

Détection de Visages par Méthode Hybride Couleur de Peau et Template Matching

Ch. Bencheriet, A/H. Boualleg & H. Tebbikh

B. Guerzize & W. Belguidoum

LAIG, Université 8 mai 45 de Guelma BP 401, Guelma 24000, Algérie

bencheriet@yahoo.fr

tebbikh@yahoo.com

boualleg_ah@yahoo.fr

Résumé :

Les visages constituent une catégorie de stimulus importante par la richesse des informations qu'ils véhiculent. Ils sont à la fois des vecteurs visuels principaux à l'identité des individus et des vecteurs essentiels de communication (verbale et non verbale). La détection automatique de visages est un problème très important. Du point de vue application, elle est à la base de tous systèmes de reconnaissance de visages, car avant de reconnaître n'importe quelle personne il est indispensable de localiser son visage.

Dans ce document nous présentons, un système de détection de visages à partir des images fixes, notre système est testé, avec une variété d'images contenant plusieurs personnes sans conditions sur l'éclairage et sans conditions sur l'angle de rotation du visage dans l'image. Il est basé sur une méthode hybride, constitué de la méthode de détection de couleur de peau et la méthode du « template matching », s'appuyant sur la création d'un modèle de visages.

Mots clé : Détection des visages, modèle de visages, détection de couleur de peau, template matching.

Introduction

Sans aucune hésitation, l'homme est capable de distinguer les visages de tout autre « objet ». Mieux, il est capable d'en reconnaître des centaines, malgré de très faibles différences structurelles. Mais qu'est ce qu'un visage ? Ceci pourrait constituer un exemple de définition :

« Un visage est une structure tridimensionnelle possédant une configuration « externe » - le contour de la face modelé par les saillies osseuses et souligné par les cheveux où s'inscrit une configuration interne » formé par l'assemblage des traits. Certaines particularités s'y ajoutent comme la pilosité, les lunettes, la texture, la coloration de la peau...etc. ».

La détection de visages dans une image sans hypothèse préalable est un problème très difficile à résoudre en raison de la très grande variabilité de la forme et la couleur à détecter (Orientation et tailles quelconques, éclairage variable, couleur de peau différente). Il existe trois types de modélisation pour détecter les visages : La première dite bas niveau :

consiste à suivre par exemple des objets de teinte chair en mouvement dont on analyse la forme globale « Blob Analysis », la seconde structurelle où des primitives tels que les yeux, le nez et la bouche sont détectés et associés à des modèles de visages. La dernière globale dans laquelle une image est assimilée à un vecteur (luminance, moment, codage). Une base d'exemples est constituée, à partir de laquelle on peut extraire des vecteurs propres « eigen-faces » ou configurer les poids d'un réseau de neurones [5].

Une grande variété de méthodes de détections de visages a été proposée ces dernières années. Cependant, la plupart d'entre elles s'attardent à répondre aux questions "Y a-t-il des visages dans cette image ? Si oui, où sont-ils ?". Sans toutefois extraire les caractéristiques du visage. Les méthodes basées sur des réseaux de neurones se résument à l'utilisation d'un classifieur à deux sorties représentant la présence ou l'absence de l'objet recherché dans une sous-région de l'image. Le principe de base consiste à balayer l'image avec une fenêtre d'attention de dimensions fixes et de réaliser la détection sur les sous-images. Néanmoins, il est encore une fois nécessaire

d'effectuer plusieurs balayages à différentes résolutions pour ainsi réaliser une détection suffisamment robuste. L'inconvénient majeur de cette méthode est que la phase d'apprentissage nécessite un redimensionnement des images pour être compatibles avec les dimensions requises par le nombre d'entrées du réseau aussi cette phase est très gourmande en temps. Un autre problème peut nuire aux résultats des RNA c'est la rotation de la tête et le changement d'échelle.

