\mathbf{v}$ est de dimension n et le vecteur $\mathbf{s}$ de dimension m. Si nous disposons d'une base de N images représentées par les matrices $\{A_1, \ldots, A_j, \ldots, A_N\}$.

Alors, trouver le vecteur $\mathbf{v}$ qui minimise la distance entre les images originales et les images projetées revient à maximiser le critère $J(\mathbf{v})$ suivant :

$$J(\mathbf{v}) = tr(\mathbf{v}^T \Sigma \mathbf{v}) \tag{3.37}$$

$$\Sigma = \frac{1}{N} \sum_{j=1}^{N} (A_j - \bar{A})^T (A_j - \bar{A}) \tag{3.38}$$

$$\bar{A} = \frac{1}{N} \sum_{j=1}^{N} A_j \tag{3.39}$$

$J(\mathbf{v})$ est la trace de la matrice de covariance des vecteurs $\mathbf{s}$. Σ est la matrice de covariance des matrices A et $\bar{A}$ est l'image moyenne de l'ensemble des N images utilisées. Le vecteur $\mathbf{v}$ qui maximise J est le vecteur propre qui correspond à la plus grande valeur propre de la matrice de covariance Σ [128]. Un seul vecteur de projection est généralement insuffisant. Ainsi comme pour la PCA, le sous-espace de projection est formé des k vecteurs propres de la matrice de covariance correspondant aux plus grandes valeurs propres. Le sous-espace de projection W s'écrit alors :

$$W = [\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}] \tag{3.40}$$

L'image est alors décrite par la matrice :

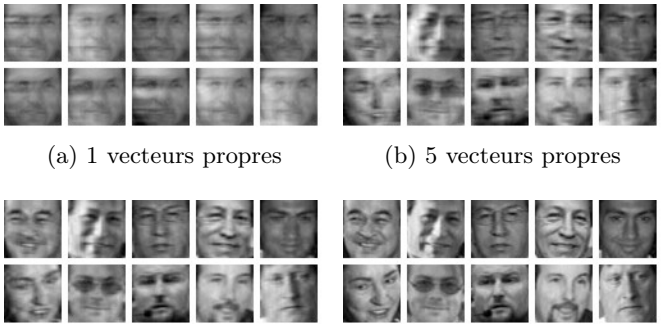
$$S = (A - \bar{A})W \tag{3.41}$$

Reconstruction d'image à partir de la 2D-PCA : La matrice S obtenue par la projection de l'image A dans le sous-espace W permet de décrire l'image originale A de dimension $m \times n$ par la matrice S de dimension $m \times k$ inférieure. A l'inverse, il est possible connaissant la matrice S, le sous-espace W et l'image moyenne $\bar{A}$ de reconstruire plus ou moins précisément en fonction de k l'image originale. Si $\tilde{A}$ est l'image reconstruite, alors, la base W étant orthonormale, nous pouvons écrire :

$$\tilde{A} = SW^T + \bar{A} \tag{3.42}$$

Dans le cas ou $k = n$ alors $\tilde{A} = A$. Plus la valeur de k diminue, plus la reconstruction de $\tilde{A}$ est approximative. Afin de mesurer l'efficacité de la 2D-PCA pour décrire et reconstruire des images, nous avons utilisé une base de données de 80 images de visages (figure : 3.1) de dimension 45×45 afin de calculer la matrice de covariance de la 2D-PCA et de déterminer le sous-espace W de projection. Nous avons ensuite reconstruit des images de visages avec différents nombres de vecteurs propres k (figure : 3.26). Si l'utilisation d'un seul vecteur propre permet de reconnaître que nous avons affaire à des visages, la reconstruction est très approximative et les visages ne sont pas identifiables. Dix vecteurs propres permettent de reconnaître les visages et vingt vecteurs entraînent une reconstruction presque parfaite.

Afin de comparer la 2D-PCA à la PCA, nous avons effectué les mêmes expérimentations pour la PCA (figure : 3.27). Nous constatons que la 2D-PCA nécessite

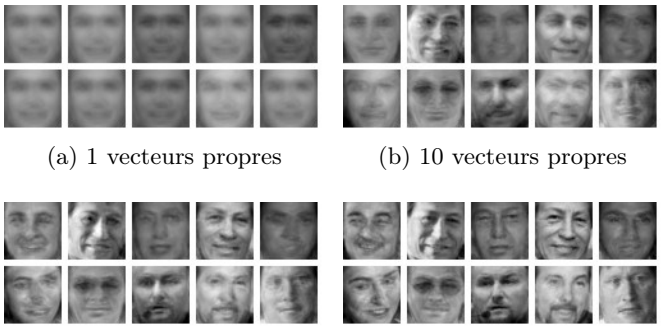


(a) 1 vecteurs propres (b) 5 vecteurs propres

(c) 10 vecteurs propres (d) 20 vecteurs propres

FIGURE 3.26 – Reconstruction d'images avec la 2D-PCA. La base de projection W est calculée à partir de 80 images de dimension 45×45. Une reconstruction parfaite nécessite donc 45 vecteurs propres. Un seul vecteur propre nous permet de reconnaître que nous avons affaire à un visage. Vingt vecteurs propres permettent une reconstruction presque parfaite des images originales.

moins de vecteurs propres pour obtenir une reconstruction avec un niveau de qualité équivalent. Cependant, les erreurs de reconstruction sont assez différentes entre la 2D-PCA et la PCA et il est ainsi difficile de comparer la qualité de reconstruction d'une image par rapport à l'autre.



(a) 1 vecteurs propres (b) 10 vecteurs propres

(c) 20 vecteurs propres (d) 40 vecteurs propres

FIGURE 3.27 – Reconstruction d'images avec la PCA. La base de projection W est calculée à partir de 80 images de dimension 45×45. Une reconstruction parfaite nécessite donc $45 \times 45 = 2025$ vecteurs propres. Une quarantaine de vecteurs propres est nécessaire pour une très bonne reconstruction des images de visages. Une vingtaine de vecteurs propres permet de reconnaître les visages mais entraîne des images relativement floues.

