Reconnaissance d'Expressions Faciales à l'Aide d'une Mémoire Associative Bidirectionnelle à Fonction de Sortie Chaotique

Karima Tabari[1], Mounir Boukadoum[1], Sylvain Chartier[2,3], Hakim Lounis[1]
[1]*Université du Québec à Montréal (canada)*
[2]*Université du Québec en Outaouais (Canada),* [3]*Institut Pinel de Montréal (Canada)*
ksamfoug@yahoo.fr boukadoum.mounir@uqam.ca chartier.sylvain@courrier.uqam.ca

Résumé

Une nouveau modèle de mémoire associative bidirectionnelle (BAM) est appliqué à la reconnaissance d'expressions faciales. Le modèle à été testé sur une base d'images dans différents contextes de bruits. Les résultats montrent une excellente performance face à du bruit gaussien ou au bruit par inversion de pixels, et une excellente capacité de catégorisation.

1. Introduction

L'explication des capacités d'association du cerveau humain est d'un grand intérêt pour les chercheurs en sciences cognitives et en intelligence artificielle. Les réseaux de neurones artificiels se sont révélés être un outil de modélisation puissant à cet égard. Depuis les travaux de pionnier de Kohonen dans les années 70 [1], plusieurs modèles de mémoires associatives ont été proposés, dont la mémoire BAM (Bidirectional Associative Memory) de Kosko [2] qui utilise une dynamique non linéaire en phase de rappel. Cependant, la plupart des modèles présentés dans la littérature apprennent hors-ligne, n'utilisent pas leur fonction de sortie durant la phase d'apprentissage, et peuvent uniquement traiter des patrons bipolaires, ce qui mine considérablement leur plausibilité biologique.

Récemment, un nouveau modèle de mémoire BAM a été introduit, qui utilise une dynamique non linéaire avec une fonction de sortie chaotique, tout en conservant une règle d'apprentissage d'inspiration hébienne [3]. En utilisant la fonction de sortie dans sa région non chaotique, les auteurs furent en mesure de traiter des patrons binaires et en tons de gris, avec une performance égale ou supérieure à celle des modèle de comparaison. Dans ce papier, nous décrivons une application de cette nouvelle mémoire associative à la classification des émotions chez l'humain, un problème important pour les systèmes tutoriels intelligents. Dans la prochaine section, l'architecture du modèle est décrite, suivie d'une description des expériences effectuées sur une banque d'images photographiques. Les résultats sont ensuite présentés et une discussion et conclusion clôt le document.

2. Architecture du réseau utilisé

Cette section résume l'architecture de la nouvelle mémoire BAM. Une description détaillée est fournie dans [3].

La topologie du réseau est indiquée à la figure 1. Il s'agit de deux réseaux de type Hopfield interconnecté tête-bêche, de manière à créer un flux récurrent d'information entre eux. Dans la figure, $X[0]$ et $Y[0]$ représentent les vecteurs-états initiaux, W et V sont les matrices des poids synaptiques entre les couches X et Y dans les deux directions, et t représente l'itération courante. Dans le but d'avoir au moins deux attracteurs, bornés par 1 et -1 (cas des patrons bipolaire), une carte logistique cubique a été utilisée [3]. Elle est définie à partir de l'équation

$$\frac{dz}{dt} = Gz\left(1-z^2\right)$$

où G est un paramètre général. En faisant une approximation et un changement de variables on aboutit à la fonction de sortie suivante [3] :

$$\forall i,\ldots,N, y_i(t+1) = \begin{cases} 1 & if \quad a_i > 1 \\ -1 & if \quad a_i < -1 \\ (\delta+1)a_i - \delta a_i^3 & else \end{cases}$$

$$\forall i,\ldots,N, x_i(t+1) = \begin{cases} 1 & if \quad b_i > 1 \\ -1 & if \quad b_i < -1 \\ (\delta+1)b_i - \delta b_i^3 & else \end{cases}$$

Dans ces équations, $y_i(t+1)$ et $x_i(t+1)$ représentent la sortie d'un neurone i de la couche X ou Y à l'instant $t+1$, et $a_i[t]$ et $b_i[t]$ sont les fonctions d'activation correspondantes à l'instant t ($a_i[t]=[W.X[t]]_i$, $b_i[t]=[V.Y[t]]_i$). Le paramètre δ est un paramètre général dont la valeur est cruciale pour la performance du réseau : si δ est trop grand, le réseau peut converger aussi bien vers des attracteurs stables, cycliques ou chaotiques. Le réseau exhibe une approche monotone à des états d'équilibres si la valeur de δ est comprise entre 0 et 0.5 [3] .

