Some applications of morphological neural networks

Manuel Graña, Bogdan Raducanu,
University Basque Country,
Dept. CCIA,
Paseo Manuel de Lardizabal, 20013 San Sebastian, Spain
e-mail: ccpgrrom@si.ehu.es

Abstract

The Heteroassociative Morphological Memories are a recently proposed neural network architecture based on the shift of the basic algebraic framework. They possess some robustness to specific noise models (erosive and dilative noise). Here we report on going work on their applicatiob to the tasks of face localization in grayscale images and visual self-localization of.a mobile robot.

1 Introduction

Morphological Associative Memories is a novel kind of neural network architectures recently proposed in [2], [1]. In these networks, the operations of multiplication and addition are replaced by addition and maximum (or minimum), respectively. In [2] and [1] the construction of the Heteroassociative and Autoassociative Morphological Memories (HMM and AMM) is done following the isomorphy with the construction of the Heteroassociative and Autoassocitive Hopfield Memories exchanging the matrix product for the min/max matrix product. The use of the minimum or maximum operator determines the erosive or dilative character of the morphological memory. The memory capacity of the AMM is not bounded by conditions of orthogonality of the input patterns. The memory capacity of the HMM, however, is conditioned to some kind of max/min orthogonality relations between the patterns. The sensitivity of both AMM and HMM to erosive and dilatative noise has been characterized. The construction of a robust HMM (insensitive to both erosive and dilative noise) is decomposed in the construction of an AMM on the input pattern kernels and an HMM that maps the input pattern kernels into the output patterns. The inconvenient of this approach, besides the difficulties in the definition of the pattern kernels, lies in the extremely high storage and computational demands imposed by the construction of an AMM of any practical utility. For this reason, in this paper the approach taken is that of defining an HMM based on eroded/dilated versions of the input at several scales in a Scale-Space framework [10].

Face detection can be defined as the problem of deciding the presence of a face in the image. Multiple face detection is the generalization to the presence of a set of several faces in the image. Multiple face detection is usually dealt with by solving many single face detection problems posed over a set of overlapping subimages extracted from the image by a sliding window. Face localization is the problem of giving the coordinates in the image of the detected faces. Face localization is answered giving the coordinates of the positive responses to the single face detection instances in the multiple face detection. From a statistical pattern recognition perspective, face detection can be considered as a two-class classification problem: given an image, it must be decided if it belongs either to the face class or to the non-face class. The main difficulty then is the appropriate characterization of the non-face class. Some works, based on neural networks, [3], [4] [6] do it through a bootstrap-

0-7803-7044-9/01/$10.00 ©2001 IEEE

ping training strategy. Others, like the ones based on linear subspace transformation like the Principal Component Analysis (PCA) [5] compute the likelihood of the face class on the basis of the distance to the face subspace characterized by the eigenfaces.. Geometrical approaches try to fit an ellipse to the face contour [7] or to detect some face elements and verify their relative distances. Detection of face elements is difficult and a research subject by itself. Finally, approaches based on color processing [8] are very easy to realize, although prone to give high false positives rates. A sensible approach to more robust face localization is the combination of several methods into a multi-cue system. Recent works in this line, integrating geometrical and color-based approaches, are reported in [9]. In this spirit we propose our work as another verification tool.

Self-Localization is the ability to determine the spatial position and orientation of the robot using the information provided by its sensors[11, 12, 13, 14, 15, 16]. Visual self-localization based on the images provided by on-board cameras is usually based on the detection of some predetermined landmarks [11, 13, 15] specifically designed to be easily recognized in real time. Landmarks either provide a complementary information to the internal state estimation or are taken as a direct source for position estimation (i.e.: via triangulation or via Bayesian classification). The stated goal is to recognize, with some degree of robustness, several scenes that characterize predetermined robot placements and orientations. Robustness must cope with some variations in lighting and small rotations and translations of the images due to the uncertainty of the robot position, which, in its turn, is due to the uncertainties in the motion of the robot. The set of views is coded using a binary valued vector, under a straightforward orthogonal binary codification. Thus the Heteroassociative network takes as input an image of an indoor scene and gives as output a binary vector that encodes the view.

The paper is structured as follows. In section 2 we review the formal definition of HMM together with their properties. In section 3 and 4 we present results on face localization and self-llocalization. And finally, in section 5, we present our conclusions and future goals.