Afin de comparer la qualité de reconstruction de la 2D-PCA et de la PCA, nous avons utilisé une mesure d'erreur de reconstruction simple qui consiste à calculer le ratio entre l'énergie de l'image originale moins l'image reconstruite et l'énergie de

l'image originale. Ainsi, si A est la matrice représentant l'image originale et $\tilde{A}$ celle représentant l'image reconstruite, alors l'erreur de reconstruction E_c se calcule ainsi :

$$En(A) = \sum_{i=j=0}^{i<m,j<n} A(i,j)^2 \tag{3.43}$$

$$E_c = \frac{En(A - \tilde{A})}{En(A)} \tag{3.44}$$

La figure 3.28 nous montre l'erreur de reconstruction E_c de la PCA et de la 2D-PCA en fonction du nombre de vecteurs propres. Nous mesurons ainsi ce que nous avons pu observer sur les figures 3.26 et 3.27, c'est à dire, que pour l'utilisation d'un même nombre de vecteurs propres, l'erreur de reconstruction pour la 2D-PCA est nettement inférieure à celle de la PCA classique.

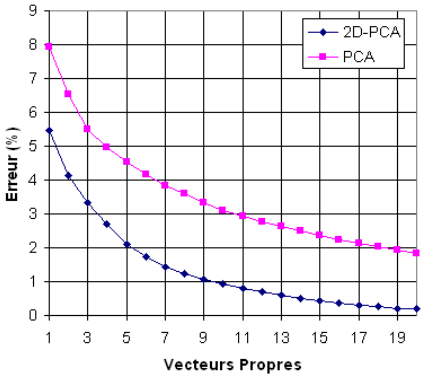


FIGURE 3.28 – Erreur de reconstruction de la PCA et de la 2D-PCA en fonction de la dimension du sous-espace de projection.

Si la 2D-PCA permet de représenter une image à partir d'un sous-espace de dimension plus réduite que la PCA classique, la 2D-PCA reste bien moins performante que la PCA pour la compression d'images. Pour la PCA, le nombre de coefficients permettant de décrire une image est égal au nombre k de vecteurs propres utilisés. Pour la 2D-PCA, le nombre de coefficients nécessaires pour décrire une image de dimension $m \times n = 45 \times 45$ est très supérieur. En effet, A chaque vecteur propre est associé un vecteur $\mathbf{s}$ de m coefficients permettant de décrire chacune des m lignes des images. La projection par la méthode de la 2D-PCA dans le sous-espace W de dimension k entraîne donc $m \times k$ coefficients nécessaires à la description de l'image.

Ainsi, pour atteindre un taux d'erreur de reconstruction de 2%, la 2D-PCA doit utiliser $k = 5$ vecteurs propres et la PCA, $k = 18$ vecteurs propres. Cependant, l'image est alors décrite par 18 coefficients pour la PCA alors que $k \times m = 5 \times 45 = 225$ coefficients sont nécessaires pour la 2D-PCA.

3.6.3.2 Généralisation de l'Analyse en Composantes Principales 2D

Nous avons dans la section précédente brièvement décrit le fonctionnement de la 2D-PCA. Il existe cependant une autre manière d'appréhender cette méthode. En effet, si l'on observe la matrice de covariance Σ de la 2D-PCA (équation 3.38), on remarque qu'elle correspond à la matrice de covariance des lignes de l'ensemble des images exemples centrées (images originales moins l'image moyenne). Si l'on pose $\mathbf{l_{ij}}$ le vecteur correspondant à la i^{ieme} ligne de la j^{ieme} image centrée $(A_j - \bar{A})$ alors :

$$\Sigma = \frac{1}{N}\sum_{j=1}^{N}(A_j - \bar{A})^T(A_j - \bar{A}) \tag{3.45}$$

$$= \frac{1}{N \times m}\sum_{j=1}^{N}\sum_{i=1}^{m}\mathbf{l_{ij}}\mathbf{l_{ij}}^T \tag{3.46}$$

Ainsi, la base de projection W de la 2D-PCA correspond au sous-espace orthogonal obtenu en effectuant l'Analyse en Composantes Principales de l'ensemble des lignes des images centrées. Chaque ligne $\mathbf{l_i}$ des images exemples est projetée dans le même sous espace et est décrite par le vecteur $\mathbf{s_i}$.

$$\mathbf{s_i} = W^T\mathbf{l_i} \tag{3.47}$$

De même, on observe que reconstruire une image par la méthode de la 2D-PCA correspond à reconstruire chaque ligne de cette image de la même façon que pour une PCA classique :

$$\tilde{\mathbf{l_i}} = W\mathbf{s_i} \tag{3.48}$$

Afin de généraliser cette méthode, nous proposons non pas de décrire seulement l'ensemble des lignes d'une image en les projetant dans le même sous-espace orthogonal W, mais de décrire tout rectangle de dimension $(p \times q)$ appartenant à une image A. Ainsi, si nous choisissons $p = 1$ et $q = n$ nous effectuons une 2D-PCA, alors que si nous choisissons $p = m$ et $q = n$ nous effectuons une PCA classique.

Description de la C-PCA : La C-PCA consiste à calculer le sous-espace orthogonal W qui puisse décrire au mieux l'ensemble des imagettes ou régions de dimension $p \times q$ des N images exemples. Pour ce faire, on extrait de chaque image A_j l'ensemble des régions de dimension $p \times q$ que l'on exprime sous forme de vecteurs $\mathbf{x_j}$ et l'on en calcule la matrice de covariance Σ (figure : 3.29). Si nous prenons le cas particulier de nos 80 images de visages de dimension 45×45 et que nous décidons d'appliquer la C-PCA avec $p = q = 5$, alors, chaque image exemple A_j génère $(m - p + 1) \times (n - q + 1) = 1681$ vecteurs $\mathbf{x_j^{uv}}$, (u, v) représentant les coordonnées supérieures gauches des régions X_j^{uv} extraites dans l'image exemple A_j. La matrice de covariance Σ se calcule ainsi :

$$\Sigma = \sum_{j=1}^{N} \sum_{u,v} \mathbf{x_j^{uv}} \mathbf{x_j^{uv}}^T \tag{3.49}$$