D'autres méthodes s'intéressent aux parties caractéristiques du visage comme le nez, la bouche et les yeux. Les positions relatives des différentes composantes du visage sont étudiées après avoir été détectées. La difficulté dans cette approche est de traduire par des règles strictes, à définir, la manière dont le chercheur représente le visage. Si ces règles sont trop précises, elles n'identifient pas certains visages (false negative). Dans le cas contraire, elles provoquent de fausses alertes (false positive). Il est évident qu'il est impossible de définir des règles qui tiennent compte de toutes les variabilités comme la position du sujet, par exemple. Le taux de réussite de cette méthode reste faible avec surtout un grand nombre de fausses alertes et ne peut donc se suffire à elle-même. En revanche, elle permet une focalisation rapide sur les parties intéressantes de l'image [5] [1].

Dans cet article, nous proposons une méthode de détection de visage basée sur l'hybridation de deux méthodes la détection de couleur de peau et l'appariement du gabarit appelé communément «template matching». Il est constitué de 4 sections : après une introduction générale à la section I, nous illustrons la méthode adoptée dans la section II. Nous présentons dans la section III quelques exemples des résultats obtenus sur une variété d'images. Nous terminons cet article par une conclusion générale et des perspectives de ce travail; objets de la section IV.

2. Méthodes de détection de visages

De nombreuses techniques de détections existent et diffèrent aussi bien par les approches qu'elles emploient que par les techniques d'apprentissages qu'elles utilisent. L'approche proposée est basée sur l'utilisation de deux méthodes : La méthode de détection de peau pour éliminer les sections n'ayant pas la couleur de la peau et celle du template matching pour éliminer les sections n'ayant pas les caractéristiques d'un visage.

Les principales étapes se résument comme suit :

1. Section des pixels représentant la peau à partir d'une image d'entrée.
2. Filtrage des régions de peau.
3. Détection du visage par appariement du gabarit "template matching".

3. Détection des sections de couleur de peau

Lorsque les images en entrée sont en couleurs, il

est avantageux d'utiliser cette information supplémentaire pour isoler les régions susceptibles de contenir des visages. En effet, plusieurs travaux ont été basés sur la détection de peau. Dans la majorité des cas, la peau est représentée par une portion d'un espace de couleurs particulier. En utilisant les frontières de cette région comme valeurs de seuillage sur une image, il est possible d'extraire les pixels dont la couleur peut s'apparenter à celle de la peau [2].

Il y a plusieurs modèles de couleurs pouvant s'appliquer à la détection de la peau cette variabilité dépend de l'espace de couleur adopté pour la représentation de la chrominance des pixels les plus utilisés sont : l'espace RGB, RGB normalisé, HSV, YCbCr.

Dans cet article nous présentons le cas de l'espace RGB (figure 1).

3.1 Détection de la couleur de peau par l'espace RGB

Le codage RGB, mis au point en 1931 par la Commission Internationale de l'Eclairage (CIE) consiste à représenter l'espace des couleurs à partir de trois rayonnements monochromatiques de couleurs :

- Rouge (de longueur d'onde égale à 700,0 nm),
- Vert (de longueur d'onde égale à 546,1 nm),
- Bleu (de longueur d'onde égale à 435,8 nm).

Etant donné que le codage RGB repose sur trois composantes proposant la même gamme de valeur, on le représente généralement graphiquement par un cube dont chacun des axes correspond à une couleur primaire (figure 1) :

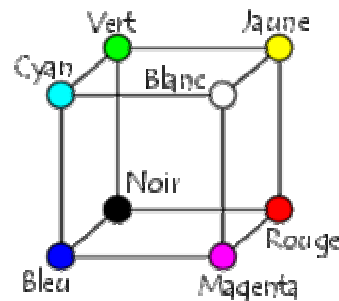


Figure 1 : Présentation des couleurs dans l'espace RGB

3.1.1. Seuillage selon l'espace RGB

La détection de la couleur de peau en utilisant l'espace RGB ne nécessite aucun modèle de peau et aucune transformation des couleurs, elle vérifie simplement une suite des contraintes pour décider si un triplet de couleur (R, G, B) représente une couleur de peau ou non, ces contraintes sont divisées en deux, selon l'état de l'éclairage, fort (jour) ou faible (nuit) [2].