2 Heteroassociative Morphological Neural Network

The work on Morphological Neural Networks stems from the consideration of an algebraic lattice structure $(\mathbb{R}, \vee, \wedge, +)$ as the alternative to the usual $(\mathbb{R}, +, \cdot)$ framework for the definition of Neural Networks computation [2], [1]. The operators $\vee$ and $\wedge$ denote, respectively, the discrete min and min operators (resp. sup and inf in a continuous setting). The approach is termed morphological neural networks because $\vee$ and $\wedge$ are the basic operators for the morphological erosion and dilation.

Following the analogy, given $(X,Y) = \left\{(\mathbf{x}^\xi, \mathbf{y}^\xi); \xi = 1, .., k\right\}$, a set of pairs of input/output patterns the heteroassociative neural network built up as $W = \sum_\xi \mathbf{y}^\xi \cdot (\mathbf{x}^\xi)'$ becomes in the setting of morphological neural networks:

$$W_{XY} = \bigwedge_{\xi=1}^{k} \left[\mathbf{y}^\xi \times \left(-\mathbf{x}^\xi\right)'\right] \quad (1)$$

$$M_{XY} = \bigvee_{\xi=1}^{k} \left[\mathbf{y}^\xi \times \left(-\mathbf{x}^\xi\right)'\right] \quad (2)$$

where $\times$ is any of $\boxed{\vee}$ or $\overline{\wedge}$. It follows that the weight matrices are lower and upper bounds of the max/min products $\forall\xi; W_{XY} \leq \mathbf{y}^\xi \times \left(-\mathbf{x}^\xi\right)' \leq M_{XY}$ and therefore the following bounds on the output patterns hold $\forall\xi; W_{XY} \underline{\vee} \mathbf{x}^\xi \leq \mathbf{y}^\xi \leq M_{XY} \overline{\wedge} \mathbf{x}^\xi$, that can be rewritten $W_{XY} \underline{\vee} X \leq Y \leq M_{XY} \overline{\wedge} X$.

A matrix A is a $\underline{\vee}$-perfect ($\overline{\wedge}$-perfect) memory for (X,Y) if $A \underline{\vee} X = Y$ ($A \overline{\wedge} X = Y$). It can be proven that if A and B are $\underline{\vee}$-perfect and $\underline{\vee}$-perfect memories for (X,Y) then

$$A \leq W_{XY} \leq M_{XY} \leq B \text{ and } W_{XY} \underline{\vee} X = Y = M_{XY} \overline{\wedge} X. \quad (3)$$

Conditions for perfect recall on the memories are given by a theorem that states that W_{XY} is $\underline{\vee}$-perfect if and only if $\forall\xi$ the matrix $\left[\mathbf{y}^\xi \times \left(-\mathbf{x}^\xi\right)'\right] - W_{XY}$ contains a zero at each row. Similarly M_{XY} is $\overline{\wedge}$-perfect if and only if $\forall\xi$ the matrix $\left[\mathbf{y}^\xi \times \left(-\mathbf{x}^\xi\right)'\right] -$

M_{XY} contains a zero at each row. These conditions are rewritten for W_{XY} and M_{XY} respectively as follows:

$$\forall\gamma\forall i\exists j; x_j^\gamma = \bigvee_{\xi=1}^{k}\left(x_j^\xi - y_i^\xi\right) + y_i^\gamma \quad (4)$$

$$\forall\gamma\forall i\exists j; x_j^\gamma = \bigwedge_{\xi=1}^{k}\left(x_j^\xi - y_i^\xi\right) + y_i^\gamma. \quad (5)$$

Finally, let it be $\widetilde{\mathbf{x}}^\gamma$ a noisy version of $\mathbf{x}^\gamma$. If $\widetilde{\mathbf{x}}^\gamma \leq \mathbf{x}^\gamma$ then $\widetilde{\mathbf{x}}^\gamma$ is an eroded version of $\mathbf{x}^\gamma$, or $\widetilde{\mathbf{x}}^\gamma$ is subjected to erosive noise. If $\widetilde{\mathbf{x}}^\gamma \geq \mathbf{x}^\gamma$ then $\widetilde{\mathbf{x}}^\gamma$ is a dilated version of $\mathbf{x}^\gamma$, or $\widetilde{\mathbf{x}}^\gamma$ is subjected to dilatative noise. Morphological memories are very sensitive to these kinds of noise. The conditions for the perfect recall of $\mathbf{x}^\gamma$ given a noisy copy $\widetilde{\mathbf{x}}^\gamma$ for W_{XY} , that is, the conditions under which $W_{XY} \veebar \widetilde{\mathbf{x}}^\gamma = \mathbf{y}^\gamma$ are as follows:

