

Détection et reconnaissance des visages basée sur les réseaux de neurones artificiels

ADJOUJ Réda

*Evolutionary Engineering and Distributed
Information Systems Laboratory, EEDIS
Computer Science Department
University of Sidi Bel-Abbès, Algeria
E-mail : AdjReda@yahoo.fr*

BOUKELIF Aoued

*Laboratoire des télécommunications et du traitement numérique de signal,
équipe de recherche des techniques vidéo Département Electronique,
Université de S.B.A.
E-mail: aboukelif@yahoo.fr*

1. Introduction

Les réseaux de neurones (Neural Networks NN) ont pour objectif la classification et la prédiction. La plupart des algorithmes de reconnaissance de visages cherchent à extraire un code descriptif réduit du visage (entre 10 et 500 paramètres), puis à rattacher ce code à une classe statistique correspondante à un individu.

La méthode employée consiste à localiser les informations utiles pour le système c'est à dire localiser seulement le visage, extraire des caractéristiques à partir de ce visage, comparer ces caractéristiques avec d'autres qui existent dans la base de données images .Au cas où ces caractéristiques se correspondent alors le visage est identifié ou reconnu.

La réduction du champ de la recherche (qui est le rôle de la procédure de classification) limite le nombre des visages à parcourir en choisissant seulement ceux qui sont susceptibles de correspondre au visage cherché, c'est à dire ceux qui ont des caractéristiques qui répondent au critère de classification puis on procédera à l'identification.

L'authentification par reconnaissance de visages se heurte à plusieurs limitations et ne sera jamais assez fiable pour des applications de haute sécurité:

Un humain moyen atteint un taux de reconnaissance de 99%. *Les NN implémentés sur ordinateur peuvent-ils le dépasser ?*

- Toute population comporte environ 1% de vrais jumeaux qui ont deux visages différents mais qui ont les mêmes distances métriques
- Les humains peuvent parvenir à les reconnaître avec de l'expérience, qu'en est-il des ordinateurs ?
- Comment l'ordinateur pourra-t-il comprendre qu'il n'a affaire qu'à une photo d'une photo, ou pire, d'un masque ?

2. Les méthodes de reconnaissance et d'identification des visages

Parmi les méthodes de reconnaissances des visages, on cite :

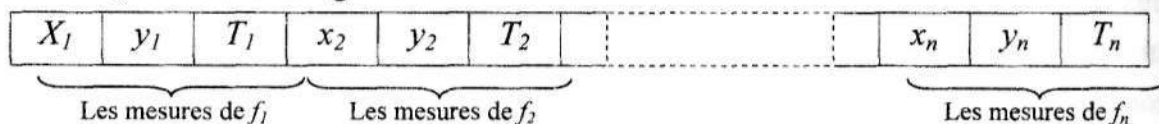
- L'approche à base de vues [*'view-based approach'*].
- L'approche fondée sur l'apparence [*'appearance-based approach'*].
- L'approche des caractéristiques invariantes [*'invariant features approach'*].
- L'approche de Reconnaissance des visages basée sur les NNs[Mitchell,1997]:

2.1. L'approche à base de vues [*'view-based approach'*]

Cette technique utilise une représentation vectorielle du visage, ce vecteur est le résultat d'une détection faite au préalable, et il contient les caractéristiques qui ont été énumérées : f_1, f_2, \dots, f_n , et que la représentation du vecteur contient en premier les mesures de f_1 puis de $f_2 \dots$ jusqu'à f_n .

Les mesures pour chaque f_i incluent son emplacement (x,y) et l'intensité ou la valeur de la couleur locale de ce point, c'est-à-dire pour chaque f_i on a un triplet (x_i, y_i, T_i) telle que :

I_a : vecteur image



(x_i, y_i) : les coordonnées de la caractéristique f_i du visage.

T_i : L'intensité du pixel (x_i, y_i) qui représente f_i .

$$V = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ \dots \\ \dots \\ x_n \\ y_n \end{bmatrix} \text{ tel que } V \subset I_a \text{ vecteur d'image}$$

$$f_i : I_a \rightarrow \mathbb{R}^+ \times \mathbb{R}^+ \times D^+$$

$$\text{pixel}_i \propto (x_i, y_i, T_i)$$

$$\text{et } \mathbb{R}^+ \times \mathbb{R}^+ : [256 \times 256]$$

$$D^+ : [0 \rightarrow 256 \text{ couleurs}]$$

V : Vecteur position des f_i , V est de taille $2n$.