Eclairage fort :

$$(R > 95) \text{ ET } (G > 40) \text{ ET } (B > 20) \text{ ET}$$

$$(\text{MAX}(R, G, B) - \text{MIN}(R, G, B) > 15) \text{ ET}$$

$$(\text{ABS}(R-G) > 15) \text{ ET } (R > G) \text{ ET } (R > B)$$

Eclairage faible :

$((R > B) \text{ ET } (G > B)) \text{ OU } ((R > 220) \text{ ET } (G > 210) \text{ ET } (B > 170) \text{ ET } (\text{ABS}(R-G) \leq 15))$

3.1.2 Segmentation des régions de peau

Puisque les régions de peau sont plus lumineuses que les autres parties des images, ces régions peuvent être segmentées du reste par un seuillage. Ce processus produit une image binaire dont les "1" représentent les Pixels supérieurs à un certain seuil, et les "0" représentent les pixels inférieurs à ce seuil. On peut formuler cette procédure par [4] :

$x < \text{Seuil} \Rightarrow x=0$
 $x \geq \text{Seuil} \Rightarrow x=1$

tel que x représente les valeurs des Pixels de l'image résultat du seuillage dans l'espace RGB.

Le seuil optimal choisi est pris dans l'intervalle 0.35 à 0.45. Un des résultats est illustré dans la figure 3.

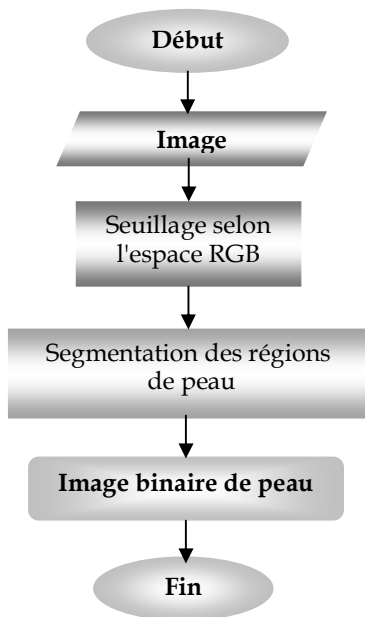


Figure 2 : Illustration du processus de détection de peau



Image originale

Image de peau

Figure 3 : Transformation de l'image originale en image de peau

3.2. Filtrage des régions de peau

En utilisant le résultat de la détection de la peau, nous procédons à l'extraction des régions qui peuvent probablement contenir un visage humain. Ce processus se résume par les étapes suivantes :

- Application des opérateurs morphologiques.
- Etiquetage de l'image binaire.
- Application des contraintes de ratio et des contraintes de surface sur chaque segment.

3.2.1 Filtrage des régions par les opérateurs morphologiques

Les opérations morphologiques que nous avons utilisé sont l'érosion et la dilatation. Le but de ces opérations est d'éliminer les petits segments inutiles [7].