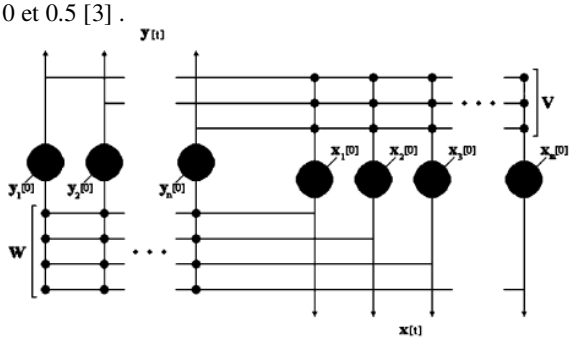


Figure 1 : Topologie de réseau utilisée

Le réseau utilise une règle d'apprentissage dérivée de l'approche hébienne/anti-hébienne [4][3] :

$$W(k+1)=W(k)+\eta[y(0)-y(t)][x(0)+x(t)]^T$$

$$V(k+1)=V(k)+\eta[x(0)-x(t)][y(0)+y(t)]^T$$

Dans ce travail, nous avons posé t=1 de manière à avoir

$$W(k+1)=W(k)+\eta[y(0)-y(1)][x(0)+x(1)]^T$$

$$V(k+1)=V(k)+\eta[x(0)-x(1)][y(0)+y(1)]^T$$

où

$X(1)= f(V[k]*Y(0))$ et $Y(1)=f(W[k]*X(0))$

Dans les équations précédentes, $\underline{W}$, V, $X(0)$ et $Y(0)$ sont tels que définis précédemment, η est le paramètre d'apprentissage et k est numéro du patron d'apprentissage. Comme la règle d'apprentissage inclut une boucle de retour à partir de la fonction de sortie non linéaire via $X(1)$ et $Y(1)$, elle permet au réseau d'effectuer un apprentissage en ligne et de faire contribuer la fonction de sortie à la convergence des poids de connections.

3. Expériences

3.1. Méthodologie

Nous avons utilisé la base de données California Facial Expressions (CAFE) [5] afin d'évaluer les capacités d'association de notre réseau. CAFE contient un ensemble d'images photographiques qui expriment diverses émotions sur le visage humain. Elle comprend 50 individus pour chacun desquels on dispose de 7 images reflétant les émotions (en colère, dégoûté, heureux, triste, craintif, neutre, surpris). La résolution de chaque photographie est 380x240 pixels en tons de gris.

Dans un premier temps, nous avons sélectionné 5 individus avec 4 émotions, pour un nombre total de 20 patrons d'apprentissage. Les associations suivantes furent établies (cf. Figure 2) :

- En colère (Anger) → lettre A
- Heureux (Happy) → lettre H
- Triste (Sad) → lettre S
- Neutre (Neutral) → lettre N

Pour les besoins de l'apprentissage, nous avons réduit la taille des images à 96x60 pixels et normalisé les 255 valeurs de tons de gris de chaque pixel en valeurs réelles entre +1 et -1. Les expériences firent menées avec les valeurs suivantes de paramètres : δ=0.1 et η=0.00115.

Nous avons étudié la performance du réseau pour des prototypes, et aussi face à plusieurs types de bruit : gaussien, inversion de pixels, et patrons partiellement masqués (cf. Figure 3). Dans tous les cas, le réseau a convergé après 15 à 17 époques d'apprentissage (300 à 340 présentations de stimuli), ce qui est remarquable étant donné la taille des vecteurs d'entrée (5700 éléments représentant les 95x60 pixels).