Si $p = q = 5$ alors, la matrice Σ est de dimension 25×25. Chaque vecteur $\mathbf{x}$ de $p \times q = 25$ éléments peut alors être décrit par un vecteur $\mathbf{s}$ de k éléments, projection du vecteur $\mathbf{x}$ dans le sous-espace orthogonal W^T formé des k premiers vecteurs propres de Σ.

$$\mathbf{s} = W^T \mathbf{x} \tag{3.50}$$

Nous avons nommé cette méthode PCA Convolutionelle car les descripteurs d'une images A sont alors extraits par convolution. En effet, les éléments s_i du vecteur $\mathbf{s}$ sont le produit scalaire du i^{ieme} vecteur propre avec le vecteur $\mathbf{x_j^{uv}}$ extrait de l'image A_j. Ceci est équivalent à corréler l'imagette X_j^{uv} avec l'image reconstituée V_i du vecteur propre $\mathbf{v_i}$. Ainsi, obtenir la valeur de s_i pour toute coordonnée (u, v) de l'image A revient à convoluer l'image originale avec le filtre V_i. L'image résultante S_i contenant alors en tout point (u, v) la valeur de s_i correspondante.

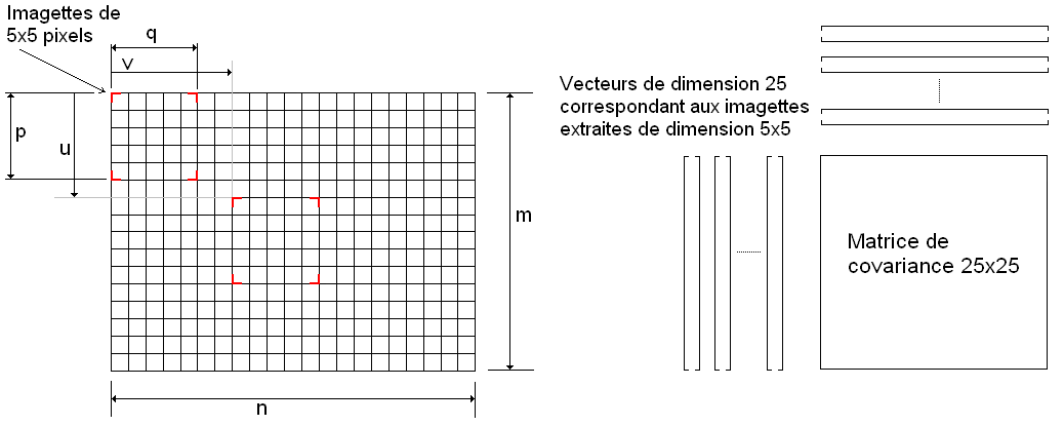


FIGURE 3.29 – Détermination de la matrice de covariance par la méthode de la C-PCA pour des imagettes de dimension $p = q = 5$. Chaque imagette définie par ses coordonnées (u, v) est extraite sous forme de vecteur afin de déterminer la matrice de covariance Σ.

Afin d'illustrer la méthode, nous avons appliquée la C-PCA à la base d'images de 80 visages utilisées précédemment (figure : 3.1). Nous avons choisi les variables $p = q = 5$. La figure 3.30 montre les quatre filtres V_i correspondant aux quatre premiers vecteurs propres obtenus avec la C-PCA. Puis, nous avons appliqué ces filtres à six images de visages représentant ainsi la projection de chaque imagette de dimension 5×5 sur un vecteur propre.

C-PCA et reconstruction d'images : La C-PCA permet de décrire des images par un simple filtrage par convolution en projetant chaque imagette de dimension

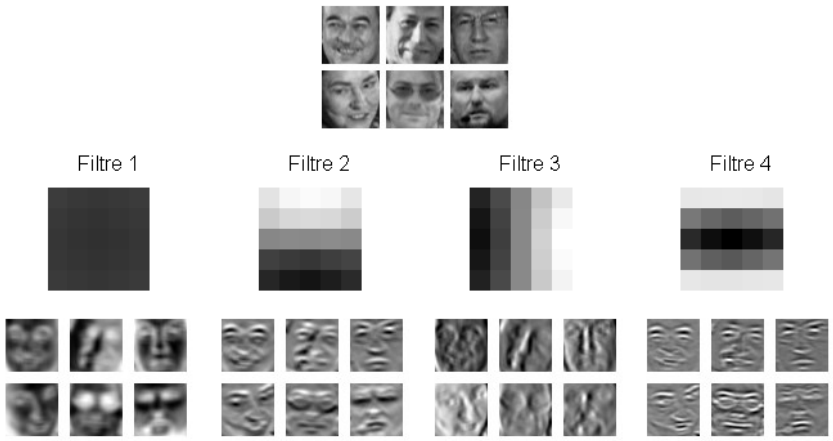


FIGURE 3.30 – Exemples des quatre premiers filtres 5×5 obtenus par la C-PCA, calculés à partir d'une base de 80 images de visages. Ces filtres sont appliqués sur six images de visages. Le premier filtre est un filtre moyenneur. Le second et le quatrième filtre sont des détecteurs de contours horizontaux et le troisième un détecteur de contours verticaux.

(p, q) sur une base de vecteurs propres orthogonaux. Comme pour la 2D-PCA il est évident que cette méthode ne réduit pas autant l'espace de description des images que la PCA classique. Cependant, la compression d'images n'est pas le but de la C-PCA. L'idée de la C-PCA est d'apporter une méthode flexible pour décrire des images. Ces descripteurs pouvant alors comme pour la 2D-PCA être utilisés avec des systèmes de classification pour des problèmes comme la reconnaissance de visages, ou dans notre cas, la détection d'objets. Le choix de la dimension $p \times q$ des filtres que nous extrayons avec la C-PCA permet de décider le type d'information que nous souhaitons extraire des images. Afin de montrer l'influence des variables p et q ainsi que le nombre k de vecteurs propres utilisés, nous avons appliqué la C-PCA en utilisant la même base de 80 images exemples que pour la PCA et la 2D-PCA, puis reconstruit les mêmes visages (figure : 3.31).