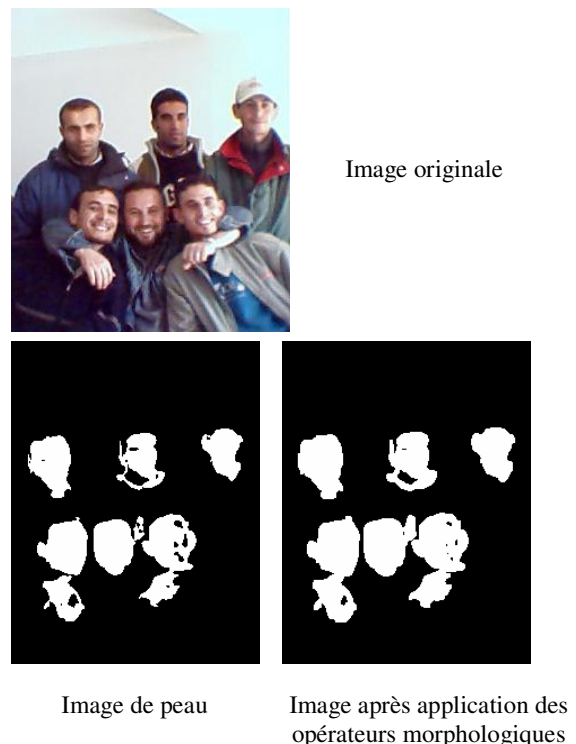


Image de peau

Image après application des opérateurs morphologiques

Figure 4 : Application des opérations morphologiques

3.2.2 Etiquetage des régions

Nous devons d'abord déterminer le nombre de régions de peau dans l'image, en associant à chaque région une valeur entière appelée étiquette. Afin de déterminer l'étiquette d'un Pixel, on utilise pour cela un voisinage d'ordre 8. Le nombre de régions dans l'image segmentée sera donc la valeur de l'étiquette maximale [4].

3.2.3. Extraction des caractéristiques des régions [4]

Cette étape consiste à séparer les régions et en extraire les différentes informations tels que :

- ✓ La position,

- ✓ Les dimensions,
- ✓ La surface,
- ✓ Le rapport entre la hauteur et la largeur "Ratio",
- ✓ Le centre de gravité,
- ✓ La rotation.

a- Séparation des régions

Afin de séparer une région particulière, nous balayons l'image segmentée à la recherche des pixels qui correspondent à l'étiquette de la région concernée. Le résultat sera une image binaire qui ne contient que la région en question.



Image étiquetée

Segment séparé

Figure 5 : Exemple de séparation d'une région

b- Calcul de la position du segment

Il consiste à déterminer les coordonnées (X, Y) du coin supérieur gauche du segment par rapport à l'image segmentée.

c- Calcul des dimensions du segment

En désigne par dimension, la largeur et la hauteur du segment, cela s'effectue en calculant la différence entre les coordonnées (X1, Y1) du coin inférieur droite du segment et ceux (X2, Y2) de son coin supérieur gauche, comme le montre la formule suivante :

$$\text{Largeur} = X1 - X2$$

$$\text{Hauteur} = Y1 - Y2$$

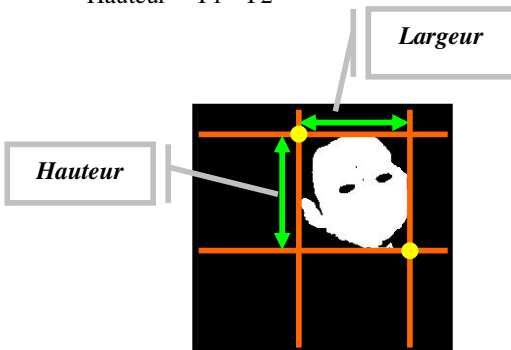


Figure 6 : Calcul des dimensions du segment

d-Calcul de la surface du segment

Il consiste à compter le nombre de pixels appartenant à cette région :

$$\text{Surface} = \sum_i^M \sum_j^N I(i, j)$$

e- Calcul du ratio segment

C'est le rapport entre la largeur et la hauteur de la région. Il est défini comme suit :

$$\text{Ratio} = \text{Hauteur} / \text{Largeur}$$

f- Calcul du centre de gravité du segment

Pour étudier une région, nous devons déterminer son centre. Une manière efficace est de calculer le centre de gravité :

$$\bar{x} = \frac{1}{A} \sum_{i=1}^n \sum_{j=1}^m jB[i, j]$$

$$\bar{y} = \frac{1}{A} \sum_{i=1}^n \sum_{j=1}^m iB[i, j]$$

Où B est la matrice [n x m] qui représente la région et A est la surface de cette région.