Nous avons aussi étudié les capacités de généralisation du réseau en utilisant quatre images de test dont deux appartenant à des individus nouveaux (2 et 4), et deux à des individus avec des trait faciaux modifiés (cf. figure 4). Ainsi l'image 3 représente un individu appartenant à la base d'apprentissage, mais avec des cheveux plus apparents, l'œil gauche plus fermé et les lèvres relâchées par rapport à l'image originale. Quant au sujet de la figure 1, il a les yeux plus ouverts.

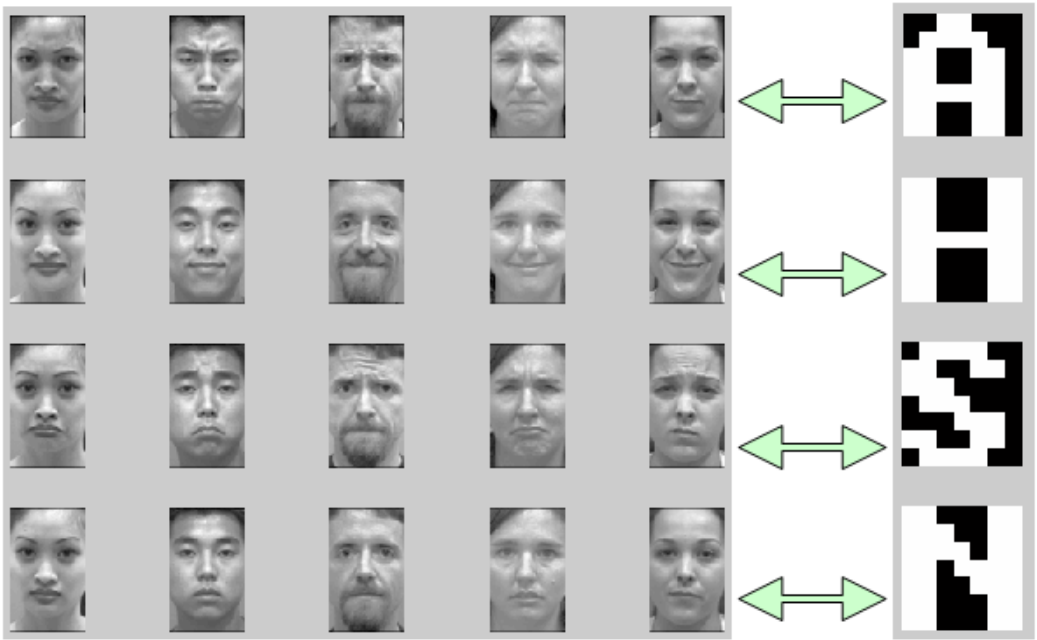
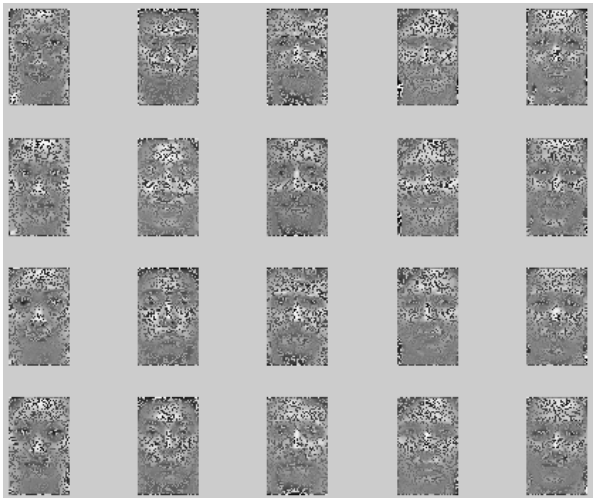


