Invariant Pattern Recognition of 2D Images Using Neural Networks and Frequency-Domain Representation*

Fernando César C. De Castro
decastro@ee.pucrs.br

José Nelson Amaral
amaral@ee.pucrs.br

Paulo Roberto G. Franco
pfranco@ee.pucrs.br

Electrical Engineering Department
Pontifícia Universidade Católica do Rio Grande do Sul
90619-900 - Porto Alegre - RS - Brazil

Abstract

Frequency domain representation of two dimensional gray-level images is used to develop a pattern recognition method that is invariant to rotation, translation and scaling. Frequency domain representation is a natural feature detector that allows the use of only few directions of highest energy as training data for a set of Artificial Neural Networks (ANNs). We developed a new algorithm that uses the spectral information stored in these ANNs to compare a given image with a known pattern, determining the relative translation between them and yielding a measure of their similarity. The representation and method we adopted has the advantage of leaving only the rotation of the object as a free parameter to be determined by the algorithm. We minimize the spectral resolution noise using Spectral Directional Filtering. Our experimental results indicate that the proposed method has excelent discriminating power.

1 Introduction

In an earlier work [3] the authors introduced a frequency-domain method for pattern recognition in two dimensional gray-level images. That method uses the image magnitude spectrum as a feature space, yielding translation invariance. Scaling invariance is attained by storing scale information in an ANN set and rotational invariance is achieved using the ANN set in a template matching procedure. Translation values are obtained from the Cross Power Spectral Density between the images. Our preliminary results [3] have shown good discrimination of patterns and an insensitivity to the effects of finite resolution. This article presents a detailed derivation for the procedure.

Our algorithm addresses the problem of evaluating the similarity between objects in two-dimensional images [9]. Given a gray-level reference image and a gray-level input image, the algorithm aims to identify if an object in the input image corresponds to the pattern object in the reference image, regardless of scaling factor, rotation angle, and translation.

We can divide image recognition techniques into spatial-domain algorithms and frequency-domain algorithms. A spatial-domain algorithm processes the image pixel values and their respective x-y coordinates. As a consequence, invariance to translation can be achieved only through costly computations. Algorithms that use spatial-domain representation include Template Matching [8], Moment Invariants [1], and Artificial Neural Networks (ANNs) [5].

A frequency-domain algorithm computes the image spectrum by applying a two dimensional Fast Fourier Transform (FFT) [7]. The magnitude spectrum of an image contains information about the object shape and the phase spectrum contains information about the object translation; therefore, the shape information in the magnitude spectrum is naturally translation-invariant. The Fourier-Mellin transform is a powerful tool for image recognition techniques that use this spectral property [12, 9]. This method represents rotation and scaling as single translations on the parameter space, and allows the use of the Phase Correlation Technique [10] to determine translation, scaling and rotation. Thus, there are four free parameters. To represent scaling and rotation as translations, the Fourier-Mellin method uses logarithmic mapping, which introduces

* This work was in part supported by a grant from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and by Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) - Brazil

distortion in the radial spectral polar coordinate. This distortion results in a non-uniform precision for the scale estimation and an accuracy loss that is specially severe for images of 128×128 pixels [9].

In the present article, we use spectral differential relations to determine directly the translation parameters without the need for the Phase Correlation Technique. Because scaling information is stored in ANNs, there is no need of logarithmic mapping for scale estimation. Therefore, our method does not suffer from the logarithmic distortion problem pointed by Chen [9]. Also, due to the intrinsic ANNs generalization capability, it is not necessary to use any interpolating scheme, even for images of 128×128 pixels. Our approach to the pattern recognition problem is naturally invariant for translation and scaling. Thus, the only free parameter is rotation.

2 General Description

In this work we use a *Feedforward Multilayer Perceptron* (FMP) trained with the backpropagation algorithm [11] to store the magnitude of a Fast Fourier Transform (FFT)[6, 7] applied to the reference image. The FMP architecture is 65-7-1, i.e., it has 65 input nodes, 7 hidden neurons and 1 output neuron. The use of one hidden layer of 7 neurons was determined empirically as a compromise to minimize the mean square error of the trained ANN and to maximize the learning speed [11]. The 65 input nodes are necessary because we are working with an image of 128×128 pixels. We need only one output neuron because each ANN is dedicated to estimate a single scaling factor value.