g- Calcul de la rotation du segment

Habituellement les visages sont orientés verticalement, cependant on peut avoir une petite inclination θ , qui peut affecter les résultats de l'étape suivante, ce qui nous amène à calculer cette inclination pour rendre droite cette région, par la formule suivante [4] :

$$\theta = \frac{1}{2} \arctan \frac{b}{a - c}$$

Où :

$$a = \sum_{i=1}^n \sum_{j=1}^m (x'_{ij})^2 B[i, j]$$

$$b = 2 \sum_{i=1}^n \sum_{j=1}^m x'_{ij} y'_{ij} B[i, j]$$

$$c = \sum_{i=1}^n \sum_{j=1}^m (y'_{ij})^2 B[i, j]$$

et

$$x' = x - \bar{x}$$

$$y' = y - \bar{y}$$

3.2.4. Sélection des régions

La sélection des régions consiste à balayer toutes les régions détecter, en vérifiant un ensemble de contraintes. Ces contraintes sont aux nombres de deux,

l'une concerne le ratio de la région et l'autre concerne sa surface. Leur vérification est précédée d'une étape préliminaire de traitement de l'image binaire par les opérateurs morphologiques qui vise essentiellement à éliminer le bruit de l'image binaire.

a) Vérification de la contrainte du ratio

Le ratio est le rapport entre la hauteur et la largeur d'un segment, habituellement, dans le cas des visages, ce rapport ne dépasse pas un certain intervalle, donc tous les segments qui ont un ratio qui ne fait pas partie de cet intervalle seront éliminés. L'intervalle convenable utilisé est le suivant : $0.8 < \text{ratio} < 1.8$



Image originale



Image de peau non filtrée

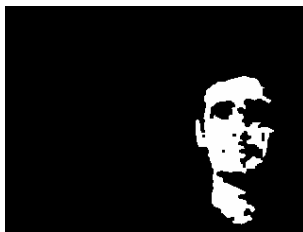


Image de peau filtrée

Figure 7 : Exemple de filtrage avec la contrainte de ratio

b) Vérification de la contrainte de surface

Pour augmenter les performances de la méthode choisie, nous avons introduit une autre contrainte pour éliminer les segments de tailles négligeables, dont il est inutile de les traiter, qui consiste à fixer un seuil minimal et tous les segments qui ont une surface inférieure à cette valeur seront supprimés. Une autre contrainte qui concerne la surface est le rapport entre la surface en question et la surface dont le segment est le plus grand, qui doit être inférieur à un certain seuil.

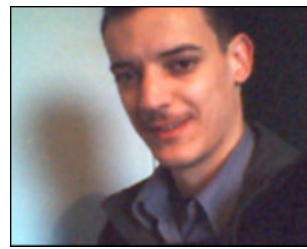


Image originale



Image de peau non filtrée



Image de peau filtrée

Figure 8 : exemple de filtrage avec la contrainte de surface

4. Détection des visages par appariement du gabarit « template matching »

L'appariement de gabarit ou template matching, consiste à comparer l'intensité des pixels entre un gabarit prédéfini et plusieurs sous-régions de l'image à analyser. Ce processus consiste en pratique à effectuer plusieurs balayages couvrant toute la superficie de l'image. Les endroits les plus propices à la présence de visages seront donc facilement identifiés par des minimums de distance entre le gabarit et l'image sous-jacente.

Parmi les distances envisageables, il y a notamment la norme L1 (distance de Manhattan) et la norme L2 (distance euclidienne), la cross-corrélation, ...etc.

La construction d'un gabarit efficace représente un défi en elle-même. En effet, l'utilisation d'un gabarit plus ou moins adapté au type d'objet recherché peut nuire à une détection efficace et diminuer la précision des résultats.

Pour ce qui est de la localisation des différentes caractéristiques du visage, leurs positions peuvent être déduites à partir des positions correspondantes sur le gabarit. Celles-ci doivent être préalablement déterminées manuellement en positions relatives par rapport aux dimensions du gabarit. Il n'est cependant pas garanti que le gabarit soit parfaitement positionné en translation, en échelle et en rotation sur le visage à détecter, ce qui produira des coordonnées légèrement

erronées.