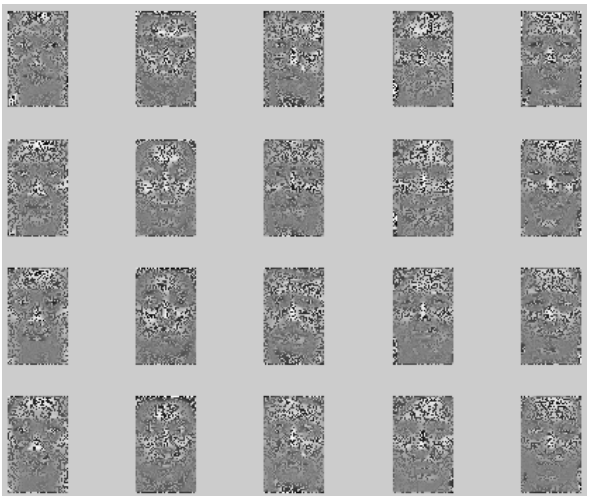
Figure 2. Association des Images et lettres utilisées lors de l'apprentissage du CNN_BAM. (a) images des individus (b) les émotions considérées

(a) Bruit gaussien (30 dBW)

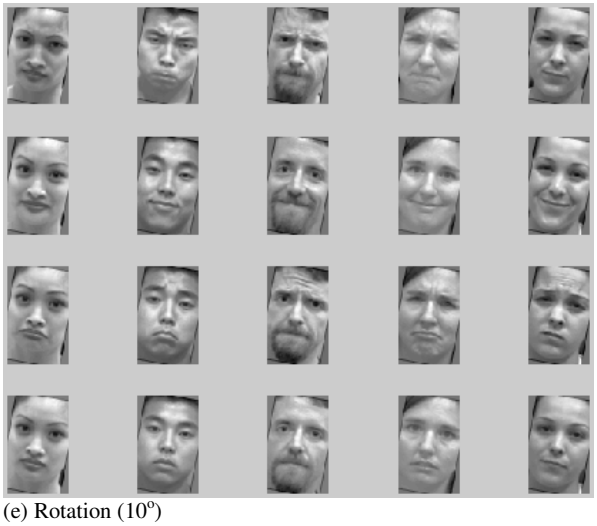
(b) Inversion de pixels (20%)



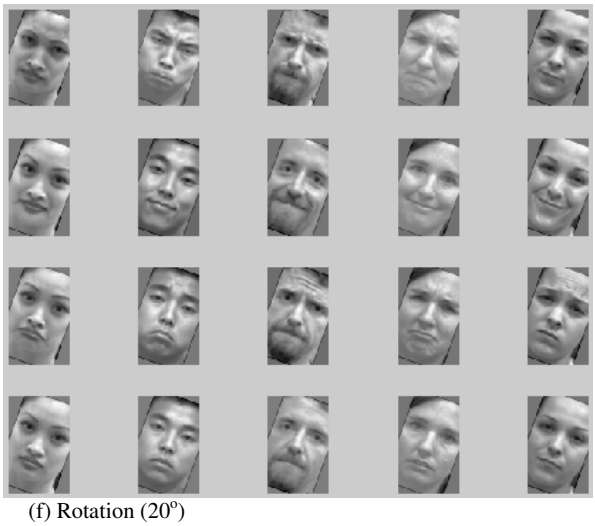
(c) Inversion de pixels (40%)



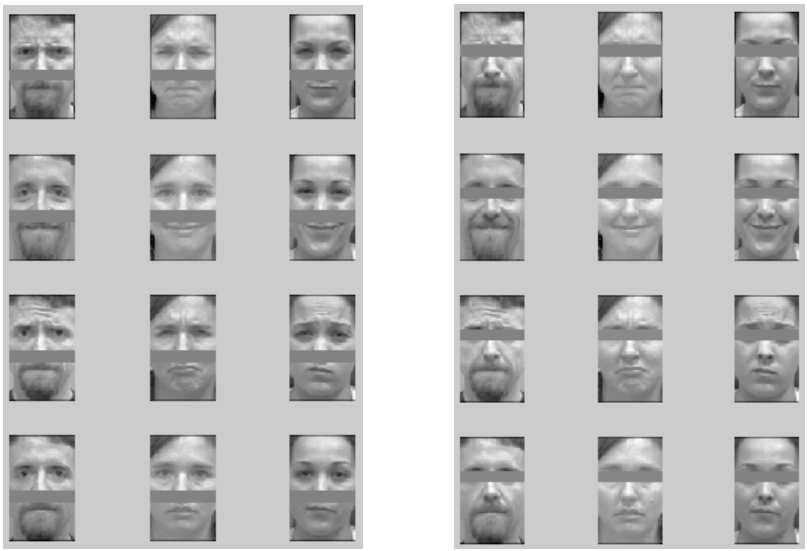
(d) Inversion de pixels (50%)



