



**HAL**  
open science

# Détection et localisation de visages sur image numérique

Gilles Burel, Dominique Carel

► **To cite this version:**

Gilles Burel, Dominique Carel. Détection et localisation de visages sur image numérique. Revue-Technique-Thomson-CSF, 1993, 25 (1). hal-03221446

**HAL Id: hal-03221446**

**<https://hal.univ-brest.fr/hal-03221446v1>**

Submitted on 19 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# Détection et localisation de visages sur image numérique

Gilles BUREL<sup>1</sup> & Dominique CAREL

Thomson CSF, Laboratoires Electroniques de Rennes

## Résumé

Un procédé automatique de détection/localisation de visages sur images numériques est proposé. La méthode repose sur les idées d'apprentissage par l'exemple, et d'analyse multi-résolution de l'image. Une gestion originale des bases d'apprentissage est proposée afin d'améliorer les performances. Divers résultats expérimentaux correspondant à une réalisation dans laquelle le classifieur est un perceptron multi-couches, entraîné par l'algorithme de rétropropagation du gradient, sont présentés.

## Abstract

We propose a method for automatic detection and location of faces on digital images. The approach is based on learning techniques and multi-resolution analysis. Special emphasis is put on management of the learning data, in order to improve the performances. Various experimental results, obtained by using a multi-layers perceptron as classifier, are provided.

**Mots Clés :** Détection et Localisation de Visages, Multi-Résolution, Classification, Perceptron Multi-Couches, Rétro-Propagation.

**Keywords :** Detection and Location of Faces, Multi-Resolution, Classification, Multi-Layers Perceptron, Back-Propagation.

**Title :** Detection and localization of faces on digital images

---

<sup>1</sup>nouvelle adresse : UBO/LEST, BP 809, 29285 Brest cedex, France

# 1 Introduction

Un procédé automatique de détection/localisation de visages sur images numériques est proposé. La détection/localisation de visages consiste, à partir d'une image numérique, à fournir les coordonnées des coins de fenêtres rectangulaires encadrant les visages présents dans l'image. Une indication de confiance peut également être associée à chaque fenêtre.

On peut trouver dans la littérature différents articles traitant de la caractérisation de visages. Le but fixé est généralement l'identification de visages, c'est à dire l'indication du nom de la personne, lorsque le visage a déjà été détecté, et donc localisé (photo d'identité, cadrage manuel, etc). Les techniques mises en œuvre sont par exemple l'analyse de cartes d'isodensité [8], la mesure de caractéristiques anthropomorphiques [2], l'analyse de symétries [4], ou l'analyse en composantes principales [11]. Une revue est fournie par [10]. Certaines de ces techniques peuvent être extrapolées à la détection de visages, mais elles n'ont généralement pas été conçues pour cette application.

Une autre méthode que l'on pourrait envisager d'utiliser indirectement à des fins de détection de visages est décrite dans [6]. Elle permet de détecter les yeux, mais nécessite des images de très grande résolution (les yeux doivent être très nettement détaillés). Son application à la détection de visages n'est donc pas viable, car dans la plupart des applications réelles pour lesquelles la détection de visages peut être nécessaire, il est irréaliste d'espérer une telle résolution.

L'absence de méthode efficace de détection de visages dans la littérature est principalement due aux problèmes suivants :

1. Il n'existe pas de modèle de visage, du fait d'une part de la forte variabilité d'un individu à l'autre, et d'autre part de la non rigidité d'un visage. On est donc confronté à un problème bien plus difficile que les problèmes classiques de reconnaissance d'objets rencontrés en traitement d'images (reconnaissance de pièces industrielles par exemple), pour lesquels on peut construire des modèles a priori.
2. Dans le cadre d'applications réalistes, on ne maîtrise pas la distance visage-caméra. La taille d'un visage sur l'image peut donc être quelconque.
3. De la même façon, en environnement réel, on ne maîtrise que partiellement les conditions d'éclairage.
4. Enfin, toujours dans le cadre d'applications réalistes, un visage n'est pas nécessairement vu de face. Il faut donc être capable de le détecter même s'il est légèrement de profil.

Pour surmonter ces problèmes, nous avons développé une méthode basée sur les concepts suivants :

- Apprentissage par l'exemple, ce qui permet de surmonter la non-existence de modèle a priori.
- Analyse multi-résolution de l'image, afin de surmonter le fait que l'on ne maîtrise pas la distance visage-caméra. Cette analyse multi-résolution est complétée par un post-traitement visant à supprimer des détections imbriquées.
- Normalisation locale en moyenne et en variance, afin d'obtenir une certaine insensibilité par rapport aux conditions d'éclairage.

- Création d'une base d'apprentissage variée, contenant notamment des visages de face et légèrement de profil.

Pour la clarté de la lecture, l'article se décompose en deux parties : une première partie, qui décrit complètement le système proposé, et une seconde partie qui fournit différents résultats expérimentaux, et permet d'illustrer différents points de la méthode.