4.1 Création du modèle de visage

La caractéristique la plus importante de cette méthode est qu'elle utilise les visages pour décider si une région de peau correspond à un visage ou non. Notre modèle a été créé en effectuant un calcul de la moyenne à partir de plusieurs images en niveaux de gris (dans notre cas), qui ne contiennent que des visages. La base de données que nous avons utilisée a été réalisée au LAIG (*Laboratoire d'Automatique et Informatique de Guelma*), elle contient 1823 visages pris dans différents profils et dans différentes conditions d'éclairage (figure 10).

Le modèle résultant sera une image en niveau de gris. La figure suivante montre des exemples de modèles [6] :



Figure 9 : Exemple de modèles de visages
(c) est le modèle utilisé



Figure 10 : Quelques exemples extraits de la base des visages

4.2 Recherche des visages dans les régions peau par « Template Matching » :

L'utilisation de la méthode du "template matching" se justifie principalement par la contrainte de temps réel qu'il faut appliquer au programme : il est nécessaire d'implémenter une fonction rapide, afin d'éviter que le système ne mette trop de temps pour réagir. Un autre avantage de cette méthode est la simplicité de l'implémentation. Le principe de cette méthode est le suivant : il s'agit, étant donné une image quelconque et un modèle prédéfini, de déterminer le degré d'appartenance (similarité) de ce modèle à cette image [4].

4.3 Mesures de similarité

Il existe plusieurs manières de mesurer la similarité entre deux images.

La première mesure appliquée est une simple différence de norme, données par la première formule pour la norme L1, et la deuxième formule pour la norme L2.

$$\text{Norme L1 : } \sum_{(i,j)} |a_{(i,j)} - b_{(i,j)}|$$

$$\text{Norme L2 : } \sqrt{\sum_{(i,j)} a_{(i,j)}^2 - b_{(i,j)}^2}$$

Où les $a_{(i,j)}$ et $b_{(i,j)}$ sont la valeur du pixel (i,j) dans les deux images à comparer. Il ressort donc que plus ces valeurs sont petites, plus les images sont proches [3].

La deuxième mesure calcule la cross-corrélation, donnée par la formule suivante :

Cross-corrélation :

$$R_x(r,c) = \sum_{j=1}^{tplRows} \sum_{i=1}^{tplCols} t(i,j) x\left(r+j-\frac{tplRows}{2}, c+i-\frac{tplCols}{2}\right)$$

Cette mesure est un peu plus complexe que la précédente, puisqu'elle n'apporte pas des résultats directs. Effectivement, le résultat est en fait une image, et la similarité peut se définir de la manière suivante : plus la probabilité d'apparition du template est grande, plus grande est l'intensité de l'image résultante [3].

Finalement, une troisième technique a été appliquée aux images binaires :

Il s'agit d'un OU Exclusif suivi d'une négation (\overline{XOR}) sur chaque pixel correspondant aux deux images. De cette manière les pixels qui ont la même valeur seront blancs dans l'image résultante et noirs aux endroits où ils sont différents [3].