Si la reconstruction d'images se fait de façon évidente pour la 2D-PCA et la PCA, certains choix sont nécessaires pour reconstruire les images originales à partir de la C-PCA.

En effet, chaque coefficient de la matrice S_i extraite de l'image A_i par filtrage convolutionnel permet de reconstruire une portion de dimension $p \times q$ de l'image originale. Ainsi, si nous utilisons l'ensemble des coefficients extraits par la C-PCA, nous devons alors tenir compte de la superposition des régions de l'image reconstruite. Nous pouvons imaginer de nombreuses solutions à ce problème. La méthode de reconstruction la plus simple consiste à sous-échantillonner l'image S_i. Si nous sous-échantillonnons S_i horizontalement avec un facteur q et verticalement avec un facteur p, alors nous pouvons reconstruire chaque région de l'image originale sans aucune superposition. De cette manière, nous diminuons aussi nettement le nombre de coefficients nécessaire pour décrire l'image originale. Si nous prenons le cas d'une

(a) p=q=3 k=1 (b) p=q=3 k=2 (c) p=q=3 k=3

(d) p=q=5 k=1 (e) p=q=5 k=3 (f) p=q=5 k=5

(g) p=q=15 k=1 (h) p=q=15 k=5 (i) p=q=15 k=15

FIGURE 3.31 – Reconstruction d'images avec la C-PCA. La base de projection W^T est calculée à partir de 80 images de dimension 45×45. Une reconstruction parfaite nécessite $p \times q$ vecteurs propres. On constate que plus la dimension $p \times q$ des filtres utilisés est importante, plus le nombre k de vecteurs propres nécessaires pour décrire une image augmente.

image de $m \times n = 45 \times 45$ pixels divisée en régions de $p \times q = 5 \times 5$ pixels non superposées, décrites par projection sur $k = 5$ vecteurs propres orthogonaux, alors le nombre de coefficients nécessaires à la reconstruction de l'image avec la méthode de la C-PCA est égal à $\left(\frac{m}{p}\right) \times \left(\frac{n}{q}\right) \times k = 9 \times 9 \times 5 = 405$, contre les 2025 coefficients de la matrice A décrivant l'image originale. Lorsque l'on observe les images de visages reconstruites avec la C-CPA, nous pouvons déjà remarquer que lorsque le nombre de vecteurs propres utilisés est insuffisant on observe aisément l'intersection entre chaque région reconstruite. D'une manière générale, on remarque expérimentalement que pour reconstruire correctement une image de visage, nous devons choisir $k \approx \sqrt{(p \times q)}$. Si nous prenons les cas particuliers de la PCA, *i.e*, $p = m$ et $q = n$, alors, nous obtenons $k = 45$ et pour la 2D-PCA $k \approx 7$ ce qui correspond approximativement à ce que nous observons sur les figures 3.27 et 3.26. De cette règle expérimentale, nous pouvons déduire que le nombre de coefficient nécessaire pour reconstruire correctement une image avec la C-PCA est proportionnel à $\frac{1}{\sqrt{pq}}$. Ainsi, cela confirme que dans le cadre de la compression d'images, la PCA donne les meilleurs résultats.

3.6.3.3 Corrélation avec filtres générés par la C-PCA

Si la PCA permet de décrire les images de visages avec un minimum de descripteurs, elle n'est que difficilement utilisable directement pour des problèmes de détections. Le premier obstacle est le nombre et la dimension des vecteurs propres

de la PCA. En effet, nous savons que pour décrire correctement un visage de 45×45 pixels, nous devons utiliser environ 45 vecteurs propres de dimension $45 \times 45 = 2025$. Projeter chaque image à classifier dans un tel sous-espace est possible mais très coûteux en terme de temps de calculs. Etant donné le très grand nombre d'images à classifier pour les problèmes de détection, cette méthode est difficilement applicable. Une autre difficulté de l'utilisation la PCA pour représenter les images dans un problème de détection est que les descripteurs utilisés doivent permettre de différencier les 'objet' des 'non objet', *i.e*, dans notre cas, un visage du reste du monde. Si la PCA permet de décrire avec très peu de descripteurs les images de visages, ce n'est pas le cas des images du 'reste du monde'. Ainsi deux images de visage et de 'non visage' pourraient avoir les mêmes descripteurs dans l'espace vectoriel de la PCA.

La C-PCA utilise un nombre restreint de vecteurs propres de dimension réduite ($p \times q$). De plus, il suffit pour décrire une image avec la C-PCA d'effectuer une convolution avec les filtres V_i de dimension $p \times q$ de la C-PCA. Nous avons vu dans les sections précédentes que si nous associons les résultats de corrélations normées centrées d'images ayant subit différents traitements, nous améliorons nettement les taux détections. Nous proposons dans cette section d'utiliser les filtres obtenus par la C-PCA en lieu et place des filtres de Sobel que nous avons arbitrairement choisis. Les filtres de la C-PCA possèdent en effet toutes les propriétés que nous recherchons. La première est l'orthogonalité des filtres qui nous permet d'extraire pour chaque filtre une information différente de l'image en Niveaux de Gris. La seconde est que les filtres utilisés correspondent aux formes prédominantes présentes dans les régions de dimension $p \times q$ dans la base d'images exemples.

Comme nous souhaitons nous focaliser sur des filtres extracteurs de contours, nous n'utiliserons pas le premier filtre extrait par la C-PCA qui est un simple filtre moyenneur. En effet, il est évident que la forme la plus répandue est l'image homogène avec très peu de variations. Les autres vecteurs propres correspondent alors aux variations de forme les plus présentes dans l'image, comme les contours horizontaux et verticaux. A titre d'exemple, la figure 3.32, présente les cinq premiers filtres correspondant aux cinq premiers vecteurs propres de la C-PCA pour la base de 80 images de visages (figure : 3.1) pour $p = q =$ 3,4,5 et 6.

Nous pouvons constater que les seconds et troisièmes filtres extraient respectivement les contours horizontaux et verticaux des images. Si nous regardons le cas des filtres 3×3 nous obtenons des filtres aux formes très proches des filtres de Sobel que nous avons précédemment utilisés.