2.2. L'approche fondée sur l'apparence [*"appearance-based approach"*].

C'est une technique fondée sur l'apparence pour caractériser et rechercher des visages dans une base de données images. Cette méthode est invariante à la translation, à la rotation plane, et au changement d'échelle et elle peut capter l'information significative de l'image sans recours à la segmentation. Elle est également très bien adaptée au faible changement de point de vue et elle est également robuste au bruit et aux occultations.

2.3. L'approche des caractéristiques invariantes [*"invariant features approach"*].

Cette approche consiste à enregistrer des caractéristiques dans la vue d'exemples qui ne changent pas, caractéristiques telles que la couleur ou les invariants géométriques, lorsque les paramètres "pose, expression et éclairage" changent. Cette approche *n'est pas applicable dans la reconnaissance de visage* alors qu'elle présente de bons résultats pour l'*indexation d'articles d'épicerie emballés* utilisant la couleur.

2.4 . L'approche de Reconnaissance des visages basée sur les NN [Mitchell, 1997]:

Le professeur Tom Mitchell avec son équipe du laboratoire de l'intelligence artificielle de CMU, ont initié le projet de reconnaissance des visages en utilisant les NN artificiels.

Le professeur Tom Mitchell a abouti enfin à une architecture énorme ,plus de 20 NN pour reconnaître 20 personnes, il en résulte une architecture colossale ; les résultats sont favorables[Mitchell,1997].

Avec cette méthode on a besoin d'incrémenter l'architecture du NN quand on veut reconnaître plus de personnes. A un moment donné, on sera limité par la technologie.

3. Principales difficultés des approches

La performance d'un algorithme de reconnaissance de visages dépend principalement du nombre de paramètres statistiquement indépendants extraits, et de leur variance.

- La variance doit être faible au sein d'une même classe
- La variance entre les classes doit être grande

On se heurte alors à trois principales source de variations au sein d'une même classe: l'éclairage, l'angle de vue ou la pose, et l'expression du visage. Généralement, ces trois paramètres engendrent des variations plus grandes que les variations entre classes distinctes, à paramètres constants !

Dans une première approche, on peut chercher à fixer les paramètres instantanés, en standardisant les modalités d'acquisition: expression neutre, sans lunettes, éclairage de face, etc. Une approche plus générale consiste à prendre tous ces paramètres en compte. Il en résultera un algorithme plus complexe, avec une phase d'apprentissage beaucoup plus importante (photos du même visage sous plusieurs éclairages, avec différentes expressions).

3.1. Approche géométrique de détection et de reconnaissance des visages

3.1.1. Localisation des visages a partir des images :

Des visages sont issus d'images de photos d'identités ; la position du visage est à peu près dans le même endroit. Lorsque le système localise les limites du visage automatiquement, il va distinguer entre ce qui se trouve à l'intérieur, qui est le visage lui-même "*foreground*", et ce qui est à l'extérieur des limites du visage qui est le "*background*". Après cette distinction, le système va opérer seulement sur le visage donc sur le "*foreground*".

3.1.2. Classification :

Une fois le visage localisé, le système peut le classifier puis extraire les points caractéristiques tels que les centres des yeux et la bouche, pour le faire passer au module d'identification. Le critère de classification est la proportion de la largeur du visage (l) sur sa hauteur (h) (Figure 1). Cette proportion implique la création de trois classes distinctes :

- La **première** contient les visages qui ont une forme ovale, tel que la proportion varie dans l'intervalle $[0.50...0.64]$.
- La **deuxième**, la classe normale, tel que la proportion varie dans l'intervalle $[0.66...0.70]$.
- La **troisième** classe celle des visages qui convergent vers la forme d'un cercle, tel que la proportion varie dans l'intervalle $[0.71...1.00]$.

