ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 369-376

ANN for facial information processing: a review of recent approaches

B. Raducanu, M. Graña, A. D´Anjou, F.X. Albizuri

Dept. CCIA, UPV/EHU+ Apartado 649, 20080 San Sebastián
e-mail: {ccbradub, ccpgrrom, ccpdadaa, ccpalirx}@si.ehu.es

Abstract: We provide in this paper a guided introduction to the recent literature on the application of Artificial Neural Networks to the diverse aspects of facial information processing and their applications. We can not claim the revision to be complete because of the explosive growth of the field, but it will provide a framework to understand the relevance of the papers presented in the special session.

0 Introduction

There is a broad spectrum of applications of techniques that involve somehow the processing of facial information. Some of the most important fields of application are:

1. Security systems, where the face localization and recognition may play the role of a non intrusive test, although not very definitive. Robust authentication systems combine face recoginition with other identification sources, such as voice authentication. Applications range from restricted area access to broad area control (airports, stations).
2. Browsing of image databases or content-based queries based on face localization and recognition primitives. The access to mugshot databases for the identification of criminal suspects is a priviliged example.
3. Human computer interaction systems [57] that show a greater sensitivity to human to human interaction protocols, such as facial and gesture recognition [30] (emotion, gaze tracking [17]) or multisource recognition such as speech-reading [46, 47]. These interfaces may also involve the synthesis of speech, faces and facial dynamics.
4. The development of standards for image and video compression (MPEGIV) involve model based compression procedures [39].

Although difficult, it is possible to isolate a set of technical problems that arise when trying to provide solutions for the some or all of the above applications:

1. Feature extraction. Both model based geometric features and wholistic features have been used for recognition, image compression, database browsing, image reconstruction and other applications.
2. Face localization. Some of the sources of complexity of the localization task are:

+ This work is being supported by the Dept. Educación, Univ. e Inv. of the Gobierno Vasco, under project PI97/12. Bogdan Raducanu profits from a predoctoral grant of the UPV/EHU

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 369-376

- Relative size of the face in the scene.
- Background clutter. The texture of the background and clothes may produce false localizations.
- Partial oclusions and the occurrence of facial details such as glasses, moustaches and beards
- The existence of several scales produced by different distances to the camera, or image planes
- Diverse head poses, inclination and rotation produce very different visual objects (e.g. profile versus frontal views)

3. Tracking the face or the face features in image sequences [18]
4. Classification poses several specific difficulties in face recognition applications:
   - There are usually a large number of classes (subjects to be recognized) so the performance of the classifier must be robust against the growth of the number of classes
   - Classes may be added or deleted in real time (operation time) so that the classifier must be able to accept or forget subjects without degradation
   - Training patterns are scarce
   - The preprocessing [52] may introduce distortion artifacts that blurr the discrimination boundaries between individuals
   - In the case of applications such as lip-reading, the recognition involves also time dependent models, with the above limitations.

Previous revisions on facial recognition that can be found in the literature are [1-3]. In the following sections we will try to categorize the works appeared in the literature in recent years, with an emphasis in those that use Artificial Neural Networks (ANN). We can advance that the most popular architectures are the Multilayer Perceptron (MLP), the Radial Basis Functions (RBF) and the Self Organizing Map (SOM). As a rather general impression, we can say that most of the works are performed on Identity like images over relatively small datasets. The most common approach to obtain robustness against transformations is to train the networks with artificially generated images.

1 Face modelling and feature extraction

There are tow broad approaches to feature extraction: wholistic and geometrical. The most succesful wholistic method is the application of the Principal Components Analysis (PCA) transformation to the face image known as the eigenface approach [2, 4, 5, 19]. The MLP have been applied to compute wholistic features as the result of a compression or autoassociation topology [2], sometimes over the result of a previous PCA reduction [11]. Geometrical features are localized characteristics such as points, lines or shapes. Approaches in this category include the elastic net [2, 7] based on the computation of Gabor filters, the Banana Wavelets of [8], the use of steerable spatial filters [12-16, 23]. Some hybrid approaches are the Local Feature Analyis [24] and other combinations of shapes and gray level information [58, 59]. There is some accepted evidence that favor wholistic features (templates) against geometric ones [27, 40]. However, there are still many people working on the latter and on the sensitivity of the recognition to the accuracy of their measurements. [45].