Afin de vérifier l'utilité de tels filtres pour notre système de détection, nous avons remplacé les filtres de Sobel par les filtres de la C-PCA pour différentes valeurs de p et q. Nous avons commencé par les valeurs $p = q = 3$ afin de comparer les résultats de la C-PCA pour des filtres de même dimension que les filtres de Sobel (figure : 3.33).

Nous pouvons alors vérifier que les résultats avec le second et troisième filtres de la C-PCA sont très proches de ceux avec le filtre horizontal et le filtre vertical de Sobel. La combinaison des filtres donne aussi des résultats très légèrement inférieurs mais comparables à ceux obtenus avec Sobel.

Nous avons ensuite comparé la combinaison des seconds et troisièmes filtres de

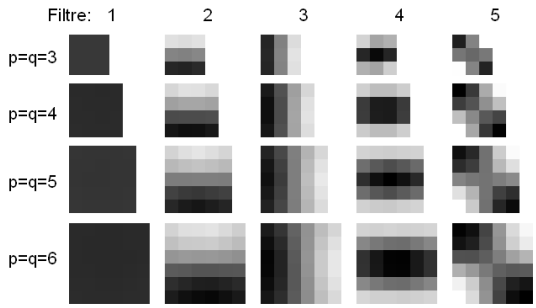
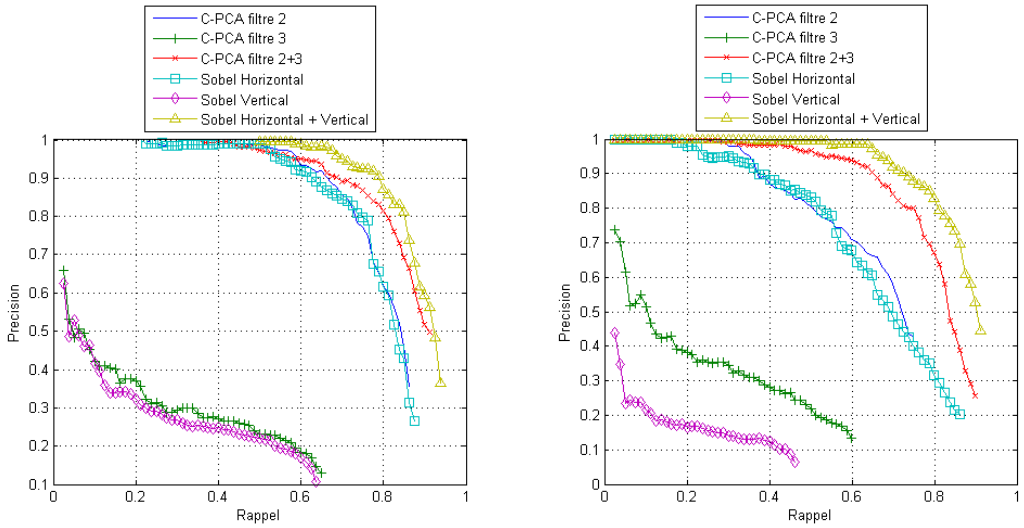


FIGURE 3.32 – Cinq premiers filtres extraits par la C-PCA pour différentes valeurs de p et q. On remarque que nous obtenons les mêmes formes pour les filtres, indépendamment de la dimension des régions extraites. Le premier filtre est un filtre moyenneur, le deuxième et le quatrième filtre extraient des contours horizontaux. Le troisième filtre extrait les contours verticaux et le dernier filtre correspond aux contours diagonaux.

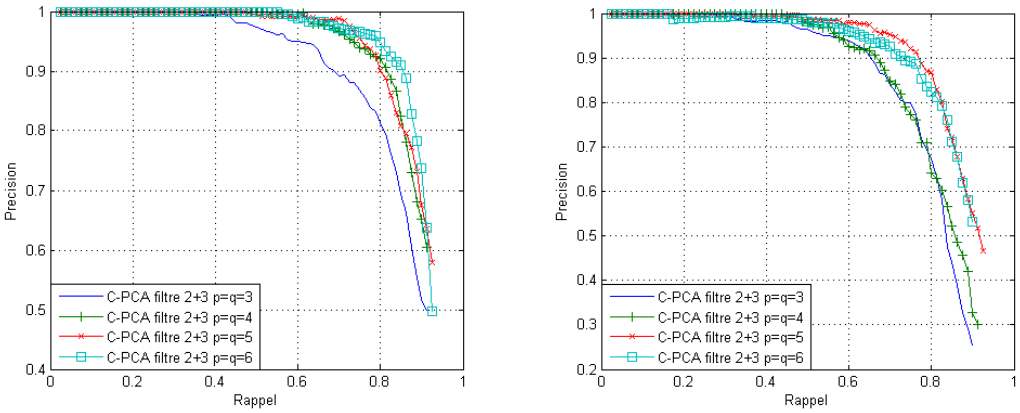


(a) 5 exemples, base (a)

(b) 10 exemples, base (a)+(b)

FIGURE 3.33 – Comparaison des filtres de dimension 3×3 de la C-PCA aux filtres de Sobel pour la détection.

la C-PCA pour des valeurs de $p = q = \{3, 4, 5, 6\}$ (figure : 3.35). L'influence de la dimension des filtres semble assez limité, les résultats sont comparables pour toutes les dimensions, une dimensions de $p = q = 5$ donne des résultats légèrement supérieurs à ceux avec $p = q = 3$ et 4 et équivalents à ceux obtenus avec des dimensions supérieures. Le temps nécessaire au filtrage d'une image étant proportionnels à $p \times q$, le meilleur rapport taux de détection sur puissance de calcul est obtenu pour des filtres de la C-PCA de dimension 5×5.