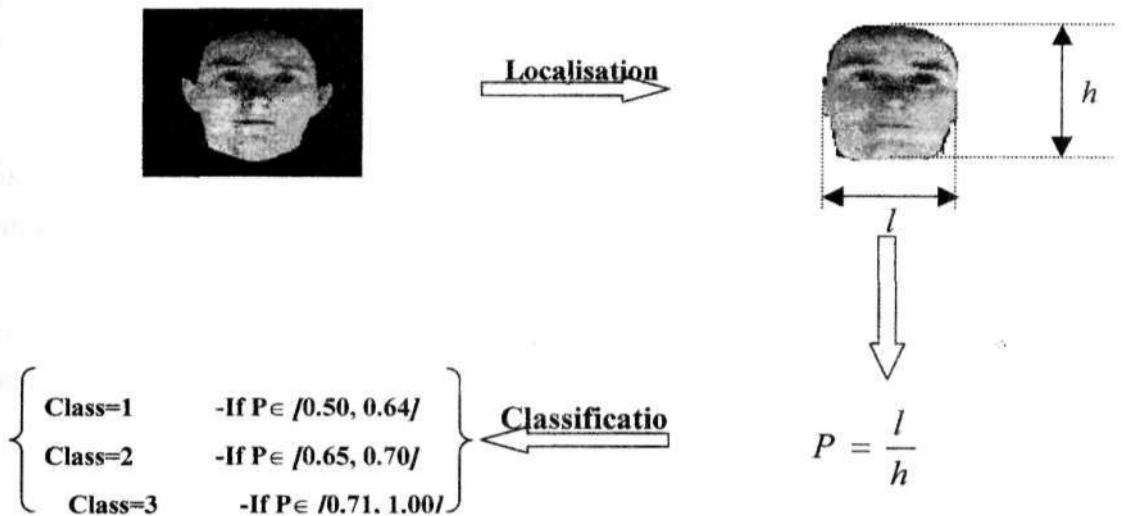


Figure 1. la procédure de classification

3.1.3. Détecteur des caractéristiques ("Feature Finder") :

La procédure précédente permet de localiser le visage, c'est à dire à partir d'une image (représentant la photo d'identité d'une personne) avoir la sous fenêtre contenant le visage. Reste à extraire les différentes caractéristiques pour ce visage. Ces caractéristiques sont être essentiellement : (Figure 2)

- Un point centre de l'œil gauche $\rightarrow P_1(x_1, y_1)$
- Un point centre de l'œil droit $\rightarrow P_2(x_2, y_2)$
- Un point centre de la bouche $\rightarrow P_3(x_3, y_3)$

- Un point centre du nez. $\rightarrow P_4(x_4, y_4)$

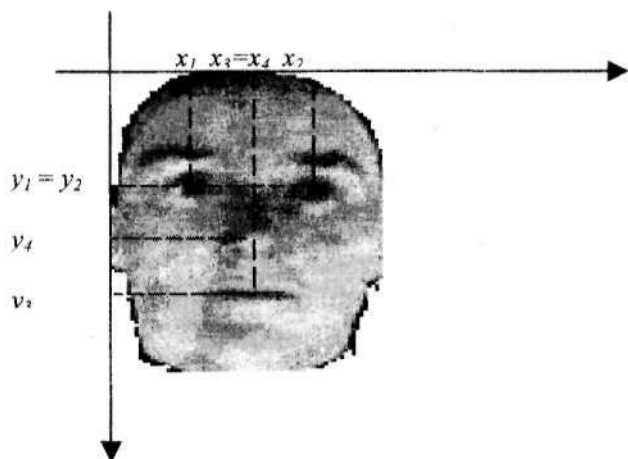


Figure 2 : Les points Caractéristiques du visage

3.1.4 .Extraction des caractéristiques faciales du visage

◆ Paramétrisation du visage :

Dans la section précédente, on a montré comment le système peut extraire le centre de l'œil. Par ce même principe le système peut extraire quatre points du visage.

La paramétrisation est basée sur les relations métriques entre ces quatre points.

Par paramétrisation, le visage est numérisé par un vecteur spécifiant les relations métriques existants entre ces différents points caractéristiques.

Le système fournit essentiellement cinq paramètres (Figure 3) :

- D_1 : Distance entre l'œil gauche et l'œil droit : $D_1 = x_2 - x_1$. (figure 2)
- D_2 , D_3 : Distances entre les yeux et le nez :
- D_4 , D_5 : Distances entre les yeux et la bouche

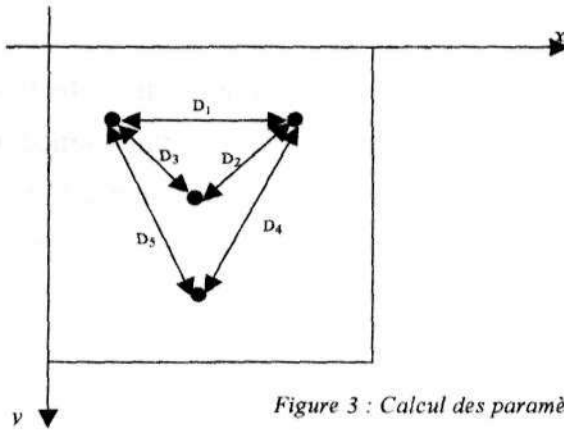


Figure 3 : Calcul des paramètres métriques

3.1.5. Identification

L'identification est le noyau de ce système : C'est la comparaison des caractéristiques du visage de l'*input* avec celles des visages de la base de données.