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 369-376

Some approaches work of face elements: the SOM was used in [32] to categorize eyes, mouth and a reduced resolution face image, and in [28] to preprocess and to categorize the images. Finally there are several exotic attempts: the sonar echoes of faces [31], the simplified optical flow of moving faces [35], a vector field of the deformations of hand localized shape features and gray scale features [37], Hidden Markov Models [36], Circular Harmonic trasformation [42] (because of its optical implementation), structured lighting was proposed in [49] to obtain real time feature extraction. Finally, in a more classical mood, local autocorrelations [54, 60], high order correlations [55], and the fourier descriptors of the face profile [61] have been also proposed.

2 Face localization

Face localization is in a certain way a recognition problem, a two class classification problem (head vs non-head). Approaches based on wholistic features (image blocks) such as eigenfaces [19], Probabilistic Decision Based Neural Network [29], MLP [38] colvolve the image with a linear or non-linear filter estimated from training data. These methods are very sensitive to the statistical characteristics of the training patterns (i.e. the existence of appropriate negative examples). Besides they generalize poorly to varying image capture conditions (illumination, pose, scale). There are several neuronal approaches that can lead to efficient solutions: the Dynamic Scaling system [6]. The work presented in [72] takes an original approach to face localization, taking for granted the localization of eyes and mouth, they determine the grouping of the parts from the peaks of the oscillations of a recurrent dynamic network. Faces that exist in different image planes are detected by the Neural Network from the geometrical relationships between their elements. It is noteworthy that very interesting localization methods arise from the consideration of geometrical features [12-16, 23] computed by steerable spatial filters. Another promising venue of progress is the consideration of color [48, 50].

3 Face recognition

Once the appropriate features are extracted, the problem of face recognition is a classification problem. Approaches based on eigenfaces [2, 4, 5, 19] solve the classification using simple nearest neighbor approaches. Some attempts have used standard MLP's to solve this classification problem [2, 9, 10, 11, 32, 53]. However, its performance and the computational complexity of its training degrades as the number of classes (individuals) to be recognized increases. Also, the addition and deletion of individuals implies retraining the whole net. A modular system based on MLP is [26]. Other neural architectures have been applyed to classification for face recognition: the Hyper-BF [27], the convolution network [28], the Probabilistic Decision Based Neural Network [29], the optical dual-scale [44] that composes focus of attention on low resolution with recognition in high resolution, the Group-Based Adaptive Tolerance Trees (GAT) [51]. In [43, 73] the authors present an extension of the RBF taking the consensus of several networks called ERBF. Besides they use decision trees to further classify the output of the ERBF, producing an hybrid schema. In [74] the effect on the recall of a linear autoassociator of adding the edge images to the patterns learnt is tested.

The construction of pose-invariant recognition of faces is a rather difficult problem, changes in face image due to varying pose are much stronger than those due to personal characteristics. The straighforward approach consists in taking as training patterns images in all the potential poses [20]. A more sophisticated approach is to derive a three dimensional model of the head and compute the matching of the face image with synthetic views of the model. Such an approach is proposed in [21] based on tree-dimensional Principal Component Analysis of ·3D scanned heads that help to regulize the shape from shading problem in the case of face images.

4 Miscellanea and speechreading

The PCA approach has been used as a building block of a content-based query system for image databases [22], for face synthesis [25]. Multisensensorial person identification that employs RBF is described in [33], the fusion of acoustic and visual evidence for the same problem using a HyperBF architecture is described in [40]. In [34] an optical implementation of MLP able to recognize faces in real time is described. An active vision realized through an active retina is proposed in [56].

For Speech reading purposes, the success of Hidden Markov Models in the acoustic domain has motivated its application to the visual recognition [66] and to synthesis of realistic talking heads [68]. Some works such as [62-65] discuss the appropriate visual features to be used for lip-reading, while others discuss the recognition engine, among which the Time Delay Neural Network are one of the preferred methods of choice [57, 46, 47, 67]. Others concentrate try to ascertain the appropriate way to combine the information from the acoustic and visual estimuli [70, 71]. The work presented in [75] uses the SOM to determine the visual speech units (visemes) and tests two ways of composing Hidden Markov Models and Neural Networks (TDNN and JNN) on a French spelling task.