## 2 Description de la méthode

La méthode se décompose en trois phases :

- Une phase d'apprentissage général, durant laquelle le dispositif règle ses paramètres internes. Les valeurs initiales des paramètres internes sont choisies aléatoirement.
- Une phase d'apprentissage adapté, durant laquelle le dispositif de détection règle ses paramètres internes, en les adaptant plus spécialement à un environnement particulier. Les valeurs initiales des paramètres internes sont celles qui résultent de l'apprentissage général.
- Une phase de détection durant laquelle les paramètres internes du dispositif sont figés aux valeurs fournies par la phase d'apprentissage adapté. Durant cette phase, le dispositif peut traiter des images numériques. Il reçoit en entrée une image numérique, et fournit en sortie les coordonnées des coins de fenêtres rectangulaires encadrant les visages trouvés sur l'image. Il fournit également avec chaque fenêtre une indication de confiance.

Les deux premières phases constituent des étapes préliminaires, dont le but est de régler les paramètres internes du système. La dernière phase constitue le fonctionnement normal du dispositif en tant que détecteur de visages.

Une vue globale de la méthode est fournie sur la figure 1. Des compléments d'information pourront être trouvés dans les documents [3] et [1].

Dans la suite de l'exposé, lorsque nous parlerons de la taille standard d'une imagerie, il s'agit de la taille sous laquelle toute imagerie est ramenée avant d'être présentée au classifieur. Cette taille est typiquement de l'ordre de 20 pixels en hauteur et 15 pixels en largeur.