Cette comparaison est basée sur des *notions géométriques* telles que la *proportion des distances*, les *euclidiennes*. Si la proportion varie d'un rapport à l'autre, ce ne sont donc pas les mêmes distances alors le visage ne correspond pas au visage de l'*input*.(figure 4) .

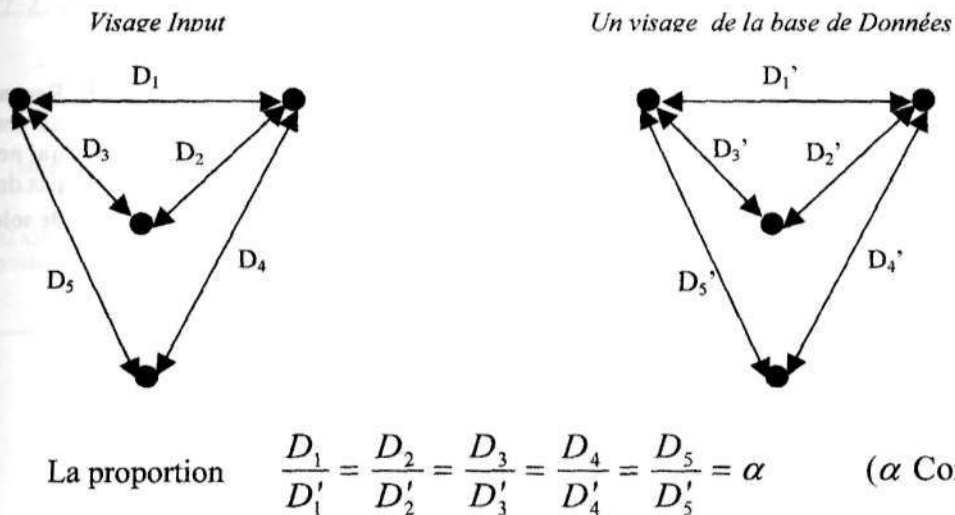


Figure 4 : Correspondance entre le visage input et un visage de la base de données

3.2. Approche de détection et de reconnaissance basée sur les réseaux neurones (NNS) artificiels

Le système est composé de cinq modules : module de détection, le module d'estimation de pose du visage, le module de reconnaissance d'émotion du visage et le module de détection de la présence ou absence des lunettes de soleil sur le visage. L'architecture du système ainsi que les principales étapes sont illustrées dans la figure 5.