References

[1] R. Chelappa, C.L. Wilson, S. Sirohey "Human and machine recognition of faces: A survey" Proc. IEEE vol 83 pp705-741, May 1995
[2] J Zhang, Y Yan, M. Lades "Face recognition: Eigenface, Elastic Matching and Neural Nets" Proc IEEE vol 85, pp 1423-1435, Sept 1997
[3] D. Valentin, H. Abdi, A.J. O'Toole, G. W. Cottrell "Connectionist models of face processing: a survey" Pattern Recognition vol 27, pp1209-1230, 1994
[4] M.A. Turk, A.P. Pentland "Face recognition using eigenfaces" in Proc. IEEE Comp. Soc. Conf. Comp. Vision and Pattern Recog. 1991, pp 586-591
[5] A.P. Pentland, B. Mohaddan, T. Starner, M.A. Turk "View based and modular eigenspaces for face recognition" in Proc. IEEE Comp. Soc. Conf. Comp. Vision and Pattern Recog. 1994, pp 84-91
[6] G. Meierfrankenfeld, K. Kopecz "Size invariance by Dynamic Scaling in Neural Vision Systems" in ESANN'97, M. Verleysen (ed) dFacto press, Brussels, pp 55-60
[7] M. Lades, J.C. Vobrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, W. Konen "Distortion invariant object recognition in the Dynamic Link Architecture" IEEE trans. Comp. vol 42 pp.300-311, 1992

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 369-376

[8] N. Krüger, G. Peters "Object recognition with Banana Wavelets" in ESANN'97, M. Verleysen (ed) dFacto press, Brussels, pp 61-66

[9] G. W. Cottrell, M.K. Fleming "Face recognition using unsupervised feature extraction" Proc. Int. Conf. Neural Networks, Paris, 1990 pp.322-325

[10] B.A. Golomb, D.T. Lawrence, T.J. Sejnowski "Sexnet: a neural network identifies sex from human faces" Adv. Neural Information Processing vol3 D.S. Touretzky, R. Lippman (eds), Kaufman, pp.572-577, 1991

[11] D. DeMers, G.W. Cottrell "Non-linear dimensionality reduction" in Adv. Neural Information Processing vol 5, S. J. Hanson, J.D. Cowan, C.L. Giles (eds), Kaufman, pp.580-587, 1993

[12] K. C. Yow, R. Cipolla, "Finding Initial Estimates of Human Face Location", in 2nd Asian Conf. on Computer Vision, Vol. 3, pp. 514-518, 1995,

[13] K. C. Yow, R. Cipolla, "Towards an Automatic Human Face Localization System", in Proc. British Machine Vision Conference, , pp. 701-710, 1995

[14] C.H. Yow, R. Cipolla "Detection of human faces under scale, orientation and viewpoint variations" 2nd Int. Conf. Autom. Face and Gesture Recogn., Vermont, USA 1996

[15] C.H. Yow, R. Cipolla "A probabilistic framework for perceptual grouping of features for human face detection" 2nd Int. Conf. Autom. Face and Gesture Recogn., Vermont, USA 1996

[16] C.H. Yow, R. Cipolla "Scale and orientation invariance in human face detection" British Mach. Vision Conf. 1996

[17] A.H. Gee, R. Cipolla "Determining the gaze of faces in images" Univ. Cambridge, Dept. Engineering Report TR 174, March 1994

[18] A.H. Gee, R. Cipolla "Fast visual tracking by temporal consensus" Univ. Cambridge, Dept. Engineering Report TR 207, Feb. 1995

[19] M. Turk, A. Pentland "Eigenfaces for recognition" J. Cognitive Neuroscience vol 3 pp.71-86, 1991

[20] D. J. Beymer "Face recognition under varying pose" MIT, AI Lab Memo No. 1461, 1993

[21] J.J. Atick, P. A. Griffin, A.N. Redlich "Statistical approach to shape from shading: reconstruction of three dimensional face surfaces from single two dimensional images" Neural Computation vol 8, pp.1321-1340, 1996

[22] A. Pentland, R.W. Picard, S. Sclaroff "Photobook: content-based manipulation of image databases" Int. Jou. Comp. Vision vol 18 pp. 233-254, 1996

[23] R. P.N. Rao, D. Ballard "Natural basis functions and topographic memory for face recognition" Proc. Int. Joint. Conf. Artif. Int, pp.10-17, 1995