Let $r(x, y)$ be the spatial representation of a reference image with a given pattern object. We want to compare the pattern object in $r(x, y)$ with a target object in an input image $s(x, y)$. In this work we assume that both $r(x, y)$ and $s(x, y)$ have a black (zero) background and a size of $N \times N$ pixels, with N being an integer power of two. Let $R(u, v)$ be the spectrum of $r(x, y)$ defined by $R(u, v) = FFT\{r(x, y)\}$ [7]. Let $R(\omega, \Psi)$ be the spectrum $R(u, v)$ in a polar coordinate system with center at $\omega = 0$, such that $u = \omega \cos \Psi$ and $v = \omega \sin \Psi$. The reference image spectral energy in direction Ψ is defined by equation (1).

$$E(\Psi) = \frac{1}{N^2} \sum_{\omega=0}^{\frac{N}{2}+1} |R(\omega, \Psi)|^2 \tag{1}$$

A highest spectral energy direction (HSED) of an image is a value of Ψ in which $E(\Psi)$ is a local maximum. The j^{th} HSED of an image is the spectral direction Ψ_j such that $E(\Psi_j)$ is the j^{th} local maximum of $E(\Psi)$, ordered such that $E(\Psi_j) > E(\Psi_{j+1})$. For instance, consider the spectral magnitude of a generic image as shown inside the square in Figure 1. The bright regions represent high values of spectral magnitude while the dark regions represent low values. This image has $m = 6$ HSEDs respectively at 168^0, 140^0, 117^0, 90^0, 37^0 and 0^0, indicated by the white lines originating at the center.

To achieve scaling invariance, the spectral magnitude variations due to spatial scaling in the reference image are used to train a set of m ANNs. We select the m HSEDs directions that concentrate most energy and train m ANNs with the spectral magnitude samples along these HSEDs.

We use P reference images to train the set of m ANNs. Typically, for a 128×128 pixels image P is between 10 and 20 [3]. The k^{th} generated image $r_k(x, y)$ is the original reference image scaled by α_k, $k = 0, \ldots, P-1$. Let $R_k(\omega, \Psi)$ be the polar coordinate spectrum of $r_k(x, y)$ with center at $\omega = 0$. The input vector of an ANN is the set of samples at the ANN input nodes [11]. The input vector for the j^{th} ANN are the values of $|R_k(\omega, \Psi_j)|$, where Ψ_j is the j^{th} HSED. For each training input vector of the j^{th} training set, we assign to the j^{th} ANN desired output the scaling factor α_k. In this way, each ANN learns the pattern object scaling variations. The ANN generalization capability enables the algorithm to recognize scaling factors that are different from those used to train the ANNs.

To solve the rotation invariance problem, we use the fact that an image rotation in the spatial domain implies the same rotation in the spectral domain [7]. Figure 1 depicts a functional block of the proposed algorithm which we call Spectral Angular Discriminator (SAD). Each triangle in the figure represents a trained ANN. The SAD is used to determine the rotation angle θ_0 between the objects in $r(x, y)$ and in $s(x, y)$. The SAD input is a set of m ANN input vectors beginning at $\omega = 0$ and having the same angular spacing as the HSEDs. In Figure 1 the SAD input is rotated by $\Psi = 0^0$, therefore the m input vectors are aligned with the respective HSEDs. If the SAD input were rotated by $\Psi = \beta$ each ANN input vector would be spaced of an angle β from their respective HSED, where $0 \leq \beta < 180^0$. The SAD output $H(\Psi)$ is a linear combination of the m ANN outputs to a given rotation angle Ψ of the SAD input on the image magnitude spectrum. This linear combination is averaged as shown in Figure 1.

Let $H_s(\Psi)$ be the SAD output to $s(x, y)$. In the results of a previous work, De Castro *et al.* [3] indicated that the global maximum of $H_s(\Psi)$ yields α_a, the first approximation to the scaling factor of the object

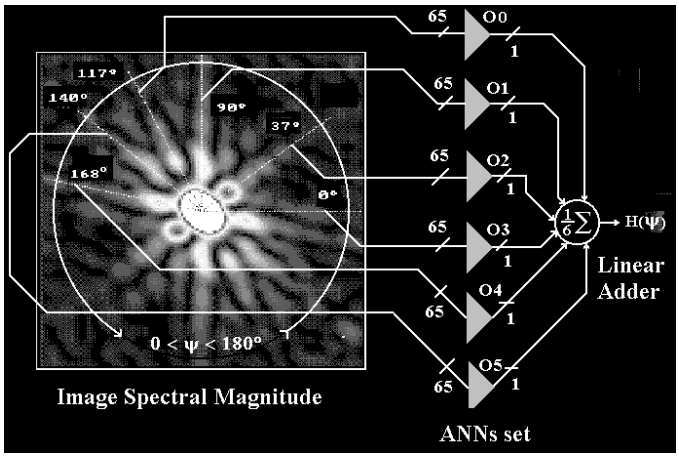

Figure 1: The SAD diagram for $m = 6$ and $N = 128$.

in $s(x, y)$. Defining $H_{r,\alpha_a}(\Psi)$ as the SAD output to the reference image scaled by α_a and $R_{rs}(\theta)$ as the circular cross-correlation function between $H_{r,\alpha_a}(\Psi)$ and $H_s(\Psi)$, that results also indicated that the global maximum of $R_{rs}(\theta)$ is $\theta = \theta_0$ and the square root of this maximum value defines the scaling factor α_0 of the target object [3].

