INCREASING THE ROBUSTNESS OF HETEROASSOCIATIVE MORPHOLOGICAL MEMORIES FOR PRACTICAL APPLICATIONS

M. Graña*, B. Raducanu
University of The Basque Country
Dept. CCIA
Paseo Manuel de Lardizabal, No.1
E-20018 San Sebastian, SPAIN

Abstract Associative Morphological Memories are a recently proposed neural networks architecture based on the shift of the basic algebraic framework. They possess some robustness to specific noise models (erosive and dilative noise). Combining the Associative Morphological Memories with erosion/dilation scale-spaces, we achieved an increased robustness against noise. Here we report ongoing work on their application to the tasks of face localization in grayscale images and visual self-localization of a mobile robot.

Keywords: associative morphological memories, erosion/dilation scale-spaces, face localization, mobile robot self-localization.

1. Introduction

Morphological Associative Memories were recently proposed in [11], [12]. In [11] and [12] the construction of the Heteroassociative and Autoassociative Morphological Memories (HMM and AMM)

is done following the analogy with the construction of the Heteroassociative and Autoassocitive Hopfield Memories, changing the conventional matrix product by the min/max matrix product. The use of the minimum or maximum operator determines the erosive or dilative character of the morphological memory. The capacity of the AMM is not bounded by any condition on the input/output stored patterns. The capacity of the HMM, however, is conditioned by a kind of max/min orthogonality relations between the patterns. The construction of a robust HMM (insensitive to both erosive and dilative noise) is decomposed in the construction of an AMM on the so called input

*Corresponding author. E-mail: ccpgrrom@si.ehu.es

Redistribution rights reserved CSIRO Publishing. ISBN 0 643 06804 X

pattern kernels and an HMM that maps the input pattern kernels into the output patterns. The inconvenience 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 we focus in a way to improve the robustness against noise of raw HMM. The approach taken is that of constructing the HMM with the eroded/dilated versions of the input at several scales in a Scale-Space framework [5].

Face detection can be defined as the problem of deciding the presence of a face in the image. Multiple face detection is usually dealt with by solving many single face detection problems posed over a set of overlapping sub-images extracted from the image by a sliding window. From a statistical pattern recognition perspective, face detection can be considered as a two-class classification problem: the face class or the non-face class. The main difficulty is the appropriate characterization of the non-face class. In our work we do not model the non-face class. Some works that present holistic approaches to face detection are either based on PCA [16] or neural networks [13], [15] and [6].

Geometrical approaches try to fit an ellipse to the face contour [19] or to detect some face elements and verify their relative distances. Finally, approaches based on color processing [17] 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 [2] and [18]. In this spirit, we propose our work with HMM as a complementary verification tool.

Self-Localization is the ability to determine the spatial position and orientation of the robot using the information provided by its sensors [1], [3], [7], [8], [10], [14]. Visual self-localization based on the images provided by on-board cameras is usually based on the detection of some predetermined landmarks [1], [7], [10] specifically designed to be easily recognized in real time. The goal of our work 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 HMM take 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 definition of a morphological scale-space. In section 3 we review the formal definition of HMM together with their properties. In section 4 and 5 we present results on face localization and self-localization. And finally, in section 6, we present our conclusions and future work.

2. Morphological Scale-Spaces

Scale-space theory deals with the formal definition of the concept 'scale' in terms of signals/images, i.e how we represent the data at a given scale and how we relate image features from one scale to another. A basic requisite for a particular collection of increasingly smooth images to be a scale-space is the *causality* property: every feature in a coarse scale (large σ) has to have a cause in a finer scale (small σ). A well-known scale-space is the linear scale-space that results from the convolution with Gaussian kernels of increasing invariance.