[24]P. Penev, J.J. Atick "Local feature analysis: a general statistical theory for object representation" Network: Comp. Neural Sys. vol 7 pp477-500, 1996

[25] R. Brunelli, O. Mich, "SpotIt! An Interactive Identikit System", in Graphical Models and Image Processing, Vol. 58, No. 5, Sept. 1996, pp. 399-404

[26] T. Yahagi, H. Takano "Face recognition using Neural Networks with multiple combinations of categories" Systems and Comp. in Japan vol26, pp55-65, 1995

[27] R. Brunelli, T. Poggio "Face recognition: features versus templates" IEEE trans. Patt. Recog. Mach. Int. vol 15 pp1042-1052, 1993

[28] S. Lawrence, C. L. Giles, A. C. Tsoi, A. D. Back, "Face Recognition: A Convolutional Approach", in IEEE Trans. on Neural Networks, Vol. 8, No. 1, , pp. 98-113 Jan. 1997

[29] S.-H. Lin, S. Y. Kung, L.-J. Lin, "Face Recognition/Detection by Probabilistic Decision-Based Neural Network", in IEEE Trans. on Neural Networks, Vol. 8, No. 1, , pp. 114-132 Jan. 1997

[30] F.K.H. Quek "Eyes in the interface" Image and Vision Computing vol13 pp.511-525

[31] I.E. Dror, F. L. Florer, D. Rios, M. Zagaeski "Using artificial bat sonar neural networks for complex pattern recognition: Recognizing faces and the speed of a moving target" Biol. Cybern. vol 74 pp.331-338, 1996

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 369-376

[32] A.J. Luckman, N.M. Allinson, A.W. Ellis, B.M. Flude "Familiar face recognition: A comparative study of a connectionist model and human performance" Neurocomputing vol 7, pp3-27, 1995
[33] R. Brunelli, D. Falavigna,T. Poggio, L. Stringa "Automatic person recognition by acoustic and geometric features" Mach. Vision and Appl. vol 8 pp317-325, 1995
[34] B. Javidi, J. Li, Q. Tang "Optical implementation of neural networks for face recognition by the use of nonlinear joint transform correlators" Appl. Opt. vol 34 pp.3950-3962
[35] P. Kruizinga, N. Petkov, "Optical Flow applied to person Identification", in Proceedings of EUROSIM'94,L.Dekker, W. Smit, J.C: Zuidervaat (eds) pp. 871-878
[36] F. Samaria, F. Fallside, "Face Identification and Feature Extraction Using Hidden Markov Models" Cambridge Univ. Eng. Dept
[37] D. Beymer "Vectorizing face images by interleaving shape and texture computations" MIT AI Memo No 1537
[38] K. K. Sung, T. Poggio, "Example-based Learning for View-based Human Face Detection", in *A.I. Memo* No. 1521, Dec. 1994
[39] D.E. Pearson "Developments in model-based video coding" Proc IEEE vol 83 pp.892-906, June 1995
[40] R. Brunelli, D. Falavigna "Person identification using multiple cues" IEEE Trans. Patt. Anal. Mach. Int. vol 17 pp.955-966, 1995
[41] D. L. Swets, J Weng "Using discriminat eigenfeatures for image retrieva" IEEE Trans. Patt. Anal. Mach. Int. vol 18 pp.831-836, 1996
[42] S. Chang, S.A. Boothroyd, J. Chrostowski "Partial rotation invariant pattern maching and face recognition with a joint transform correlator" Appl. Opt. vol 36 pp2380-2387, 1997
[43] S. Gutta, H. Wechsler "Face recognition using hybrid classifiers" Patt. Recog. vol 30, pp.539-553, 1997
[44] M. A. Neifeld "Optical dual-scale architecture for neural image recognition" Appl. Opt. vol 34, pp.5920-5927, 1995
[45] N. Roeder, X. Li "Accuracy analysis for facial feature detection" Patt. Recog. vol 29 pp143-157, 1996
[46] D. Stork, M.E. Hennecke "Speechreading by humans and machines" Springer Verlag , Berlin 1996
[47] C. Benoît, R. Campbell Proc. Workshop on Audio-visual speech processing, Rhodes, ESCA press, Sept 1997
[48] C.H. Lee, J.S. Kim, K.H. Park "Automatic human face location in a complex background using motion and color information" Patt. Recog. vol 29 pp.1863-1876, 1996
[49] R.A. Andrade, B.S: Gilbert, D.W.W. Dawson, C. Hart, SP. Kozaitis, J.H. Blatt "Real time optically processed face recognition system based on arbitrary moire contours" Opt. Eng. vol 35 pp
[50] E. Littmann, H. Ritter "Adaptive color segmentation- A comparison of neural and statistical methods" IEEE trans. Neural Networks vol 8 pp.175-185, 1997
[51] M. Zhang, J. Fulcher "Face recognition using Artificial Neural Network Group-Based Adaptive Tolerance (GAT) Trees" IEEE trans. Neural Networks vol 7 pp.555-567, 1996
[52] P.J. Philips, Y. Bardi "Efficient illumination normalization of facial images" Patt. Recog. Letters vol 17 pp.921-927, 1996
[53] N. Intrator, D. Reisfeld, Y. Yeshurun "Face recognition using a hybrid supervised/unsupervised neural network" Patt. Recog. Lett. vol 17 pp. 67-76, 1996
[54] F. Goudail, E. Lange, T. Iwamoto "Face recognition system using local autocorrelations and multiscale integration" IEEE PAMI vol 18 pp.1024- 1028
[55] M. Kreutz, B. Volpel, H. Janssen "Scale invariant image recognition based on higher order autocorrelation features" Patt. Recog. vol 29 pp.19-26, 1996
[56] M. Tistarelli "Active space variant object recognition" Im. Vision Comp. vol 13 pp.215-226, 1995
[57] A. Waibel, M.T. Vo, P. Duchnowski, S. Manke "Multimodal interfaces" Artif nt. Rev. vol 10 pp.299-319, 1996