(a) 5 exemples, base (a)

(b) 10 exemples, base (a)+(b)

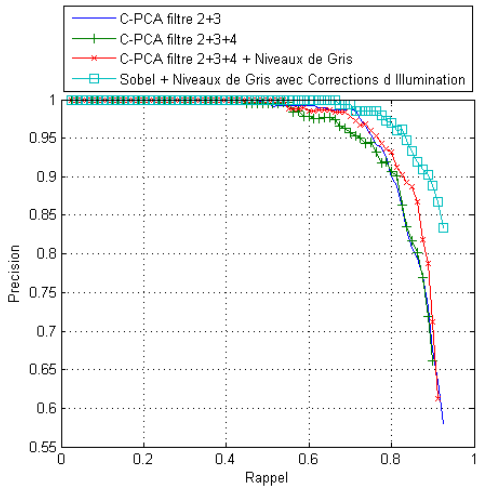
FIGURE 3.34 – Comparaison des filtres de dimension $p = q = 3,4,5,6$ de la C-PCA appliquée au système de détection. Les filtres de dimension $p = q = 5$ sont un bon compromis entre la puissance de calcul nécessaire et les taux de détection obtenus

Enfin, nous avons utilisé les filtres 5×5 de la C-PCA avec différentes bases d'exemples dans le but de mesurer l'efficacité et la sensibilité au nombre d'exemples du détecteur basé sur la C-PCA, de mesurer l'influence du nombre de filtres de la C-PCA sur les taux de détection et enfin, de permettre de comparer les résultats aux autres méthodes basées sur la corrélation (figure : 3.35).

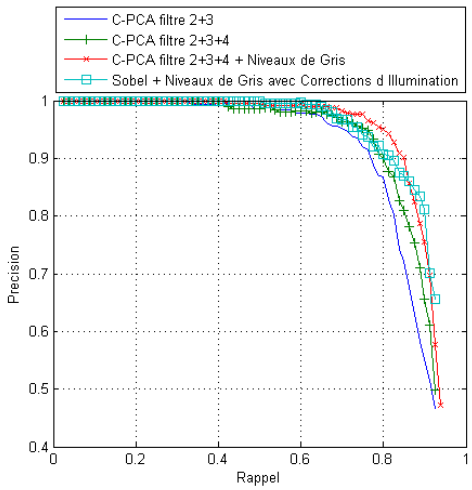
Nous constatons d'abord que plus nous utilisons de filtres de la C-PCA, meilleurs sont les résultats, et ceci quel que soit le nombre d'exemples. L'utilisation des images en Niveaux de Gris en plus des images filtrées permet une légère amélioration. L'utilisation de seulement deux filtres donne des résultats assez variables en fonction de la base d'exemples. L'utilisation d'un troisième filtre permet de rendre le système beaucoup plus robuste à la base utilisée. En effet, lorsque les résultats avec deux filtres sont assez faibles (figure : 3.35e, 3.35) le troisième filtre permet une très nette amélioration des résultats. Lorsque les résultats avec deux filtres sont plus proches de ceux obtenus avec les systèmes de détection précédents, l'apport du troisième filtre sur les taux de détection est bien plus faible. Enfin, l'utilisation de l'image en Niveaux de Gris, qui nous emmène donc à utiliser une combinaison de quatre mesures de similarité, obtient les meilleurs résultats que nous ayons obtenus avec les bases de 40 et 80 exemples.

Au final, nous pouvons dire que l'utilisation des filtres générés par la C-PCA permet d'obtenir un système de détection basé sur la corrélation, particulièrement robuste à la base d'exemples utilisée, et ceci sans effectuer de traitements d'image complexes comme une déformation affine ou des corrections d'illumination. Les filtres générés étant assez proches de ceux utilisés par Sobel, nous obtenons des résultats similaires, avec l'avantages que nous ne sommes pas limités à seulement deux filtres pour la C-PCA, ce qui nous permet une légère amélioration des résultats par rapport

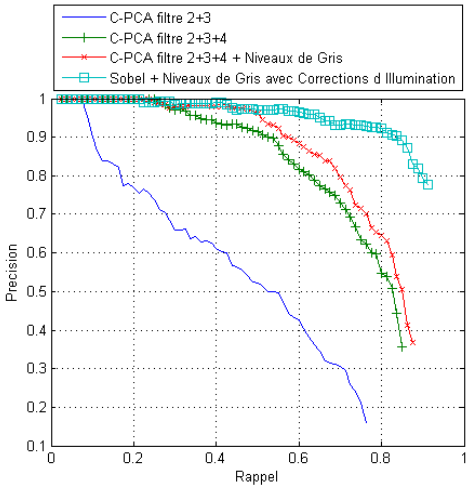
à l'utilisation des filtres de Sobel.



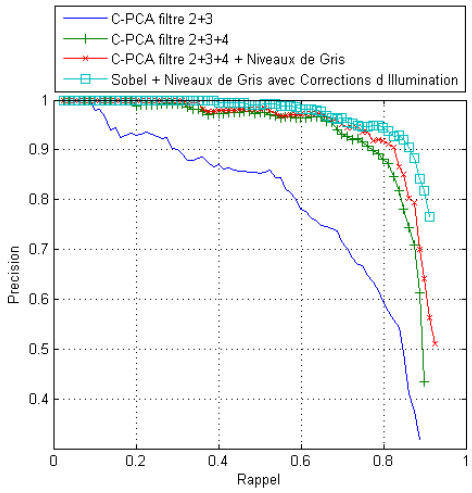
(a) 5 exemples, base (a)



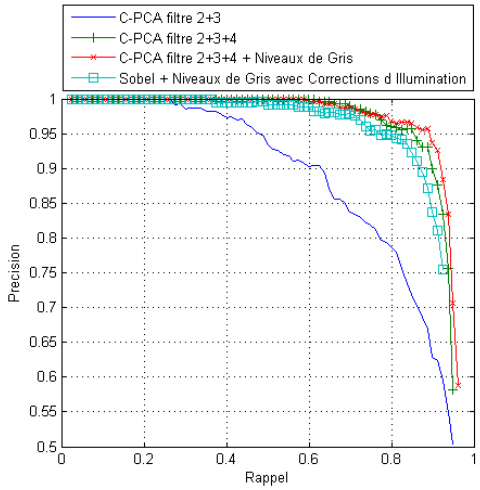
(b) 10 exemples, base (a)+(b)