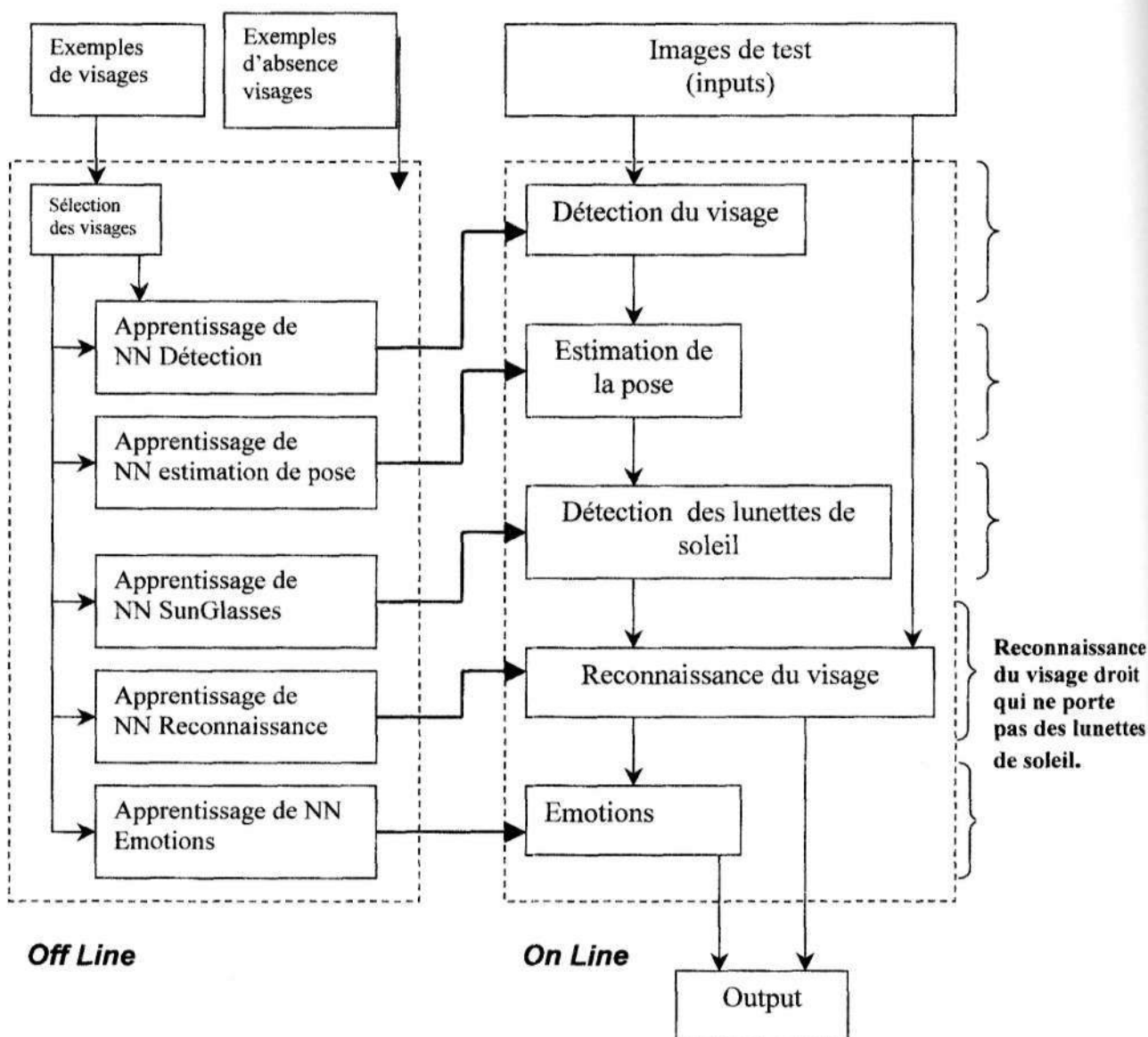


Figure 5 : Diagramme schématisant des principales étapes du système de détection et de reconnaissance basé sur les NN .

3.2.1. NN de reconnaissance :

Le NN est implémenté avec une couche cachée de 100 neurones et une couche pour l'input de 960 neurones qui reflètent le nombre total des pixels (32 x 30 pixels) et une couche d'output avec 20 neurones codés de façon à représenter les 20 personnes (figure 7). On fait l'apprentissage, puis le test du NN simple avec des images simples sans variations majeures ,prises dans des conditions favorables. On aboutit au NN après plusieurs tentatives d'exploration et de changement des paramètres de ce dernier et d'exploration des architecture des NNs.

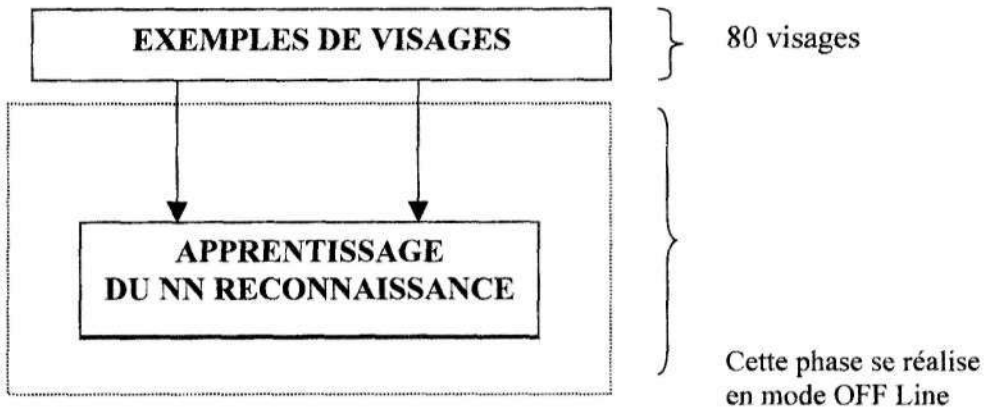


Figure 6 : Schéma de l'étape d'apprentissage de NN reconnaissance .