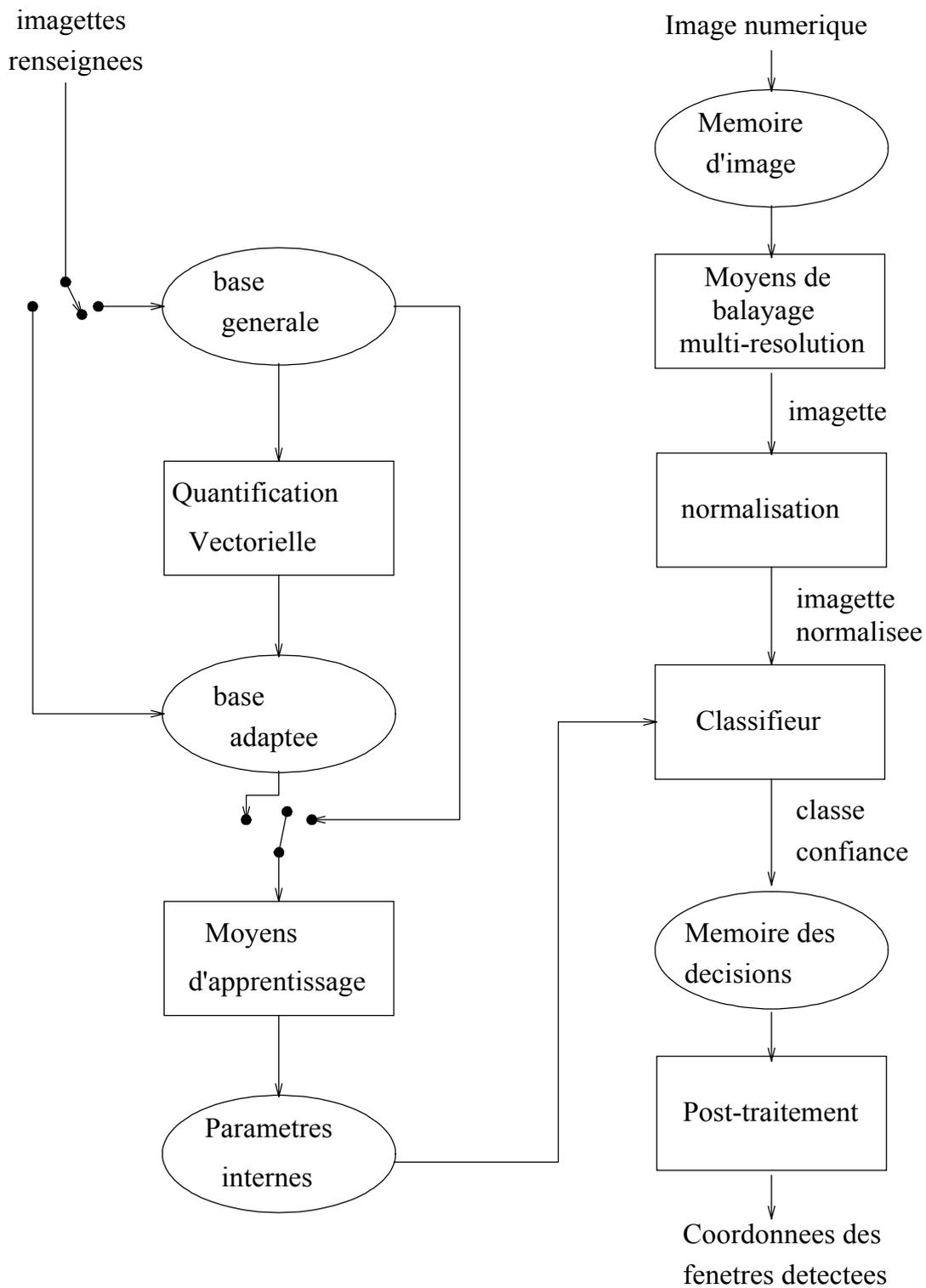


FIG. 1: Principe de la méthode

## 2.1 Apprentissage général

Cette phase consiste :

1. A créer une base d'apprentissage générale
2. A initialiser les paramètres internes du dispositif à des valeurs aléatoires.
3. A exécuter un algorithme d'apprentissage permettant d'ajuster au mieux les paramètres internes compte tenu de la base d'apprentissage créée en (1).

Pour créer une base d'apprentissage générale, on doit acquérir et enregistrer des images numériques représentant différentes personnes dans des environnements aussi variés que possible, et sous diverses apparences (visages de face et légèrement de profil). Puis on encadre ces visages de manière interactive, et pour chaque encadrement, on normalise l'imagette correspondante en moyenne et en variance, on ramène sa taille à la taille standard, et on l'enregistre.

Il est également nécessaire de disposer de contre-exemples. Pour cela, on encadre certaines zones du fond, et pour chaque encadrement, on normalise l'imagette correspondante en moyenne et en variance, on ramène sa taille à la taille standard, et on l'enregistre.

Cet ensemble d'imagettes de visages et de fonds constitue ce que nous nommerons la base d'apprentissage générale (pour chaque imagette, il faut bien entendu avoir mémorisé s'il s'agit d'un visage ou d'un fond).

Il est possible de créer des imagettes supplémentaires en effectuant de légères translation et homothéties sur chaque fenêtre de départ.

A titre indicatif, il est souhaitable d'avoir une base d'apprentissage constituée d'au moins 2000 imagettes de visages et 2000 imagettes de fonds. Il est également souhaitable que le nombre d'imagettes de fonds soit du même ordre de grandeur que le nombre d'imagettes de visages, afin d'éviter un déséquilibre entre les exemples et les contre-exemples.

## 2.2 Apprentissage adapté

Le dispositif de détection de visages est normalement appelé à être installé en un lieu particulier, et il est également possible que les personnes susceptible d'apparaître sur l'image appartiennent à un groupe connu à priori. Il est alors intéressant d'adapter le dispositif à ces conditions particulières afin d'améliorer ses performances.

La phase d'apprentissage adapté consiste :

1. A comprimer la base d'apprentissage générale.
2. A créer une base d'apprentissage adaptée à un certain environnement, en ajoutant à la base d'apprentissage générale comprimée des exemples et contre-exemples extraits sur un lieu particulier.

3. A initialiser les paramètres internes du dispositif avec les valeurs fournies par la phase d'apprentissage général.
4. A exécuter un algorithme d'apprentissage permettant d'ajuster au mieux les paramètres internes compte tenu de la base d'apprentissage créée en (2).

Pour comprimer la base d'apprentissage générale, On peut par exemple utiliser un algorithme de quantification vectorielle tel que l'algorithme LBG [7], ou l'algorithme des cartes topologiques [5]. Un tel algorithme transforme un ensemble de M imagerie en un ensemble de N imagerie, ou N est inférieur à M. Les N imagerie sont celles qui représentent le mieux les M imagerie de départ au sens de l'erreur quadratique moyenne. L'algorithme LBG nécessite que N soit une puissance de 2. L'algorithme des cartes topologiques admet une valeur de N quelconque. On applique donc un algorithme de quantification vectorielle aux imagerie de visages de la base générale, puis aux imagerie de fonds. A titre d'exemple, on peut conserver 256 imagerie de visages et 256 imagerie de fonds.

On installe ensuite l'appareil sur site, et on réalise sur place l'acquisition d'une ou plusieurs images sans personnages présents. On charge les paramètres internes du dispositif avec les valeurs fournies par l'apprentissage général. On balaye chaque image avec le procédé décrit dans la section suivante, et on mémorise les fenêtres indiquées comme étant des visages. Pour chaque fenêtre, on normalise l'imagerie correspondante en moyenne et en variance, on ramène sa taille à la taille standard, et on l'enregistre. Comme il n'y avait pas de personnages sur l'image, ces imagerie sont ce que nous appellerons des "fausses détections". Si le nombre de fausses détections est supérieur à N/2, on le comprime à l'aide d'un algorithme de quantification vectorielle pour obtenir N/2 imagerie. On rajoute alors ces imagerie dans la base d'exemples de fonds.