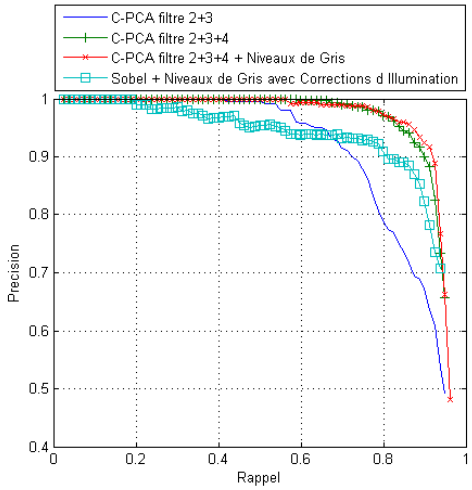
(c) 10 exemples, base (a)+(d)



(d) 20 exemples, base (i ... l)



(e) 40 exemples, base (i ... p)



(f) 80 exemples, base (a ... p)

FIGURE 3.35 – Courbes Rappel Précision du système de détection par association de corrélations croisées des images filtrées par les filtres de la C-PCA. Nous observons que cette association permet d'obtenir un système de détection fonctionnel avec très peu d'exemples.

3.7 Evaluation du système de détection sur la base de test CMU

De part la complexité de la tâche consistant à détecter des visages avec une mesure de similarité basée sur la corrélation, nous avons choisi d'évaluer notre système de détection sur une base de test plus simple que les bases de test standards. Cependant, nous avons montré qu'en associant plusieurs mesures de similarité basées sur la corrélation, nous pouvons obtenir d'assez bons résultats. Nous proposons dans cette section d'appliquer les systèmes de détection basés sur l'association de corrélations à la base de test la plus couramment utilisée en détection de visages, *i.e*, CMU.

Dans ce chapitre nous pouvons distinguer trois méthodes qui ont permis d'obtenir des résultats satisfaisants.

- La première méthode que nous nommerons 'Association 1' présentée dans la section 3.5 est l'association de la corrélation d'images de contours extraits par la méthode de Sobel et d'images en Niveaux de Gris traitées de façon à minimiser l'effet des variations d'illumination.
- La seconde méthode que nous nommerons 'Association 2' présentée dans la section 3.6.2 utilise les filtres de Sobel verticaux et horizontaux afin d'associer la corrélation des contours verticaux, horizontaux et les Niveaux de Gris des images.
- La troisième méthode que nous nommerons 'Association 3' présentée dans la section 3.6.3.3 remplace les filtres de Sobel par les seconds, troisièmes et quatrièmes filtres déterminés à partir de la méthode de la C-PCA.

La première difficulté est que pour effectuer ces tests, notre système de détection doit être capable de détecter des visages de dimension 25×25 alors que nous nous sommes jusqu'à présent limités à des images exemples de dimension 44×44. Nous avons vu dans les sections 3.3.3 et 3.4.2 que plus la dimension des images exemples est réduite, plus les taux de détection sont faibles. Nous avons ensuite fixé la dimension des images exemples à 44×44 pour les expérimentations sur les systèmes de détection basés sur l'association de corrélations. C'est pourquoi, nous commençons dans cette section par évaluer sur la base de test 'Face 1999', l'influence de la diminution de la dimension à 25×25 pixels des images de référence. Nous utilisons une base de 5 et 80 exemples afin de déterminer le comportement du système avec très peu d'exemples puis avec un nombre important. Les systèmes 'Association 2' et 'Association 3' sont utilisés exactement de la même manière que précédemment en ne changeant que la dimension des images exemples. Pour le système 'Association 1', nous modifions la valeur du seuil de prédétection s_1. En effet, nous avons vu que pour cette méthode ce seuil doit être réglé de manière assez précise. Il était égal à 0.72 pour les exemples de 44×44 pixels et est réglé à une valeur de 0.79 pour les images de 25×25 pixels.

Nous constatons sur la figure 3.36 que si les trois systèmes testés donnaient des résultats du même ordre pour des images exemples de 44×44 pixels, ceci n'est pas le cas avec les images exemples de tailles plus réduites (25×25). En effet, les systèmes basés sur l'association de filtres détecteurs de contours orientés 'Association 2 et 3'

donnent des résultats supérieurs au système 'Association 1' basé sur la corrélation d'images traitées par l'algorithme de Sobel et des images en Niveaux de Gris traitées de manière à minimiser l'effet des variations d'illumination. Ainsi, l'association de plusieurs mesures de similarité basées sur la corrélation et des filtres convolutionnels permet d'obtenir des résultats supérieurs aux systèmes basés sur des traitements d'image plus complexes tels que la correction du gradient d'illumination ou l'égalisation d'histogramme.

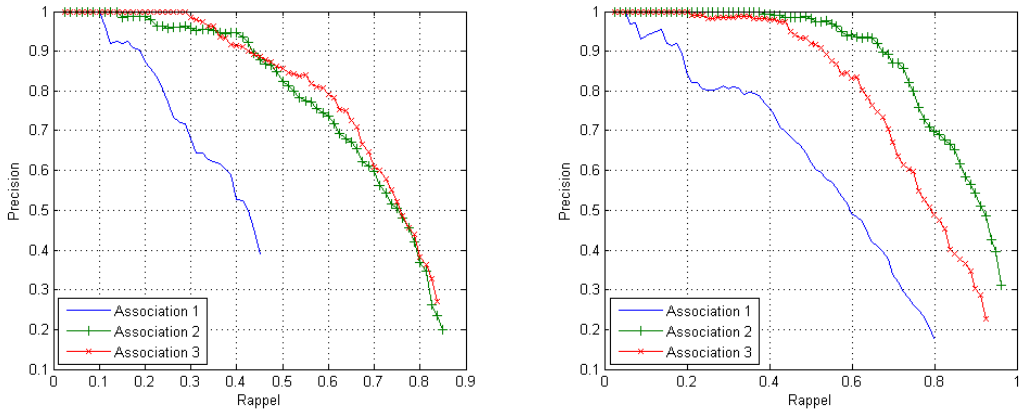
Dans une certaine mesure, nous pouvons rapprocher un tel résultat avec le fonctionnement des réseaux de neurones convolutionnels. En effet, le système de détection de Garcia et Delakis [39] effectue sur la première couche du réseau de neurones plusieurs filtrages convolutionnels avec des filtres de dimension 5×5. Ce système est le système de détection de visages le plus performant tout en étant le seul système de détection qui ne nécessite aucun prétraitement d'image corrigeant les variations d'illumination.