3.2.2. Architecture du NN de reconnaissance :

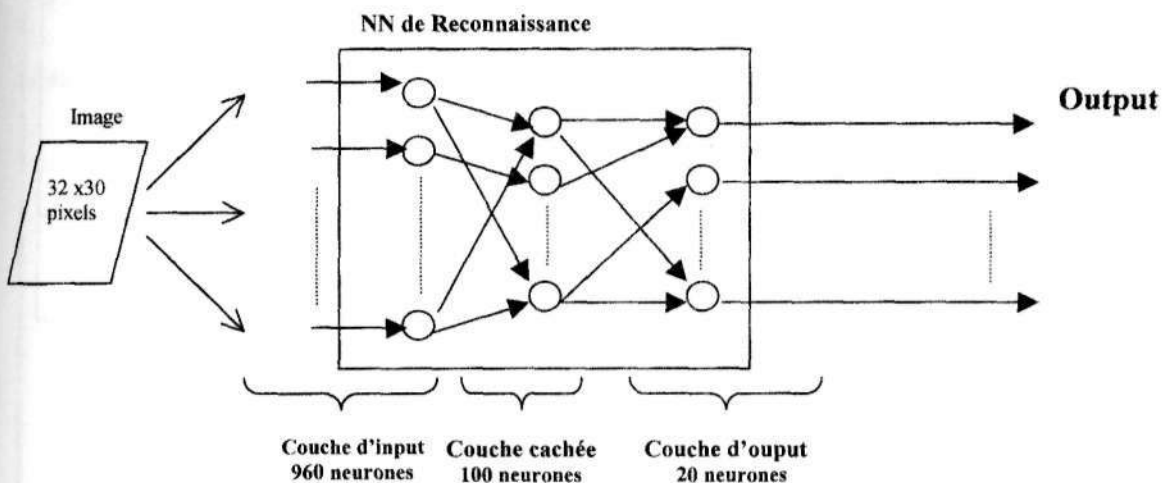


Figure 7 : Réseaux de neurones de reconnaissance .

3.2.3. Description de l'architecture :

Le NN de reconnaissance est un réseau de neurones à 2 couches ; la 1^{ère} couche correspond à la couche cachée avec 100 neurones, la 2^{ème} couche aux sorties avec 20 neurones. La fonction de transfert utilisée dans chaque neurone de cette architecture c'est la Log-Sigmoïd. La fonction Log-Sigmoïd est choisie parce que le rang des outputs du NN est de 0 à 1. Elle est parfaite pour l'apprentissage des valeurs booléennes comme outputs [Nnet,1998]. La couche cachée a 100 neurones; ce nombre est désigné par expérience. Les neurones des 2 couches sont massivement connectés au mode Feedforward qui utilise l'algorithme de Backpropagation (rétro-propagation) pour régler ses poids synaptiques.

4. Résultats de la simulation

4.1. La reconnaissance :

Le NN a montré des meilleurs résultats de reconnaissance surtout pour l'ensemble des images d'apprentissage d'autant plus que l'ensemble ne comporte que 38 images; le NN a reconnu toutes les images de l'ensemble (tableau 1).

Architecture (module)	Ensemble de test 217 images	Taux de Reconnaissance	Erreur de Reconnaissance
Le NN de Reconnaissance	Les images de MIT :38 images	97.36 %	01/38
	Les images de CMU : 80 Images (20 x 4 Visages) 20 personnes avec 4 images différentes (4 émotions)	93.75%	05/80
	Les images "Divers" :39 images	97.74%	02/39

Tableau 1 : Résultats sur la reconnaissance.

4.2. NN de détection

Après l'apprentissage du NN, on passe au teste, et à la récupération des résultats :

L'architecture (module)	Ensemble de test	Taux de détection	Erreur de détection
Le NN de détection	Les images de CMU 100 Images (80 avec visages +20 sans Visages)	100 % *	00/100
	Les images de MIT 58 images (38 avec visages +20 images sans Visages)	98.27%	01/58
	Les images divers (59 images) (39 avec visage + 20 images sans visages)	98.31%	01/59 **

Tableau 2 : Résultats sur la détection .

4.3. Résultats du test de performance du NN estimation de pose

Une fois que le NN Émotion a fait l'apprentissage, vient le rôle des exemples de test et de validation de l'architecture .

Systèmes	Ensemble de test	Taux de reconnaître d'Émotion	Erreur de reconnaître d'Émotion			
			neutre	heureux	triste	en colère
Le module NN De reconnaissance l'Émotion du visage dans l'approche neuronale	Les images de CMU 80 Images de visages (20 en état neutre, 20 heureux, 20 tristes, 20 en colère)	70 %	24/80			
			14/20	02/20	05/20	03/20
	Les images de MIT 38 images. (il sont tous en état neutre)	*** (le système a reconnue qu'ils sont tous en colère)	38/38			
			/	/	/	38/38
Les images diverses 39 images	61.54%	15/39				
		/	09/39	03/39	03/39	

Tableau 3 : Taux et erreur de reconnaissance de l'émotion du visage dans l'image d'input .