De la même façon, on réalise l'acquisition d'images dans lesquelles les personnes qui sont susceptibles d'apparaître ultérieurement sont présentes. On encadre leurs visages, on normalise chaque imagerie, et on rajoute les imagerie obtenues à la base d'imagerie de visages, en ne dépassant toutefois pas un nombre d'imagerie rajoutées égal à N/2.

## 2.3 Détection

Les paramètres internes du dispositif sont maintenant figés aux valeurs fournies par la phase d'apprentissage adapté. On réalise un balayage multirésolution de l'image avec des fenêtres rectangulaires dont la taille varie par exemple d'un facteur  $2^{\frac{1}{3}}$  lorsque l'on passe d'une résolution à la suivante. Le rapport hauteur sur largeur de chaque fenêtre est le même et est fixé a priori pour correspondre approximativement aux proportions d'un visage (ce rapport doit être de l'ordre de  $\frac{4}{3}$ ). Pour chaque position de la fenêtre, et pour chaque résolution, on normalise en moyenne et en variance le contenu de la fenêtre, puis on ramène sa taille à la taille standard. Le résultat est fourni au classifieur qui donne en sortie une indication de classe (visage ou fond) et une confiance associée à cette indication. Si le classifieur indique qu'il s'agit d'un visage avec une confiance supérieure à un seuil fixé a priori, on mémorise alors les coordonnées des coins de la fenêtre, et la confiance.

Une amélioration possible consiste à classer systématiquement comme fond les fenêtres dont la variance avant normalisation est inférieure à un seuil fixé a priori.

Dans un deuxième temps (post-traitement), on traite l'ensemble des fenêtres obtenues afin de supprimer certaines fenêtres imbriquées. Deux fenêtres sont considérées comme imbriquées si le centre de l'une est à l'intérieur de l'autre. Une fenêtre est supprimée si elle est imbriquée dans une fenêtre de confiance supérieure.

On peut également, en se basant sur un raisonnement géométrique élémentaire, supprimer toute fenêtre se trouvant dans l'image en dessous d'une fenêtre de taille nettement plus grande, et de confiance supérieure. En effet, un petit visage serait plus loin qu'un grand visage, et devrait donc être masqué par le corps de la personne la plus proche.

## **2.4 Possibilités d'améliorations**

Il est possible d'améliorer le procédé en réalisant un prétraitement qui fournit des zones d'intérêt sur l'image. On ne balaye alors que ces zones d'intérêt et non pas l'image entière.

Par exemple, on peut utiliser un détecteur de mouvement, et considérer que seules les zones en mouvement sont dignes d'intérêt. L'idée sous jacente étant qu'un visage ne reste pas immobile sur une longue période.

# **3 Résultats expérimentaux**

## **3.1 Création d'une base d'apprentissage générale et premières expérimentations**

### **3.1.1 Conditions expérimentales**

Nous utilisons une caméra d'angle d'ouverture  $90^\circ$  en horizontal, et qui fournit une image noir et blanc de  $720 \times 576$  pixels. On souhaite détecter des visages dont la distance est comprise entre 1.5 et 5m, ce qui correspond à la fourchette de distance moyenne pour les applications typiques de ce type de méthode. En conséquence, compte tenu de l'angle d'ouverture et de la résolution de la caméra, la taille d'un visage sur l'image est approximativement comprise entre  $18 \times 25$  pixels (pour une distance de 5m) et  $60 \times 80$  pixels (pour une distance de 1.5 m).

Nous avons placé la caméra dans une grande salle. Une quarantaine de personnes ont été invitées à venir se placer dans le champ de la caméra, par groupes de 4 à 8 personnes, sans contrainte d'attitude. Diverses séquences d'images ont été acquises en changeant les conditions d'éclairage et en modifiant le fond. La base d'images a ensuite été découpée en deux parties : une base d'apprentissage et une base de test, de telle sorte que les personnes et les fonds présents dans la base d'apprentissage ne le soient pas dans la base de test.

### **3.1.2 Choix des paramètres du système**

Nous avons testé une réalisation de l'approche proposée dans laquelle le classifieur est un perceptron multi-couches (ou réseau de neurones multi-couches) entraîné par l'algorithme de rétropropagation du gradient [9]. La fenêtre d'analyse subit systématiquement une homothétie qui la ramène à une taille standard de 15x20 pixels. Le perceptron multi-couches possède donc 300 entrées, et 2 sorties (nombre de classes : fond/visages). Dans ce cas particulier, le nombre de sorties aurait pu être ramené à 1, mais ceci n'a pas été fait afin de maintenir une compatibilité avec divers outils d'évaluation de classifieurs, qui supposent dans le cas général que le nombre de sorties est égal au nombre de classes. Nous avons testé un perceptron 2 couches (300+2) et un perceptron 3 couches (300+10+2).