Nous pouvons aussi constater que les taux de détection sont très nettement inférieurs à ceux obtenus avec des images exemples de 44×44 pixels (figures : 3.19, 3.25 et 3.35). Il semble donc difficile d'atteindre les même taux de détection que l'état de l'art avec une mesure de similarité basée sur la corrélation.

Lorsque nous appliquons les trois systèmes de détection à la base de visages CMU, nous pouvons déjà constater, que comme sur la base 'Face 1999', les systèmes 'Association 2 et 3' sont plus performants que le système 'Association 1'. La principale différence entre les résultats obtenus sur les deux bases de test est l'influence du nombre d'exemples de la base de référence. En effet, sur la base de test 'Face 1999' il est presque équivalent d'utiliser 5 exemples (figure : 3.37a) ou 80 exemples (figure : 3.37b). Sur la base de test CMU, les résultats sont nettement inférieurs avec la base de référence de 5 exemples comparés à ceux obtenus avec la base de 80 exemples. Ceci s'explique par la plus grande diversité des visages présents dans la base CMU qui ne peuvent être correctement représentés par seulement 5 exemples.

L'intérêt de la base de test CMU est qu'elle nous permet de nous comparer aux systèmes de détection de l'état de l'art. La plupart de ces méthodes ne donnent que le Rappel pour un nombre donné de fausses détections. Nous reportons dans le tableau 3.1 les résultats publiés pour les principaux systèmes de l'état de l'art, ainsi que le nombre d'exemples utilisés dans la phase d'apprentissage et incluons les résultats obtenus par nos trois systèmes de détection basés sur la corrélation. Nos résultats sont très loin de ceux de l'état de l'art. Cependant, il faut noter que nous utilisons beaucoup moins d'exemples d'apprentissage que ces méthodes et aucun classifieur complexe. Ainsi notre système fonctionne avec près de cinquante fois moins d'exemples que les systèmes traditionnels et est capable de détecter des visages de dimension réduite dans des environnements complexes. La figure 3.38 montre des exemples de détections sur la base CMU et met en évidence à la fois, la capacité de la corrélation à détecter des visages dans un environnement complexe, mais aussi les limites des méthodes utilisant la corrélation.

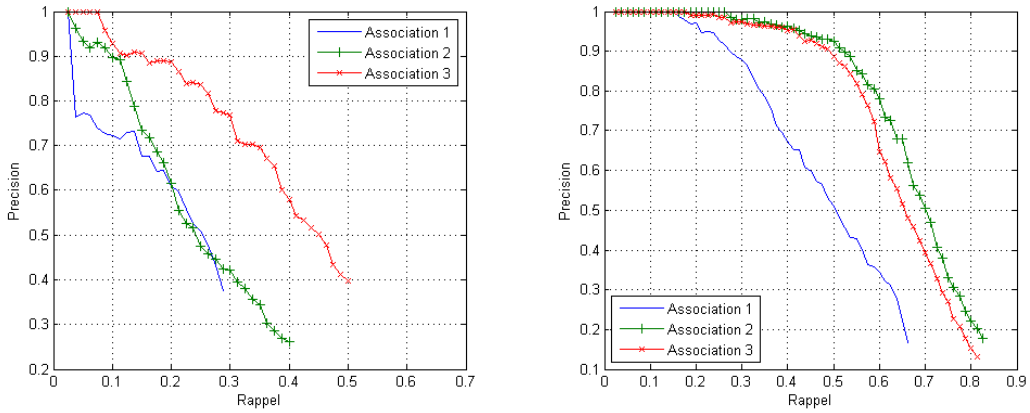
Un autre avantage de la méthode basée sur la corrélation et les plus proches voisins, est que comme nous pouvons le voir sur la figure 3.38, nous somme capables de repérer assez précisément la position de points d'intérêt tels que le nez, les yeux



(a) 5 exemples, base (a)

(b) 80 exemples, base (a ... p)

FIGURE 3.36 – Courbes Rappel Précision des trois systèmes de détection par association de corrélations sur la base 'Face 1999' avec des images exemples de 25×25 pixels.



(a) 5 exemples, base (a)

(b) 80 exemples, base (a ... p)

FIGURE 3.37 – Courbes Rappel Précision des trois systèmes de détection par association de corrélations sur la base CMU avec des images exemples de 25×25 pixels.

ou la bouche. En effet la position de ces points d'intérêt ayant manuellement été annotée sur chaque image exemple, nous pouvons la reporter sur chaque détection correspondante.

Ainsi l'Association de plusieurs corrélations permet d'obtenir des résultat intéressants, en particulier si on les compare à ceux obtenus par la corrélation croisée normée sans association (figure : 3.11) qui ne semblait pas permettre l'application de la corrélation pour détecter des visages sur des bases de test complexes telles que

Système de détection	Fausses détections				
	0	10	31	65	167
Rowley *et al* [42] (4000)	-	83,2%	86,0%	-	90,1%
Viola-Jones [122] (4916)	-	78,3%	85,2%	89,8%	91,8%
CFF [39] (3702)	88,8%	90,5%	91,5%	92,3%	93,1%
Association 1 (80)	14,7%	24,7%	30,7%	35,2%	44,4%
Association 2 (80)	25,2%	40,9%	51,7%	56,2%	64,2%
Association 3 (80)	17,0%	40,1%	48,1%	54,6%	58,7%

TABLE 3.1 – Comparaison du Rappel des systèmes de détection pour divers nombre de fausses détections sur la base de test CMU. Les nombres entre parenthèses correspondent au nombre d'exemple de la base d'apprentissage.

CMU.

Dans le chapitre suivant, nous proposons de garder le principe de l'association de mesures de similarité en remplaçant la corrélation par une mesure plus complexe basée sur les réseaux de neurones qui ont montré leur efficacité en détection d'objets.

FIGURE 3.38 – Exemples de détections effectuées sur la base CMU à partir du système basée sur la C-PCA et l'association de corrélations croisées normées centrées