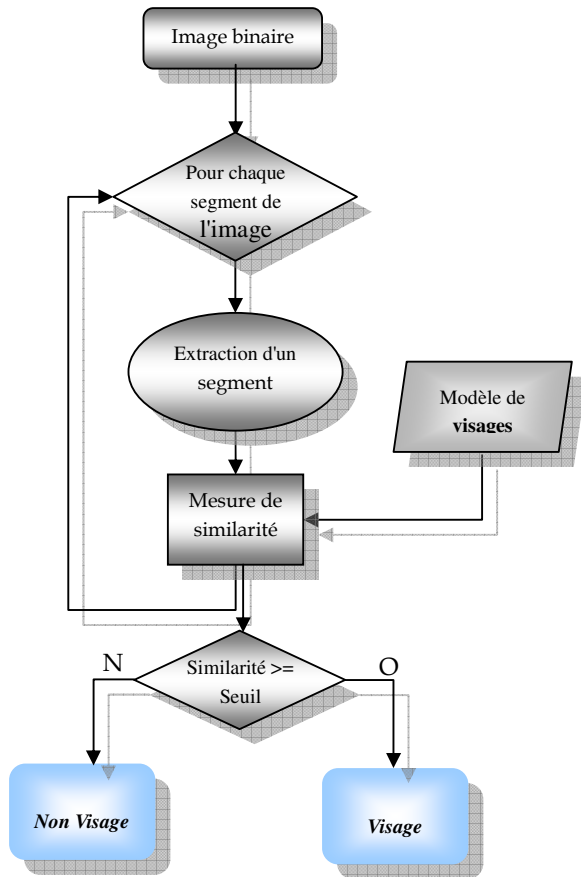


Figure 11 : Schéma illustratif du template matching

4.4. Application de la méthode du « Template Matching »

Pour effectuer l'appariement avec un segment détecté en utilisant le modèle du visage créé, nous passons par les étapes suivantes :

a) Extraction de l'image niveau de gris

D'abord, nous devons extraire l'image binaire du segment et remplir les trous existants. Nous multiplions l'image résultante par l'image originale en niveau de gris, le résultat sera une image en niveau de gris qui représente le segment en question. En fin, nous réduisons la taille de l'image pour qu'elle ne contienne que les pixels du segment. La figure 12 illustre un résultat de cette opération.

b) Normalisation du segment

Cette procédure de normalisation porte sur l'image en niveau de gris du segment, afin de rendre possible et d'améliorer l'opération d'appariement. La normalisation consiste à ajuster l'inclinaison du segment selon son orientation et rendre les dimensions du segment égales à celles du modèle.

c) Comparaison du segment avec le modèle

Cette comparaison s'effectue en balayant la surface

du segment et en calculant la cross-corrélation entre le modèle et la région sélectionnée du segment normalisé. Le résultat de cette procédure est la valeur maximale calculée ainsi que la position de la région correspondante. La valeur est comprise dans l'intervalle des valeurs réelles : [0.0, 1.0].

d) Décision

En fixant un seuil de similarité, nous pouvons juger un segment de peau comme étant visage ou non.

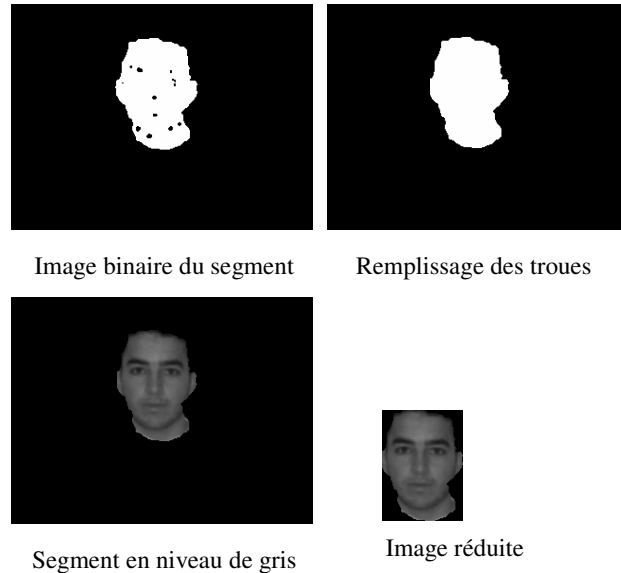


Figure 12 : Extraction du segment

5. Illustration et interprétation des résultats

Nous avons effectué quelques tests sur des images que nous avons prises par une caméra personnelle et d'autres téléchargées sur le Net. Les tests sont effectués sur des visages qui se présentent dans différentes couleurs et différentes orientations et échelles. Nous remarquons dans la :

Figure 13 (a) 1 détection, 0 non détection et 0 fausse alerte.

Figure 13 (b) 2 détections, 0 non détection et 1 fausse alerte.

Figure 13 (c) 5 détections, 0 non détection et 1 fausse alerte.

Figure 13 (d) 0 détections, 1 non détection et 0 fausse alerte.