$$\forall j; \widetilde{x}_j^\gamma \leq x_j^\gamma \vee \bigwedge_i \left(\bigvee_{\xi\neq\gamma}\left(y_i^\gamma - y_i^\xi + x_j^\xi\right)\right) \quad (6)$$

$$\forall i\exists j_i; \widetilde{x}_{j_i}^\gamma = x_{j_i}^\gamma \vee \left(\bigvee_{\xi\neq\gamma}\left(y_i^\gamma - y_i^\xi + x_{j_i}^\xi\right)\right).$$

Similarly for the perfect recall of $\mathbf{x}^\gamma$ given a noisy copy $\widetilde{\mathbf{x}}^\gamma$ for M_{XY}, that is, the conditions under which $M_{XY} \barwedge \widetilde{\mathbf{x}}^\gamma = \mathbf{y}^\gamma$ are as follows:

$$\forall j; \widetilde{x}_j^\gamma \leq x_j^\gamma \wedge \bigvee_i \left(\bigwedge_{\xi\neq\gamma}\left(y_i^\gamma - y_i^\xi + x_j^\xi\right)\right) \quad (7)$$

$$\forall i\exists j_i; \widetilde{x}_{j_i}^\gamma = x_{j_i}^\gamma \wedge \left(\bigwedge_{\xi\neq\gamma}\left(y_i^\gamma - y_i^\xi + x_{j_i}^\xi\right)\right).$$

These conditions (7), (6) are the basis for our approach. The conditions in (6) and (7) state that the matrix W_{XY} is robust against controlled erosions of the input patterns while the matrix M_{XY} is robust against controlled dilations of the input patterns. Therefore if we store in the W matrix a set of eroded patterns, the input could considered as a dilation of the stored pattern most of the times. The dual assertion holds for the M matrix. It also holds that when the output of both M and W memories are the same, then the output of both corresponds to the desired output. This holds in the case of interactions between the stored patterns. We will consider these matrices as approximations to the ideal memory of all the distorted versions of the input data, so that their output is an approximation to the response of this ideal memory. We applye a scale space approach to increase the robustness of the process.

Given a set of input patterns X and a set of output class enconding Y. We built a set of HMM $\{M_{XY}^\sigma, W_{XY}^\sigma; \sigma = 1, 2, ...\mathrm{s}\}$ where each M_{XY}^σ is constructed from outputd and the input patterns eroded with an spherical structural object of scale σ, and each W_{XY}^σ is constructed from the outputs and input patterns dilated with an spherical structural object of scale σ. Given a test input pattern $\mathbf{x}$, the memories at the different scales are applied giving $\mathbf{y}^M = \bigcup_{\sigma=1}^{\mathrm{s}}\left(M_{XY}^\sigma \barwedge \mathbf{x}\right)$ and $\mathbf{y}^W = \bigcup_{\sigma=1}^{\mathrm{s}}\left(W_{XY}^\sigma \veebar \mathbf{x}\right)$. The final output is the intersection of these multiscale responses:

$$\mathbf{y} = \mathbf{y}^M \bigcap \mathbf{y}^W. \quad (8)$$

In the case of face localization, the output is the classification of the image block as a face, which is given as a block of white pixels whenever the input image block is identified with any of the stored face patterns. In the mobile self-localization the M memories are build from eroded versions of the scene, and the test images are dilated before being applied to the memory for recognition.

3 Experiments on face localization

As stated in the introduction, the target application is face localization. For this purpose a set face patterns is selected as the representatives of the face class. In the experiments reported here the set of face patterns is the one presented in figure 1. This small set shows several interesting features: faces are of different sizes, background has been manually removed, there is no precise registration of face features (some of the faces are rotated), and there is no intensitity normalization (equalization or any other illumination compensation). Therefore, building this set of patterns corresponds to an almost casual browsing and picking of face images in the database.