(e) Rotation (10^o)



(f) Rotation (20^o)



(g) Image masquée 1 (h) Image masquée 2

Figure 3. Effet de différents types de bruit sur les prototypes

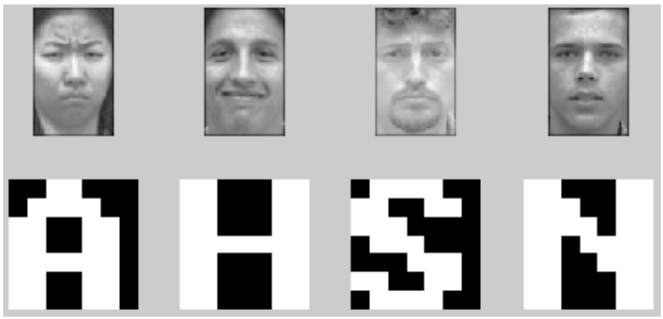


Figure 4. Paires de test pour la généralisation

Suite aux étapes précédentes, nous avons étendu notre étude aux 7 émotions représentées dans CAFE, et avons considéré l'ensemble des sujets féminins. Les 10 sujets ont été divisés en 9 sujets d'entraînement et un sujet de test en phase de rappel. Ce processus a été répété 10 fois en changeant à chaque expérience le sujet à exclure des patrons d'apprentissage.

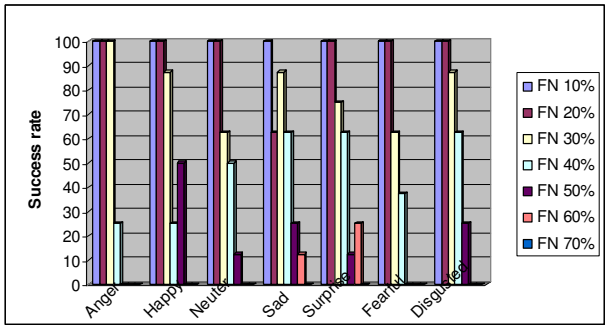
3.2. Résultats

Le tableau 1 donne les résultats obtenus dans le cas de 5 individus et 4 émotions. Il montre un taux de classification parfait pour les prototypes, le bruit gaussien, les images partiellement masquées, et l'inversion de pixels lorsque la proportion de pixels affectés est inférieure ou égale à 40%. Par contre, la performance du réseau a été faible pour les images tournées ou lorsque l'inversion de pixels était majeure. Quant aux images de la figure 4, elles ont toutes été classées correctement, bien que ne faisant pas partie de l'ensemble d'apprentissage.

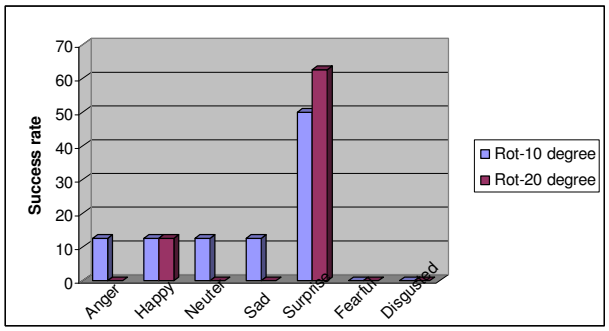
Les Figures 5a à 5c donnent les résultats pour l'ensemble des 7 émotions et les 10 sujets féminins, lorsqu'on soumet les photos à l'inversion de pixels, aux rotations, et au bruit gaussien. Les résultats sont similaires à ceux obtenus précédemment.