L'image est analysée à 7 résolutions différentes. Les tailles des fenêtres d'analyse multi-résolution sont les suivantes : 18x25, 23x31, 30x40, 37x50, 47x63, 60x80. Ceci permet la détection de tout visage se trouvant dans une gamme de distance allant de 1.5m à 5m. Comme on peut le constater, le facteur d'homothétie entre deux résolutions successives est  $2^{\frac{1}{3}}$ . Bien entendu, pour une implantation efficace, c'est la taille de la fenêtre d'analyse qui est fixe, et l'image qui est progressivement réduite.

La figure 2 représente un visage ramené à la taille standard de 15x20 pixels. On remarquera que l'on est proche de la taille minimale nécessaire à la détection visuelle d'un visage. Il n'est donc pas concevable de descendre beaucoup plus bas en taille standard (ceci aurait permis d'utiliser un classifieur plus petit, d'où un apprentissage et un traitement plus rapide).

Le pas de balayage de l'image est fixé à 2 pixels. Pour toute imagerie extraite au cours du balayage, on ramène sa luminance moyenne à 128 et son écart type à 40, avant de lui appliquer l'homothétie qui la ramène à la taille standard.

### **3.1.3 Création d'une base d'apprentissage générale**

Pour réaliser l'apprentissage, une base d'images de visages a été créée en encadrant interactivement les visages présents dans les séquences destinées à l'apprentissage. Un total de 121



FIG. 2: 27 imageries d'un visage, ramenées à la taille standard

visages, correspondant à une vingtaine de personnes différentes, ont ainsi été encadrés. Pour chaque encadrement ainsi défini, on extrait 27 imageries, en faisant subir à la fenêtre de départ des combinaisons des translations et des homothéties suivantes :

$$\Delta X = -\frac{T_X}{10}, 0, \frac{T_X}{10}$$

$$\Delta Y = -\frac{T_Y}{10}, 0, \frac{T_Y}{10}$$

$$\rho = 2^{-\frac{1}{6}}, 1, 2^{\frac{1}{6}}$$

où l'on a noté  $T_X$  et  $T_Y$  la taille de la fenêtre indiquée par l'opérateur,  $(\Delta X, \Delta Y)$  la translation, et  $\rho$  le facteur d'homothétie. On remarquera que les valeurs du facteur d'homothétie ont été choisies de telle sorte qu'elles occupent au mieux l'espace entre deux résolutions d'analyse successives (on rappelle que la résolution d'analyse varie d'un facteur  $2^{\frac{1}{3}}$  entre 2 résolutions successives). La figure 2 représente un exemple de 27 imageries générées à partir d'un seul encadrement manuel.

On obtient ainsi un total de  $121 \times 27 = 3267$  imageries de visages.

Le même nombre d'imageries de fond a été automatiquement extrait sur des séquences sans personnes présentes, d'où un total de l'ordre de 6000 exemples (visages + fonds).

Un premier réseau a appris sur cette base. Nous avons ensuite complété cette base en ajoutant 3000 fonds détectés par ce réseau comme visages sur des images diverses (scènes d'extérieur ou d'intérieur).

On a donc finalement un total de 9000 exemples (environ 3000 visages et 6000 fonds), qui constitue la base d'apprentissage générale.

### **3.1.4 Comparaison des réseaux 2 et 3 couches**

Nous avons réalisé l'apprentissage sur la base générale du réseau 2 couches et du réseau 3 couches. Le traitement d'images nouvelles avec chacun de ces réseaux ne fait pas apparaître de différence en termes de performances.

Ceci conduit à penser que les poids du réseau trois couches à l'issue de l'apprentissage sont tels qu'il se comporte en fait comme un deux couches. Pour vérifier ceci, nous avons transformé le réseau 3 couches en réseau 2 couches en réalisant le produit matriciel des matrices de poids correspondant respectivement au passage de la couche cachée à la couche de sortie et au passage de la couche d'entrée à la couche cachée. On constate alors que ce nouveau réseau donne les mêmes sorties sur les imagerie de la base d'apprentissage que le 3 couches initial. Ceci confirme le fait que les activités des neurones de la couche intermédiaire de ce réseau restaient confinées à la zone quasi-linéaire. Il est donc bien équivalent à un 2 couches.

En conséquence, comme un réseau 2 couches permet un apprentissage et un traitement plus rapide, c'est la structure que nous avons retenue pour la suite de l'étude.

### **3.1.5 Exemple de résultat**

La figure 3 présente un exemple de résultat obtenu avec le classifieur ayant appris sur la base générale. Cette image a été acquise dans la conditions signalées plus haut (§3.1.1). Les personnes présentes sur cette image ne faisaient pas partie de la base d'apprentissage.