Another way to generate a scale-space is using mathematical morphology. Although recent works have proposed several kinds of morphological scale-spaces (erosion/dilation and opening/closing), we will use the erosion/dilation morphological scale-space proposed by Jackway [5]. In the following definitions, we assume that f is the original grayscale image and g is the structuring function, namely $f : D \subset R^n \rightarrow R$ and $g : E \subset R^n \rightarrow R$.. The dilation is defined as: $(f \oplus g)(x) = \sup_{t \in E} \{f(x-t) + g(t)\}$ and the erosion as: $(f \ominus g)(x) = \inf_{t \in E} \{f(x+t) - g(t)\}$. In order to dismiss the *lateral shifting effect*, in [5] they impose for the origin of the structuring function the following conditions: $\sup_{t \in E} \{g(t)\} = 0$ and $g(0) = 0$. A suitable scale-space structuring function is the *sphere* defined by the following equation: $g_\sigma(x) = |x| \left(\left(1 - \|x/\sigma\|^2\right)^{1/2} - 1 \right)$, $\|x\| \leq \sigma$. Now, the multiscale dilation-erosion is defined as:

$$(f \,\tilde{}\, g_\sigma)(x) = \begin{cases} (f \oplus g_\sigma)(x) & \text{if } \sigma > 0 \\ f(x) & \text{if } \sigma = 0 \\ (f \ominus g_\sigma)(x) & \text{if } \sigma < 0 \end{cases} \tag{1}$$

For positive scales ($\sigma > 0$), the operation corresponds to a dilation, and for negative scales ($\sigma < 0$), the operation corresponds to an erosion. As $|\sigma|$ increases, the image tends to have less *structure*. When $|\sigma| \rightarrow 0$, the image converges to the original one.

In an erosion/dilation scale-space, the interesting features for image analysis are the local extrema of the intensity function, in practice the set of *reduced fingerprints*:

$$E_\sigma^*(f, g_\sigma) = \begin{cases} E_{\sigma\,\max}, & if \;\; \sigma > 0 \\ E_\sigma(f, g_\sigma), & if \;\; \sigma = 0 \\ E_{\sigma\,\min}, & if \;\; \sigma < 0 \end{cases} \tag{2}$$

where

$$E_{\sigma\,\min} = \left\{ x \left| (f \,\tilde{}\, g_\sigma)(x) = \min_{\varepsilon \in B_\sigma} \{(f \,\tilde{}\, g_\sigma)(x+\varepsilon)\} \right. \right\} \tag{3}$$

$$E_{\sigma\,\max} = \left\{ x \left| (f \,\tilde{}\, g_\sigma)(x) = \max_{\varepsilon \in B_\sigma} \{(f \,\tilde{}\, g_\sigma)(x+\varepsilon)\} \right. \right\} \tag{4}$$

and B_σ is the sphere of radius σ. The causality property in the erosion/dilation scale-spaces ensures that the Reduced fingerprints are preserved up to a given scale and that no new fingerprints can appear at a non-zero scale. These properties are important in what follows.

3. Heteroassociative Morphological Memories

The work on Morphological Neural Networks stems from the consideration of an algebraic lattice structure $(\mathbf{R}, Y, Z, +)$ as the alternative to the usual $(\mathbf{R}, +, \cdot)$ framework for the definition of Neural Networks computation [11] [12]. The operators Y and Z denote, respectively, the discrete max and min operators (respective. sup and inf in a continuous setting). The approach is termed morphological neural networks because Y and Z are the basic operators for the morphological erosion and dilation.

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

$$W_{XY} = \bigwedge_{\xi=1}^{k} \left[\mathbf{y}^\xi \times \left(-\mathbf{x}^\xi\right)'\right] \qquad M_{XY} = \bigvee_{\xi=1}^{k} \left[\mathbf{y}^\xi \times \left(-\mathbf{x}^\xi\right)'\right] \tag{5}$$

where $\times$ is any of Y or Z. 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 $\forall\xi; W_{XY}\ Y\ \mathbf{x}^\xi \leq \mathbf{y}^\xi \leq M_{XY}\ Z\ \mathbf{x}^\xi$, that can be rewritten $W_{XY}\ Y\ X \leq Y \leq M_{XY}\ Z\ X$.