Tableau 1 Performance du réseau pour 5 individus et 4 émotions

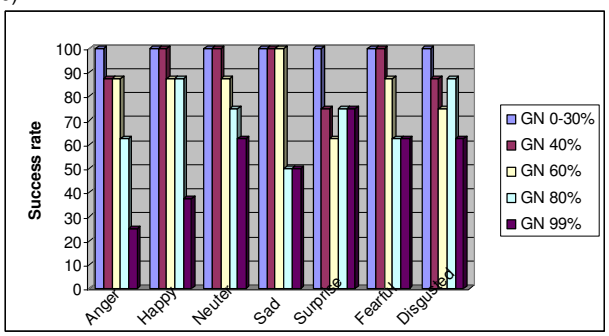
Type de bruit	Taux de classification réussie (%)
Nil	100
Gaussien (30 dBW)	100
Inversion de pixels (20%)	100
Inversion de pixels (40%)	100
Inversion de pixels (50%)	60
Inversion de pixels (60%)	15
Inversion de pixels (>60%)	0
Masque 1	100
Masque 2	100
Rotation (10°)	30
Rotation (20°)	35



a)



b)



c)

Figure 5. Performance du réseau pour l'ensemble complet d'émotions et tous le sujets féminins : a) condition d'inversion de pixels ; b) conditions de rotation l c) conditions de bruit gaussien.

4. Discussion et conclusion

On constate que pour le bruit gaussien, et le bruit par inversion de pixels d'intensité faible à modérée (<40%), le taux de reconnaissance est 100 %. Il est aussi intéressant de remarquer que pour les images masquées 2, le réseau a pu identifier la bonne lettre à associer malgré la suppression d'une partie importante du visage pour l'identification des émotions, la région oculaire.

Par ailleurs, les résultats pour l'ensemble complet des émotions et un nombre plus élevé de sujets révèlent une excellente capacité de mémoire. Cette propriété peut être mise à profit pour pallier à la faible performance du réseau pour des images tournées. Ces dernières peuvent être apprises comme des patrons distincts en fonction de l'angle de rotation.

Les résultats obtenus sont surprenants en égard à la simplicité relative de notre architecture. Ils se comparent favorablement avec des études qui ont utilisé des algorithmes substantiellement plus complexes. Par exemple, [6][7] utilise un perceptron multicouche en conjonction avec un classificateur à composantes principales. Malgré un prétraitement substantiel des données à catégoriser, incluant l'isolation des régions oculaires et buccales à des fins de catégorisation, le meilleur taux de classification correcte obtenu dans cette autre étude, pour des prototypes sans bruit, fut de 86 %.

En conclusion, le nouveau modèle de mémoire associative présenté est non seulement supérieur à d'autres mémoires BAM tel que montré dans [3][8], mais peut aussi surpasser la performance d'autres architectures neuronales pour certaines applications comme le montrent nos résultats.

5. Références

[1] T. Kohonen, "Correlation Matrix Memories," *IEEE Trans. Computers*, vol. C-21, pp. 353-359, 1972.

[2] B. Kosko, "Bidirectional Associative Memories," *IEEE Trans. SMC*, vol. 18, pp. 49-60, 1988.

[3] S. Chartier & M. Boukadoum, "A Bidirectional Heteroassociative Memory for Binary and Grey-Level Patterns," *IEEE Trans. Neural Networks*, vol. 18-2, March 2006.

[4] J. Bégin and R. Proulx, "Categorization in Unsupervised Neural Networks: The Eidos Model," *IEEE Trans. Neural Networks*, vol. 7, pp. 147-154, 1996.

[5] M.N. Dailey, G.W. Cottrell, and J. Reilly, California Facial Expressions (CAFE) [http://www.cs.ucsd.edu/users/gary/CAFE/] La Jolla, CA: Computer Science and Engineering Department, UCSD, 2001.

[6] Padgett, Curtis and Cottrell, Garrison W., Representing face images for emotion classification, In Advances in Neural Information Processing Systems 9, pp. 894-900. MIT Press, Cambridge, MA, 1997.

[7] Dailey, Matthew N., Cottrell, Garrison W., Padgett, Curtis, and Ralph Adolphs, "EMPATH: A neural network that categorizes facial expressions,"*Journal of Cognitive Neuroscience* 14(8):1158-1173, 2002.

[8] S. Chartier & M. Boukadoum, "A sequential dynamic heteroassociative memory for multistep pattern recognition and one-to-many association," *IEEE Trans. Neural Networks,* vol. 17-1, pp. 59-68, 2006.