4.4. Apprentissage du NN Sunglasses

Systèmes	Ensemble de test	Taux de Détection	Erreur de détection
Le module NN de SunGlasses (détection des lunettes de soleil) dans l'approche neuronale	Les images de CMU 60 Images de visages (40 porte SunGlasses + 20 sans SunGlasses)	96.67%	02/60
	Les images de MIT 38 images. (ils sont tous sans SunGlasses, lunettes de soleil)	94.74%	02/38
	Les images diverses 39 images (ils sont tous sans lunettes de soleil, Sunglasses)	20.51% ***	31/39

Tableau 4 : Taux et erreur de détection des visages qui portent des lunettes de soleil.

4.5. Résultats de test de performance du NN Estimation de pose:

Systèmes	Ensemble de test	Taux d'estimation de la pose	Erreur d' Estimation de la pose			
			En Face	En haut	A gauche	A droite
Le module NN d' Estimation de la pose du visage dans l'approche neuronale	Les images de CMU 80 Images de visages (20 en face, 20 en haut , 20 à gauche,20 à droite,)	90.00%	08/80			
			00/20	00/ 20	03/20	05/20
			En Face	En haut	A gauche	A droite
	Les images de MIT 38 images. (il sont tous en vue de face)	73.68%	10/38			
			00/38	07/38	00/38	03/38
			En Face	En haut	A gauche	A droite
	Les images divers 39 image (il sont tous en vue de face)	30% ***	26/39			
			00/39	20/39	01/39	05/39
			En Face	En haut	A gauche	A droite

Tableau 5 : Taux & l'erreur d'Estimation de la pose du visage dans les deux ensemble de CMU et d' MIT.

5. Comparaison entre les caractéristiques des deux approches

Géométrique/Neuronale)

Approches géométrique classique	Approches neuronale
<p>Pré-traitement :</p> <ul style="list-style-type: none"> -Rendre l'image en Noir & blanc . (valeurs des pixels soit 1,soit 0) - Basé sur des photos en 2D ou 3D, transformées ensuite en 2D,pour qu'elles soient acceptable par le système. <p>Détection du visage :</p> <ul style="list-style-type: none"> -Manuelle. -Automatique :recherche du visage par le parcours pixel par pixel (les pixels de valeurs = 1) . (temps assez important de recherche) -Le traitement de bas niveau se réalise en noir et blanc (image binaire) <p>Reconnaissance :</p> <ul style="list-style-type: none"> -1^{er} : Extraire les points caractéristique du visage détecté. -2^{ième} :calcul des distances qui relient ces points caractéristiques. -3^{ième} :calcul des rapports de ces distances -4^{ième} :comparaison des rapport calculés avec les rapports des visages stockés dans la base de données visages. L'erreur la plus faible restitue le visage comme le plus proche du visage d'<i>input</i>, alors ce dernier est reconnu. 	<p>Pré-traitement :</p> <ul style="list-style-type: none"> -Pas de contraintes sur les valeur de pixels .(l'image en niveaux de gris, et même en couleur est accepté). -Nécessite un ensemble d'images d'apprentissage exhaustif et en nombre important et un ensemble d'images test pour valider l'architecture du NN conçu. <p>Détection du visage :</p> <p>Automatique : par simulation du NN (après apprentissage)et en <i>Online</i>. (temps négligeable) dans le cas de simulation, par contre un temps assez important dans le cas d'apprentissage(mode <i>Offline</i>)</p> <p>Reconnaissance :</p> <p>Se fait automatiquement en une seule tâche, par Le NN reconnaissance, et par simulation du NN en <i>On line</i>(après apprentissage bien sûre en <i>off Line</i>)</p> <p>Extension du système vers d'autre possibilités :</p> <ul style="list-style-type: none"> -Estimation de la pose en profondeur (par NN estimation pose) . -Détection de la présence ou l'absence des lunettes de soleil sur le visage (par NN SunGlsses) - Reconnaissance de l'émotion de la personne (par NN Emotions)

Remarques :	Remarques :
<p>-Une seule photo de la personne doit être disponible au niveau de la base de données images pour qu'il soit reconnu.</p> <p>-Aucune contrainte n'est posée au préalable sur les photos de l'input.</p> <p>-Identification non liée à la texture du visage</p> <p>-La base de données des images est théoriquement illimitée.</p> <p>-Applicable pour des photos contenant des personnes.</p> <p>-Les dimensions géométriques peuvent présenter une incohérence au niveau de l'identification.</p>	<p>- Nécessite d'une interprétation des outputs des NN pour avoir les résultats .</p> <p>-Plus de personnes à reconnaître, plus de neurones, ce qui implique une architecture ascendante du NN.</p> <p>-Image de dimension différente de 32 x 30, implique redimensionnement du vecteur d'input des NNs.</p>

6. Conclusion

Les NNs offrent l'avantage d'une forte tolérance aux pannes d'un certain nombre de neurones :le système continue à fonctionner avec des performances dégradées, et on parle de *dégradation harmonique ou progressive et non pas brusque* ,mais sans la possibilité d'interprétation du NN à cause son l'aspect boîte noire .Ils ne fournissent pas l'explication de leur résultats. Ce qui peut être gênant si on cherche à comprendre un phénomène.