Le type de résultat présenté sur cette image se confirme sur le traitement de séquences complètes prises à des instants différents. Une à trois fausses détections apparaissent parfois ponctuellement sur le fond en fonction des variations de l'éclairage (la salle est largement éclairée par la lumière du jour, et la nature de l'éclairage varie en fonction du passage des nuages devant le soleil). L'absence de détection de certains personnages se produit ponctuellement, par exemple lorsqu'une personne tourne très nettement la tête pour parler à son voisin (on a alors un visage de profil, tandis que la base d'apprentissage ne contenait que des visages de face ou de profil).

## **3.2 Création d'une base d'apprentissage adaptée**

Nous nous plaçons à présent dans une nouvelle salle, plus petite, avec une caméra différente, fournissant une image 512x356pixels. Mais l'angle d'ouverture est également plus faible qu'avec la caméra précédente, d'où une résolution angulaire du même ordre.



FIG. 3: Exemple de résultat (réseau général, visages non appris)

La caméra est fixée sur un robot 6 axes. Le robot peut ainsi analyser différentes zones de la pièce.

Comme indiqué dans la description du procédé, la base d'apprentissage générale a été comprimée à l'aide d'un réseau de Kohonen pour donner une base réduite de 256 visages et 256 fonds. Le robot a ensuite acquis différentes vues de la pièce, de manière aléatoire, en ayant connaissance du fait qu'il n'y avait pas de personnages dans la pièce. Chacune de ces vues a été traitée avec le réseau ayant appris sur la base générale. On sait que toute détection de visage sur ces images est nécessairement une fausse détection. Ces fausses détections ont été rajoutées à la base d'apprentissage générale comprimée afin de constituer la base d'apprentissage adaptée. La figure 4 montre une vue sur laquelle trois fausses alarmes sont présentes (zones encadrées). On remarquera au passage que le taux de confiance correspondant à ces fausses alarmes est faible (la confiance est indiquée en haut à gauche des fenêtres incrustées).

Pour certaines applications, les personnes susceptibles d'apparaître sont connues a priori, et l'on pourrait donc également rajouter des imageries de visages de ces personnes dans la base d'apprentissage adaptée. Cela n'a pas été fait ici car la dispositif est actuellement utilisé à des fins

de démonstration, et on ne connaît donc pas a priori les personnes susceptibles d'apparaître sur les images.

La base d'apprentissage adaptée étant constituée, on a réalisé quelques itérations d'apprentissage, en initialisant les poids du réseau de neurones aux valeurs issues de l'apprentissage sur la base générale. Ceci permet un apprentissage rapide, car le réseau part ainsi d'un état proche de la bonne valeur des poids.

Les figures 5 et 6 montrent quelques résultats de traitement dans ces nouvelles conditions. Notons que le fond peut avoir localement changé depuis la constitution de la base adaptée, à cause du déplacement d'objets. De même, les conditions d'éclairage ne sont plus les mêmes, car cette salle est éclairée par la lumière du jour. Enfin, le robot ne se trouve pas exactement dans l'une des positions qu'il a prises au cours de son acquisition de vues de la pièce destinées à la constitution de la base adaptée. A titre de curiosité, on peut signaler qu'il semble que le réseau ait atteint un certain niveau d'abstraction, ainsi que le montre la figure 7 !



FIG. 4: Exemple de fausses alarmes (réseau général)

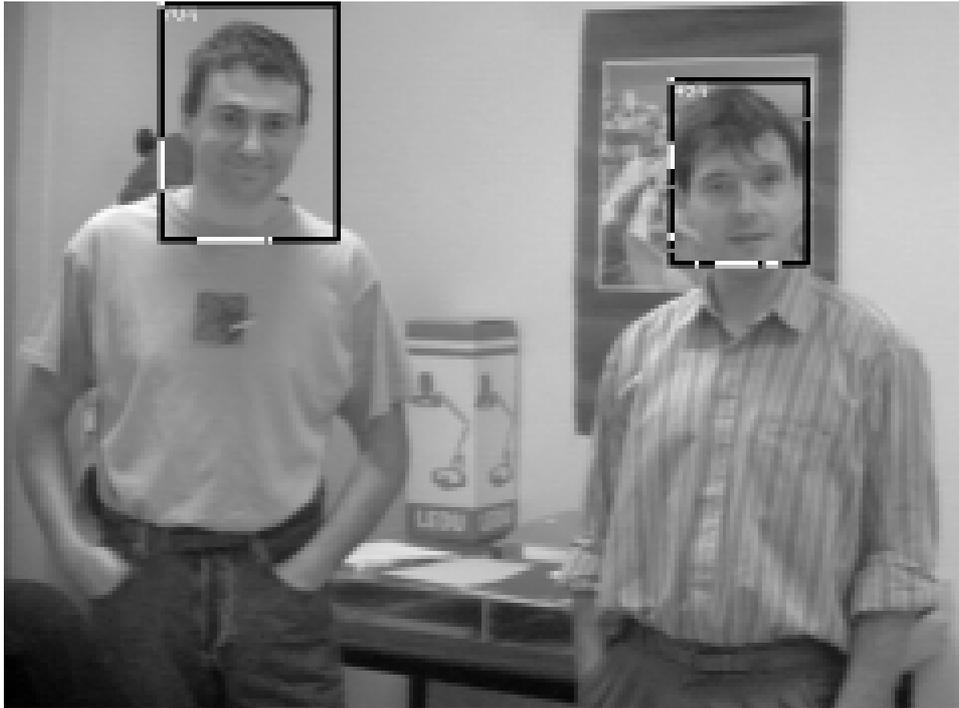


FIG. 5: Exemple de résultat (réseau adapté à la salle)