Les fausses alertes sont parfois dues à la couleur de l'objet qui prend la couleur de peau tel que dans la figure 13 (b), 13 (c) alors que dans certaines scènes les visages ne sont pas détectés tel que dans la figure 13 (d) où le visage du bébé n'a pu être détecté car ce dernier fait partie d'un segment de peau de taille dépassant la norme et ce dernier a été éliminé dans la phase de filtrage de cette image.

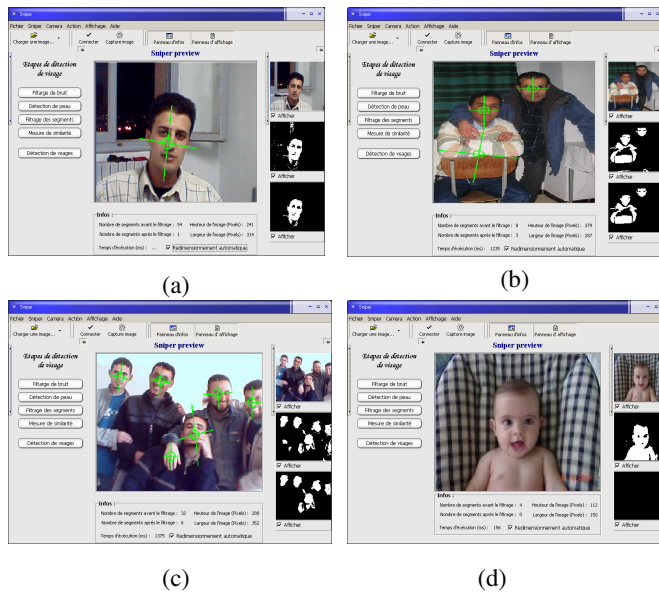


Figure 13 : Illustration de quelques résultats

6. Conclusion

L'application que nous venons de présenter dans cet article est une étape préliminaire et d'une grande importance dans tout système de reconnaissance de visages car avant de lancer le système pour reconnaître les personnes, une vérification de la présence ou pas d'un visage dans la scène ne peut être que très bénéfique. Cette étape permettra au système de détourner l'opération de reconnaissance si la scène ne contient aucun visage.

Notre application est basée sur la recherche de la couleur de peau dans la scène, l'espace de représentation des couleurs choisi est l'espace RGB. Les résultats obtenus sont satisfaisants, car ils nous ont permis d'augmenter les performances de la méthode de détection des couleurs et celle du template matching utilisées séparément. Cependant, pour mieux contourner le problème de détection des visages, nous proposons comme perspectives, à cours terme, d'essayer d'autres espaces de couleurs tel que l'espace RGB normalisé et l'espace HSV et de comparer les résultats obtenus.

7. Références

- [1] Diane FERGUSON, François LEVASSEUR, "Détection de visages" LAPI - ISICO 2004
- [2] Garcia C. , Tziritas G., "Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis", *IEEE Transactions on Multimedia*, 1(3), September 1999, p.264-277
- [3] G. Heusch: "Détection automatique de visages dans une séquence vidéo", SSC semestre 5 - 18 Février 2002.
- [4] H. Chang and U. Robles, "Face Detection", May 25, 2000

- [5] A. Iemieux, "Système d'identification de personnes par Vision numérique", Décembre 2003.
- [6] J. Serra, "Image analysis and mathematical morphology", 1989
- [7] SM. Bileschi, "Advances in Component-based Face Detection", MASSACHUSETTS INSTITUTE OF TECHNOLOGY January 2003.
- [8] C.C. Chiang, W.N. Tai, M.T. Yang, Y.T. Huang, et C.J. Huang. A novel method for detecting lips, eyes and faces in real time. *Real-Time Imaging*, 9, 2003.
- [9] K. Toennies, F. Behrens, et M. Aurhammer. Feasibility of hough-transform-based iris localisation for realtime-application. Dans *International Conference on Pattern Recognition*, 2002.
- [10] M.J. Chen, M.C. Chi, c.T. Hsu, et J.W. Chen. Roi video coding based on h.263+ with robust skin-color detection technique. *IEEE Transactions on Consumer Electronics*, 49(3), 2003.