Face localization is a two class classification problem, however, we formulate it as a response to a n-class classification problem. As described in the previous section, the M and W HMM´s are built up to classify each image block as one of the face patterns. If it fails, the response is arbitrary and we consider the image block as a non-face block. The HMM´s output are orthogonal binary vectors encoding the face pattern. Both memories are convolved with the image to search for faces. At each pixel the positive classification with one of the M^{σ} memories produces a square of face pixels of a size that is the half of the image block, and centered at this pixel position. The recognition at the different scales is added into an M-recognition binary image. The same process applies to the W^{σ} memories. The intersection of the face pixels recognized with each HMM is the final result, which is superimposed to the original image.

We have performed initial studies over a small database of 20 images with a varying range of scales. The average ROC curve over all the images relating the true and false positives obtained with scale ranges varying from $s = 1$ up to $s = 13$ is shown in figure 2. It can be appreciated that the approach obtains a high recognition rate (over 85%) with very small false recognition rates (less than 5%). As the scale range increases we reach the 100% of face recognition at the pixel level. Face pixels were labelled manually in a process which is independent of the selection of the face patterns. These results are very promising and we are planning the application of this approach to larger face databases, like the well-known CMU database. As a final result, we give in figure 3 some images with recognition results at scale 5.

4 Experiments on self-localization

As stated in the introduction, one of the target applications is Self-Localization for a mobile robot. For this purpose, we have tested the robustness of the HMM to small translations and rotations of the stored views. The results of this experiment are published elsewhere. Here we will report preliminary results on the next step leading to the use of HMM for self-localization. From a mobile robot B-21 (iRobot Corp.), we have taken with the on-board camera a sequence of pictures of a round trip of a laboratory. The following process has been performed in order to select the most representative shots: (1) Each image has been used to build an M memory whose desired output is a single 1. To add robustness the image was eroded with at structural objet of scale predefined. (2) Each M memory constructed was tested against the entire secuence. The recognition corresponds to the output of an 1 by the HMM. For robustness each test image was dilated before application of the HMM. (3) The representative shots were selected as those with greater no intersecting supports in the sequence. The support of an image are the images in the sequence that output an 1 when applied as the input to its M memory.

Figure 4 shows the supports for the images in the sequence. Rows correspond to the image used to build the HMM. Columns correspond to the images as test of the HMM. Obviously the diagonal must be white. The images were eroded with structural elements of scale 6 and dilated with structural elements of scale 4.The figure 5 shows the results of trying to recognize the images in the sequence with the shots identified in the previous step. The recognition shows high spatial coherence. That means that within small spatial displacements the same scene is recognized and a physical position and orientation can be attached to the recognized image. It must taken into account that the images are taken in indoor conditions with artificial light, the proposed approach could be of interest in these conditions.

5 Conclusions and further work

We propose the application of HMM for two tasks: (1) a realization of face localization that can be competitive with other graylevel based procedures, and

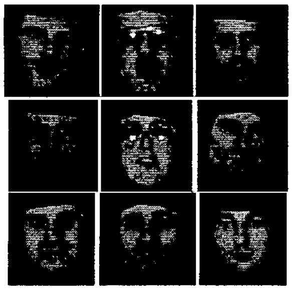

Figure 1: Face patterns used in the experiments

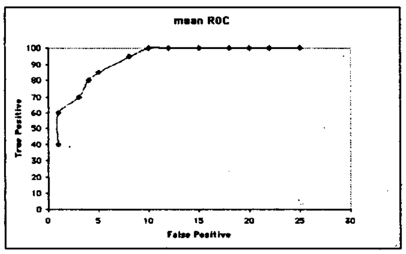

Figure 2: Mean ROC of the face localization across the set of images

Figure 3: Some results of face localization using patterns eroded/dilated to scales up to 5

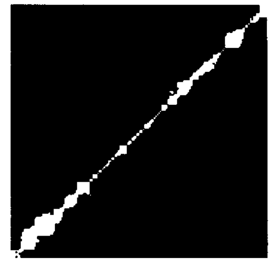

Figure 4: The supports of the images in the sequence. Erosion of scale 6 and dilation of scale 4.

(2) the self-localization of mobile robots based on visual information. The. HMM give a relatively fast response because they only perform integer and max/min operations and its response does not imply the computation of an energy minimum. The main drawback of the HMM in general is their sensitivity to morphological noise: erosions and dilations of the image. We have applied multiscale morphological ideas to overcome this sensitivity, inspired in inspired in the construction of the kernels in [2] and [1]. For the face localization task dual HMM were constructed and applied simultaneously to the images. For the self-localization task, the robust recognition was achieved applying morphological erosion to the images before constructing the M HMM and dilating the images before applying them for recognition. Further work on these appliations over extended data sets are on the way.