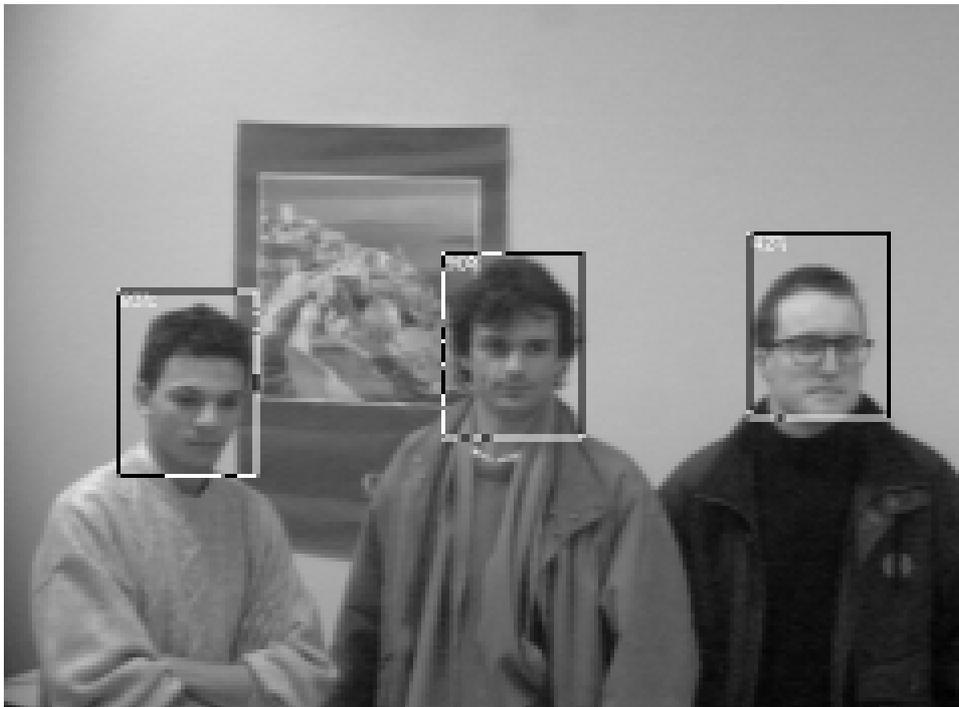


FIG. 6: Exemple de résultat (réseau adapté à la salle)

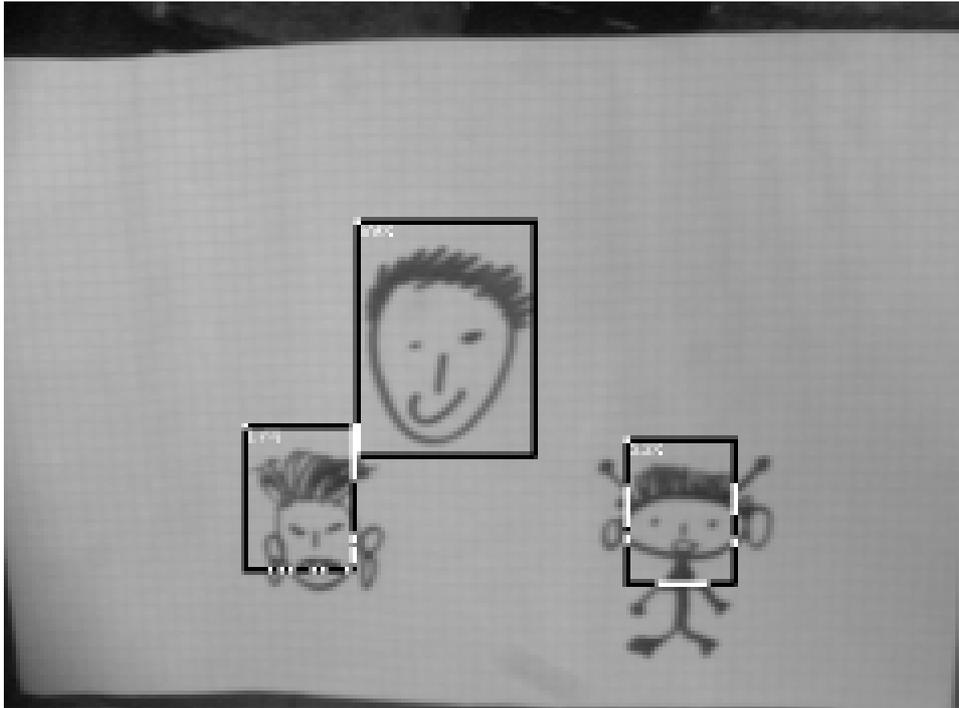


FIG. 7: Exemple de résultat

## 4 Conclusion

Nous avons proposé un procédé de détection de visages sur images numériques qui consiste :