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 369-376

[58] X.G. Jia, M.S. Nixon, "Extending the feature vector for automatic face recognition" IEEE PAMI vol 17 pp.1167-1176, 1995

[59] A. Lanitis, C.J. Taylor, T.F: Cootes "Automatic face identification system using flexible appearance models" Im. Vision Comp. vol 13, pp. 393-401

[60] R. Brunelli, S. Messelodi " Robust estimation of correlation with applications to computer vision" Patt. Recog. vol 28 pp833-841, 1995

[61] A.A. Somaie, S.S. Ipson "A human face profile identification system using 1D real fourier descriptors" Int. Jou. Infrared Millimeter Waves ,vol 16 pp.1285-1298, 1995

[62] M. S. Gray, J. R. Movellan, T. J. Sejnowski "Dynamic features for visual speechreading: A systematic comparison" NIPS'96

[63] B. Dalton, R. Kaucic, A. Blake "Automatic Speechreading using dynamic contours" pp. 373- 382 in [46]

[64] J. Luettin, A. Thacker, S. W. BEET "Active shape models for visual speech feature extraction" pp383-390 in [46]

[65] T. Coianiz, L. Torresani, B. Caprile "2D deformable models for visual speech analysis" pp.391-398 in [46]

[66] B. Jacob, N. Parlangeau "Audio-visual speech recognition and segmental master-slave HMM" pp.49-52 in [47]

[67] F. Lavagetto, P. Lavagetto "Time Delay Neural Networks for articulatory estimation from speech: suitable subjective evaluation protocols" pp. 437-444 in [46]

[68] E. Yamammoto, S. Nakamura, K. Shikano "Speech to lip movement synthesis by HMM" pp.137-140

[69] P.L. Silsbee, Q. Su " Audiovisual sensory integration using hidden markov models" pp.489-504 in [46]

[70] A. Rogozan-Foucault, P. Deleglise, M. Alissali "Adaptive determination of audio and visual weights for automatic speech recognition" pp. 61-64 in [47]

[71] J. Robert-Ribes, M. Piquemal, J. Swartz, P. Escudier "Exploiting sensor fusion architectures and stimuli complementarity in AV speech recognition" pp.193-210 in [46]

[72] S. Oka, M. Kitabata, Y. Ajioka, Y. Takefuji "Grouping complex face parts by nonlinear oscillations" this volume

[73] S. Gutta, H. Wechsler "Facial image retrieval using sequential classifiers" this volume

[74] E. Drege, F. Yang, M. Paindavoine, H. Abdi "Face Recognition: Pre-processing techniques for Linear Autoassociators" this volume

[75] A. Rogozan, P. Deléglise "Hybrid Hidden Markov Model/Neural Network Models for Speechreading" this volume