Acknowledgements The work has been developped under grants UE-1999-1 and PI-98-21 of the Gobierno Vasco (GV/EJ), Ministerio de Ciencia y Tecnología under grant TIC2000-0739-C04-02. B.Raducanu has the benefit of a predoctoral grant from the University of The Basque Country (UPV/EHU)

References

[1] G. X. Ritter, P. Sussner and J. L. Diaz-de-Leon, "Morphological Associative Memories", IEEE

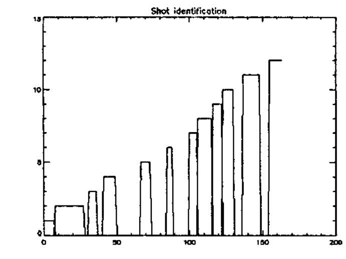

Figure 5: Recognition results along the sequence. Shots are recognized consistently by spatially close views.

Trans. on Neural Networks, 9(2), (1998), pp. 281-292

[2] G. X. Ritter, J. L. Diaz-de-Leon and P. Sussner, "Morphological Bidirectional Associative Memories", Neural Networks, Vol. 12, (1999), pp. 851-867

[3] Rowley H. A., Baluja S., Kanade T., Neural Network-Based Face Detection, IEEE Trans. Patt. Anal. Mach. Int., vol. 20, no. 1, (1998), pp. 23-38

[4] Sung K. K., Poggio T., Example-Based Learning for View-Based Human Face Detection, IEEE Trans. Patt. Anal. Mach. Int., vol. 20, no. 1, (1998), pp. 39-50

[5] Turk M., Pentland A., Eigenfaces for Recognition. Journal of Cognitive Neuroscience, vol. 3, no. 1, (1991), pp. 71-8

[6] Lin S. H., Kung S. Y., Lin L. J., Face Recognition/Detection by Probabilistic Decision-Based Neural Network, IEEE Trans. on Neural Networks, vol. 8, no. 1, (1997), pp. 114-132

[7] Wang J., Tan T., A New Face Detection Method Based on Shape Information, Pattern Recognition Letters, vol. 21, no. 6-7, (2000), pp. 463-471

[8] Yoo T.-W., Oh I.-S., A Fast Algorithm for Tracking Human Faces Based on Chromatic Histograms, Pattern Recognition Letters, vol. 20, no. 10, (1999), pp. 967-978

[9] Wang J.-G., Sung E., Frontal-View Face Detection and Facial Feature Extraction Using Color and Morphological Operators, Pattern Recognition Letters, vol. 20, no. 10, (1999), pp. 1053-1068

[10] Jackway P. T., Deriche M., Scale-Space Properties of the Multiscale Morphological Dilation-Erosion, IEEE Trans. on Patt Anal. and Mach. Int., vol. 18, no. 1, (1996), pp. 38-516

[11] C. Balkenius and L. Kopp, "Robust Self-Localization Using Elastic Templates", in T. Lindberg (ed.), Proceedings of Swedish Symposium on Image Analysis, 1997

[12] D. Fox, "Markov Localization: A Probabilistic Framework for Mobile Robot Localization and Navigation", Ph. D. Thesis, University of Bonn, Germany,December 1998

[13] S. Livatino and C. Madsen, "Optimization of Robot Self-Localization Accuracy by Automatic Visual-Landmark Selection", Proceedings of 11th Scabdinavian Conference on Image Analysis (SCIA), 1999, pp. 501-506

[14] C. F. Olson, "Mobile Robot Self-Localization by Iconic matching of Range Maps", Proceedings of the 8th International Conference on Advanced Robotics (ICAR), 1997, pp. 447-452

[15] J. Reuter, "Mobile Robot Self-Localization Using PDAB", Proceedings of International Conference on Robotics and Automation (ICRA), 2000

[16] A. Saffiotti and L. P. Wesley, "Perception-Based Self-Localization Using Fuzzy Location", in L. Dorst, M. Van Lambalgen and F. Voorbraak (eds.), Lecture Notes in Artificial Intelligence 1093, Springer-Verlag, 1996, pp. 368-385