- dans une première étape préliminaire, à créer une base d'images de visages et de fonds, représentant une forte variabilité (grand nombre de personnes et de fonds), à prétraiter chaque image, à la normaliser en taille, et à exécuter un algorithme d'apprentissage sur cette base d'images afin d'obtenir les valeurs des paramètres internes d'un classifieur en partant de valeurs aléatoires,
- dans une deuxième étape préliminaire, à compresser la base générale à l'aide d'un algorithme de quantification vectorielle, puis à compléter la base comprimée par des fausses détections obtenues sur des images acquises sur site, ces fausses détections étant prétraitées et normalisées en taille, et à exécuter un algorithme d'apprentissage sur cette base d'images afin d'obtenir les valeurs de paramètres internes d'un classifieur en partant des valeurs fournies par l'étape précédente,
- dans une étape de détection, à charger les paramètres internes du classifieur avec les valeurs fournies par la seconde étape préliminaire, puis, pour chaque image à traiter, à balayer en multi-résolution cette image, à prétraiter et à normaliser en taille chaque image, à propager le résultat dans le classifieur, et enfin à traiter les décisions fournies par le classifieur afin de supprimer les fenêtres imbriquées.

Divers résultats expérimentaux correspondant à une réalisation dans laquelle le classifieur est un perceptron multi-couches, entraîné par l'algorithme de rétropropagation du gradient [9] ont été

présentés, et semblent concluants. Diverses améliorations sont possibles, notamment la limitation du balayage de l'image à des zones d'intérêt obtenues par exemple par détection du mouvement, et le filtrage temporel des fausses alarmes. Signalons également que la base d'apprentissage ne contenait que les visages d'une vingtaine de personnes, ce qui est probablement insuffisant pour bien représenter toute la diversité des visages. L'acquisition d'une base plus importante est donc également un facteur d'amélioration possible.

## Références

- [1] Gilles BUREL, Dominique CAREL  
"Procédé de détection automatique de visages sur une image numérique ainsi que l'application à la mesure d'audience télévisuelle et à la télésurveillance".  
Brevet n° 92-05500 déposé le 5 mai 1992
  
- [2] Ian CRAW, David TOCK, Alan BENNETT  
"Finding face features"  
ECCV92 (European Conference on Computer Vision), Genova, Italy
  
- [3] Sébastien DOMMANGET  
"Détection de visages par méthode connexionniste"  
Stage de fin d'étude INT effectué à TCSF-LER  
Responsables de stage : G. BUREL & D. CAREL  
Rapport de stage, 31 janvier 1992
  
- [4] Shimon EDELMAN, Daniel REISFELD, Yechezkel YESHURUN  
"Learning to recognize faces from examples"  
ECCV92 (European Conference on Computer Vision), Genova, Italy
  
- [5] Teuvo KOHONEN  
"Self-Organization and Associative Memory"  
Springer-Verlag, 1984
  
- [6] R.A. HUTCHINSON  
"Development of an MLP feature location technique using preprocessed images"  
Congrès INNC90, Paris, 9-13 Juillet 1990, pp 67-70.
  
- [7] Y.LINDE, A.BUZO, R.M.GRAY  
"An algorithm for Vector Quantizer design"  
IEEE Trans. on Communications  
vol COM-28, n°1, January 1980

- [8] Osamu NAKAMURA, Shailendra MATHUR, Toshi MINAMI  
“Identification of human face based on isodensity maps”  
Pattern Recognition, vol. 24, *n*<sup>o</sup> 3, 1991
- [9] D.E. RUMELHART, G.E. HINTON, R.J. WILLIAMS  
“Learning internal representations by error backpropagation”  
Parallel Distributed Processing, D.E. RUMELHART and J.L. Mc CLELLAND  
Chap8, Bradford book - MIT Press - 1986
- [10] Ashok SAMAL, Prasana A. IYENGAR  
“Automatic recognition and analysis of human face and facial expression : a survey”  
Pattern Recognition, vol 25, *n*<sup>o</sup>1, pp 65-77, 1992
- [11] Matthew TURK, Alex PENTLAND  
“Eigenfaces for recognition”  
Journal of Cognitive Neuroscience, vol 3, *n*<sup>o</sup>1