A matrix A is a Y-perfect (Z-perfect) memory for (X, Y) if $A\ Y\ X = Y$ ($A\ Z\ X = Y$). It can be proven that if A and B are Y-perfect and Z-perfect memories for (X, Y) then $A \leq W_{XY} \leq M_{XY} \leq B$ and $W_{XY}\ Y\ X = Y = M_{XY}\ Z\ X$.

Conditions of perfect recall of the stored patterns are given by a theorem [11],[12] that states that W_{XY} is Y -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 Z-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:

$$\forall\gamma\forall i\exists j; x_j^\gamma = \bigvee_{\xi=1}^{k} \left(x_j^\xi - y_i^\xi\right) + y_i^\gamma, \quad \text{and} \tag{6}$$

$$\forall\gamma\forall i\exists j; x_j^\gamma = \bigwedge_{\xi=1}^{k} \left(x_j^\xi - y_i^\xi\right) + y_i^\gamma. \tag{7}$$

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 dilative noise. Morphological memories are very sensitive to these kinds of noise. The conditions of perfect recall for W_{XY}, i.e. the retrieval of $\mathbf{y}^\gamma$ given a noisy copy $\widetilde{\mathbf{x}}^\gamma$, that is, the conditions under which $W_{XY}\ Y\ \widetilde{\mathbf{x}}^\gamma = \mathbf{y}^\gamma$, are as follows:

$$\begin{array}{l} \forall j; \widetilde{x}_j^\gamma \leq x_j^\gamma\ Y \bigwedge_i \left(\bigvee_{\xi\neq\gamma}\left(y_i^\gamma - y_i^\xi + x_j^\xi\right)\right) \\ \forall i\exists j_i; \widetilde{x}_{j_i}^\gamma = x_{j_i}^\gamma\ Y \left(\bigvee_{\xi\neq\gamma}\left(y_i^\gamma - y_i^\xi + x_{j_i}^\xi\right)\right) \end{array} \tag{8}$$

This condition (8) and the dual one for the M_{XY} memory are the basis for our approach. They 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.

In essence, disregarding some small effects of the output components y_i^γ, the condition for perfect recall of (x^γ, y^γ) is that at least one of the x_i^γ is the maximum/minimum in position i across all patterns ξ. We may call the set of such i-s for a given γ the pseudo kernel of this pattern.

Let us translate this for the special case output vectors being orthogonal binary codes, i.e. $y^\xi \cdot y^\gamma = \delta_{\xi\gamma}$ and $y^\xi, y^\gamma \in \{0,1\}^N$. We will assume a coding of the form $\left(y_\xi^\xi = 1, y_\gamma^\xi = 0, \gamma \neq \xi\right)$. Then, condition 6 for perfect recall of the W_{XY} matrix becomes the following condition:

$$\forall i \, \exists j; \;\; x_j^\gamma > \left(\bigvee_{\xi=1}^{k} x_j^\xi \right) \tag{9}$$

Therefore, in order that image and output code $\left(x^\xi, y^\xi\right)$ can be properly stored in a W_{XY} memory, and that the output y^ξ can be recalled by $W_{XY} \; Y \, x^\xi$ there must exist some pixel positions whose values are greater in this image than the maximum value of these pixels in the remaining images. The dual assertion is true for M_{XY} memory. In order for the pair $\left(x^\xi, y^\xi\right)$ to be stored in M_{XY} and the output y^ξ recalled by $M_{XY} \; Z \, x^\xi$, then there must exist some pixels whose values are lower in this image than the minimum value found in the remaining images.