Les approches neuronale et géométrique pourraient trouver autant d'applications, telles que :

- Identification des gens suspects à partir de leurs photos d'identité ou leurs portraits,
- Identification parmi les voyageurs (contrôle de l'immigration), afin de prévenir les actes terroristes,
- Au niveau des établissements de haute surveillance, l'autorisation n'est donnée qu'aux personnels du service identifiés au préalable (la photo est soit prise par une carte personnelle ou par acquisition du visage à partir d'une caméra de surveillance),
- Dans le sondage, pour éviter la redondance des personnes
- Recherche des enfants disparus (à partir d'une caméra de surveillance placée dans une zone piétonnière)

7. Bibliographie :

[Boukadoum,2002] :Mounir boukadoum, Phd, Spécialité utilisation des NN dans l'industrie , Conférence intitulée : ‘‘utilisation des réseaux neurones artificiels pour résoudre les problèmes complexes ,qui ne suit aucun modèles’’, à l'université Abou-bakr-Belkaïd de Tlemcen, par Mr Mounir Boukadoum , Mai 2002.

[Nnet,1998] : *Howard Demuth, Mark Beale, ‘‘Neural Network Toolbox User's Guide’’* For Use with MATLAB® by The MathWorks, Inc.1998.

[Daugman,1997] : *John DAUGMAN, Face and Gesture Recognition: Overview, IEEE PAMI, vol.19, no.7, July 1997.*

[Jeffrey,1999] : *Jeffrey S. Norris, ‘‘ Face Detection and Recognition in Office Environments’’, Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Electrical Engineering and Computer Science at the Massachusetts Institute of Technology , M.I.T, USA, May 21, 1999.*

[Mitchell,1997] :Tom Mitchell,Home work 3,‘‘neural net &Face images’’CMU,October 1997.

[Lacaille,1996] : Jérôme Lacaille, ‘‘ Réseaux de neurones’’ DIAM -- CMLA,ENS de Cachan, septembre 1996.

[Winner,1948] : N. Wiener. Cybernetics. NY Wiley, 1948.

[Hebb,1949]: D.D. Hebb. The Organization of Behavior. NY Wiley, 1949.

[Holland,1956]: N. Rochester, J.H. Holland, L.H. Habit, and W.L. Duda. *Tests on a cell assembly theory of the action of the brain, using a large digital computer.IRE Transactions on Information Theory, 2 : 80–93, 1956.*

[Ashby,1952]: W.R. Ashby. *Design for a Brain*. NY Wiley, 1952.

[Minsky,1954] : M.L. Minsky. *Theory of neural-analog reinforcement systems and its applications to the brain-model problem*. PhD thesis, Princeton NJ,1954.

[Gabor,1954] : D. Gabor. *Communication theory and Cybernetics*. *IRE Transactions on Circuit Theory*, 1 : 19–31, 1954.

[Taylor,1956] :W.K. Taylor. *Electrical simulation of some nervous system functional activities*. *Information Theory (London)*, 3 : 314–328, 1956.

[Rosenblatt,1958]: F Rosenblatt. *The perceptron : a probabilistic model for information storage and organization in the brain*. *Psychological Review*, 65 : 386–408, 1958.

[Minsky,1969] : M.L. Minsky and S.A. Paper. *Perceptrons*. MIT Press (Cambridge), 1969.

[Malsburg,1973] :C. von der Malsburg. *Self organization on orientation sensitive cells in the striate cortex*. *Kybernetik*, 14 : 85–100, 1973.

[Grossberg,1973] : S. Grossberg. *How does a brain build a cognitive code*. *Psychological Review*, 87 : 1–51, 1973.

[Hopfield,1982] : J.J. Hopfield. *Neural networks and physical systems with emergent collective computational abilities*. *Proceedings of the National Academy of Science of the USA*, 79 : 2554–2558, 1982.

[Kohonen,1984] :T. Kohonen. *Self-Organization and Associative memory*. Springer Verlag, 1984.

[Ackley,1983] : D.H. Ackley, G.E. Hinton, and T.J. Sejnowski. *A learning algorithm for Boltzmann machines*. *Cognitive Science*, 9 : 147–169, 1983.