We will call pseudo-kernel the set of indices that ensure the recall of any input/output pair with an HMM. $K_W\left(x^\gamma\right) = \left\{ j \left| x_j^\gamma > x_j^\xi, \; \gamma \neq \xi \right. \right\}$ is the set of distinctive pixels that ensure that image x^γ and its class code can be stored and retrieved in a noise-free setting with W_{XY}. The dual pseudo-kernel $K_M\left(x^\gamma\right) = \left\{ j \left| x_j^\gamma < x_j^\xi, \; \gamma \neq \xi \right. \right\}$ corresponds to the set of indices that ensure storage and recall with the M_{XY} heteroassociative memory.

For the special case of orthogonal binary output, the conditions on perfect recall of HMM can be rewritten as: $K_W\left(x^\gamma\right) \neq \emptyset$, $\forall \gamma$ for the W_{XY} memory and, respectively $K_M\left(x^\gamma\right) \neq \emptyset, \forall \gamma$ for the M_{XY} memory.

Our approach is based on the following reasoning. The pseudo kernels will correspond with high probability to local extrema of the image and will be preserved, up to a scale, by the multiscale erosion or dilation (the causality property). That is, if $K_M\left(x^\gamma\right) \cap E_{\min}\left(x^\gamma\right) \neq \emptyset$ and the M memory constructed from the eroded patterns up to scale σ will be robust against erosive noise up to scale σ, besides its natural robustness to dilative noise. The dual assertion holds for the W memory. Therefore, the M/W HMM constructed from eroded/dilated patterns will have the same behavior on the noiseless patterns as the HMM constructed from the original ones and will be robust to general noise up to a certain scale.

4. Experiments on Face Localization

Given a set of input face patterns X and a set of output class encoding Y, where the encoding are given by a set of orthogonal binary vectors. We built a set of HMM $\{M^{\sigma}_{XY}, W^{\sigma}_{XY}; \sigma = 1, 2, ..., \mathbf{s}\}$ where each M^{σ}_{XY} is constructed from the output and the input patterns eroded with an spherical structural object of scale σ, and each W^{σ}_{XY} 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 $\mathbf{y}^M = \bigcup_{\sigma=1}^{\mathbf{s}} (M^{\sigma}_{XY} \; Z \, \mathbf{x})$ and $\mathbf{y}^W = \bigcup_{\sigma=1}^{\mathbf{s}} (W^{\sigma}_{XY} \; Y \, \mathbf{x})$. The final output is the intersection of these multiscale responses:

$$\mathbf{y} = \mathbf{y}^M \bigcap \mathbf{y}^W. \tag{10}$$

The morphological memories are applied taking as the input a sliding window over the image. The output is the classification of the image window. A block of white pixels is introduced whenever the input image window is identified with any of the stored face patterns.

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 intensity 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.

We have performed initial studies over a small database of 20 images with a varying range of scales. Face pixels were labelled manually in a process which is independent of the selection of the face patterns. The average ROC curve over all the images relating the true and false positives obtained with scale 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 increases we reach the 100% of face recognition at the pixel level. These results are very promising and we are working on the application of this approach to larger face databases, like the well-known CMU database [13]. As a final result, we give in figure 3 some images with detection results from patterns eroded/dilated at scale 5.

5. 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 [4] and [9] . 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. For increased robustness the image was eroded with at structural object of scale predefined. (2) Each M memory constructed was tested against the

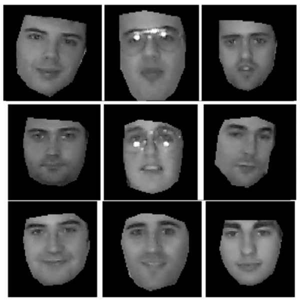

Figure 1. Face patterns used in the experiments

entire sequence. The recognition corresponds to the output of an 1 by the HMM. For increased robustness each test image was dilated before application of the HMM. (3) The representative shots were selected as those with greater non 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 4a shows, as a binary image, 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 and dilated with structural elements of scale 6. In figure 4b, we present for comparison the image supports computed from the correlation based *distance map* between images (binarized with an optimal threshold). This support image is very ambiguous and unsuitable for analysis. Besides being faster, HMMs give a better means to characterize significative shots. The figure 5a shows the results of trying to recognize the images in the sequence with the shots identified in the previous step, i.e. based on figure 4a. The recognition shows high spatial coherence. That means that within small spatial displacements the same view is recognized and the selected views can be used as landmarks with an attached physical position and orientation. To test that, the selected views are used for recognition over a video sequence that corresponds to a second walk of the robot around the lab, following a path close to the first pass. The results in figure 6b show that the spatial coherence of the recognition is preserved.

6. 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 (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 noise (in a morphological sense): erosions and dilations of the image. We have applied multiscale morphological ideas to overcome this sensitivity,

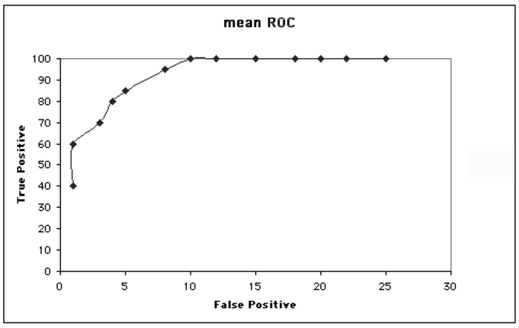

Figure 2. Mean ROC of the face localization across the set of images

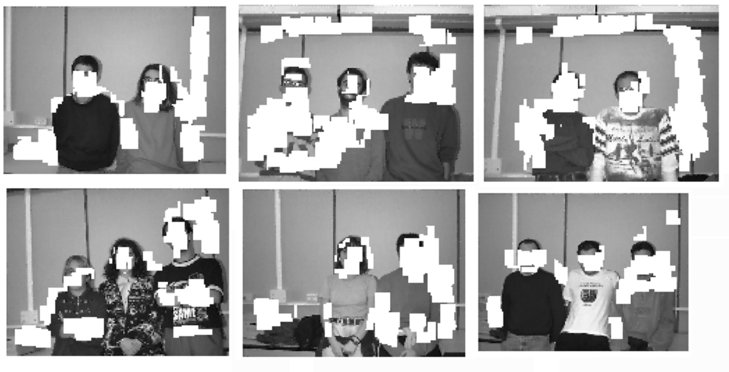

Figure 3. Some results of face localization using patterns eroded/dilated to scales up to 5

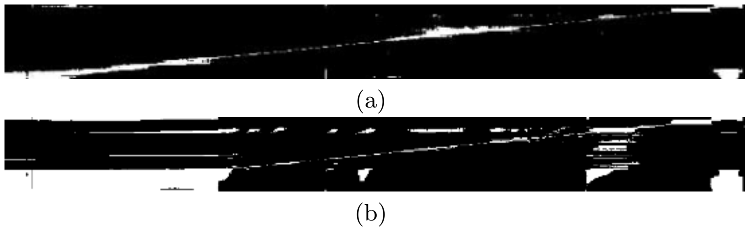

Figure 4. The supports of the images in the sequence based on: (a) the M memory with erosion/dilation of scale 6 and (b) correlation after and optimal thresholding

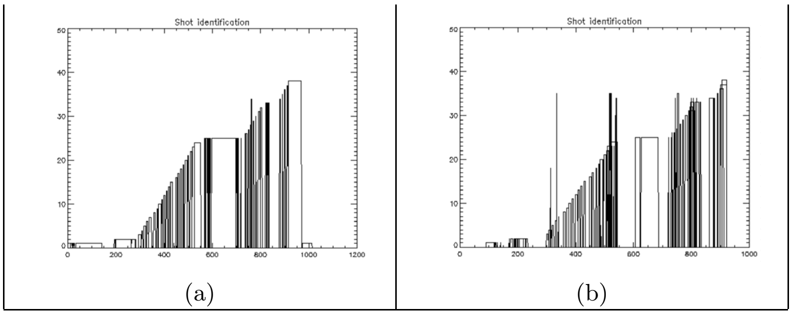

Figure 5. Recognition results along two sequences. Shots are recognized consistently by spatial close views. (a) the same sequence from where the shots have been extracted (b) a similar sequence

inspired in the construction of the kernels in [11] and [12]. 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 memory and dilating the images before applying them for recognition. Further work on these applications over extended data sets are on the way.

Acknowledgments

The work has been developed under grants UE-1999-1 and PI-98-21 of the Gobierno Vasco (GJ/EV), Ministerio de Ciencia y Tecnologia under grant TIC2000-0739-C04-02. B. Raducanu benefitted of a predoctoral grant from the University of The Basque Country (UPV/EHU).

References

[1] C. Balkenius and L. Kopp. Robust self-localization using elastic templates. In T. Lindberg, editor, *Proceedings of Swedish Symposium on Image Analysis*, 1997

[2] R. Féraud, O. J. Bernier, J.-E. Viallet and M. Collobert. A fast and accurate face detector based on neural networks. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(1):42-53, 2001

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

[4] M. Graña and B. Raducanu. On the application of morphological heteroassociative neural networks. In: *Proceedings of International Conference on Image Processing* (ICIP), I. Pitas (ed.), pp.501-504, Thessaloniki, Greece, October 2001

[5] P. T. Jackway and M. Deriche. Scale-ppace properties of the multiscale morphological dilation-erosion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(1):38-51, 1996

[6] S. H. Lin, S. Y. Kung and L. J. Lin. Face recognition/detection by probabilistic decision-based neural network. *IEEE Trans. on Neural Networks*, 8(1):114-132, 1997

[7] S. Livatino and C. Madsen. Optimization of robot self-localization accuracy by automatic visual-landmark selection. *Proceedings of the 11th Scandinavian Conference on Image Analysis (SCIA)*, pages 501-506, 1999

[8] C. F. Olson. Mobile robot self-localization by iconic matching of range maps. *Proceedings of the 8th International Conference on Advanced Robotics*, pages 447-452, 1997

[9] B. Raducanu, M. Graña and P. Sussner. Morphological neural networks for vision based self-localization. *Proceedings of ICRA2001, International Conference on Robotics and Automation*, pp. 2059-2064, Seoul, Korea, May 2001

[10] J. Reuter. Mobile robot self-localization using PDAB. *Proceedings of International Conference on Robotics and Automation* (ICRA), pages not available, 2000

[11] G. X. Ritter, J. L. Diaz-de-Leon and P. Sussner. Morphological bidirectional associative memories. *Neural Networks*, Volume 12, pages 851-867, 1999

[12] G. X. Ritter, P. Sussner and J. L. Diaz-de-Leon. Morphological associative memories. *IEEE Trans. on Neural Networks*, 9(2):281-292, 1998

[13] H. A. Rowley, S. Baluja and T. Kanade. Neural network-based face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. 20(1):23-38, 1998

[14] A. Saffiotti and L. P. Wesley. Perception-based self-localization using fuzzy location. In L. Dorst, M. van Lambalgen and F. Voorbraak, editors, *Lecture Notes in Artificial Intelligence 1093*, pages 368-385, Springer-Verlag, 1996

[15] K. K. Sung and T. Poggio. Example-based learning for view-based human face detection. IEEE Trans. on Pattern Analysis and Machine Intelligence, 20(1):39-50, 1998

[16] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71-86, 1991

[17] T.-W. Yoo, I.-S. Oh. A fast algorithm for tracking human faces based on chromatic histograms. *Pattern Recognition Letters*, 20(10):967-978. 1999

[18] J.-G. Wang and E. Sung. Frontal-view face detection and facial feature extraction using color and morphological operators. *Pattern Recognition Letters*, 20(10):1053-1068, 1999

[19] J. Wang and T. Tan. A new face detection method based on shape information. *Pattern Recognition Letters*, 21(6-7):463-471, 2000