3 The Relative Translation and the Decision Criterion

Let $r_0(x, y)$ be the reference image scaled by α_0 and rotated by θ_0. $S(u, v)$ and $R_0(u, v)$ are the spectra of $s(x, y)$ and of $r_0(x, y)$. If the shape of the target object matches the shape of the reference object, the only difference between $s(x, y)$ and $r_0(x, y)$ is a translation (x_0, y_0). In this situation their spectra are related by equation (2) [7].

$$R_0(u, v) = S(u, v)e^{-i\frac{2\pi}{N}(ux_0+vy_0)} \tag{2}$$

Equation (2) can be expressed as

$$\left|\frac{R_0(u, v)}{S(u, v)}\right| e^{i(\angle R_0(u,v)-\angle S(u,v))} = e^{-i\frac{2\pi}{N}(ux_0+vy_0)} \tag{3}$$

Using the phase information in equation (3) we can write

$$\angle D = \arctan\left(\frac{\Im\{D\}}{\Re\{D\}}\right) = -i\frac{2\pi}{N}(ux_0 + vy_0) \tag{4}$$

where

$$D = D(u, v) = R_0(u, v)S^*(u, v) \tag{5}$$

is the cross power spectral density between $r_0(x, y)$ and $s(x, y)$ [2]. $S^*(u, v)$ is the complex conjugate of $S(u, v)$. $\Re\{D\}$ is the real part of D and $\Im\{D\}$ is the imaginary part of D.

As long as (2) holds, the differentiation of equation (4) with relation to u and v yields

$$x_0(u, v) = \frac{N}{2\pi}\frac{\left[\Im\{D\}\frac{\partial}{\partial u}\Re\{D\} - \Re\{D\}\frac{\partial}{\partial u}\Im\{D\}\right]}{|D|^2} \tag{6}$$

$$y_0(u, v) = \frac{N}{2\pi}\frac{\left[\Im\{D\}\frac{\partial}{\partial v}\Re\{D\} - \Re\{D\}\frac{\partial}{\partial v}\Im\{D\}\right]}{|D|^2} \tag{7}$$

If the relationship between $r_0(x, y)$ and $s(x, y)$ is a spatial translation, equation (2) holds, resulting in $x_0(u, v)$ and $y_0(u, v)$ constant for $u = 0, \ldots, N - 1$ and $v = 0, \ldots, N - 1$. If equation (2) does not hold, then we can express the relationship between $R_0(u, v)$ and $S(u, v)$ by equation (8). In equation (8) we made

explicit the dependence of x_0 and y_0 on u and v. This suggests that the variances of $x_0(u,v)$ and $y_0(u,v)$ indicate the degree of shape mismatch between the objects in the reference and in the input images.

$$R_0(u,v) = S(u,v)e^{-i\frac{2\pi}{N}(ux_0(u,v)+vy_0(u,v))} \tag{8}$$

In order to minimize the spectral resolution noise resulting from the image finite resolution in the frequency domain, we apply Spectral Directional Filtering to $x_0(u,v)$ and $y_0(u,v)$ [7, 3]. Let $x_0(\omega,\Psi)$, $y_0(\omega,\Psi)$ and $R_0(\omega,\Psi)$ represent $x_0(u,v)$, $y_0(u,v)$ and $R_0(u,v)$ in a polar coordinate system with center at $\omega = 0$, such that $u = \omega\cos\Psi$ and $v = \omega\sin\Psi$. Let $\Psi_j + \theta_0$ be the j^{th} HSED of $R_0(\omega,\Psi)$, $j = 0,\ldots,m-1$. Since $x_0(\omega,\Psi)$ and $y_0(\omega,\Psi)$ are translations along orthogonal directions, the inverse of the vector sum of their variances computed along the HSEDs can be used as a measure of the shape similarity between the objects in $s(x,y)$ and in $r(x,y)$. Thus, we define a similarity ratio $\widehat{\lambda}$ according to equation (9),

$$\widehat{\lambda} = 10\log\left(\frac{1}{1+\sqrt{(\widehat{\sigma}_x^2)^2+(\widehat{\sigma}_y^2)^2}}\right) \quad [\text{dB}] \tag{9}$$

$$\widehat{\sigma}_x^2 = \frac{1}{\Omega}\sum_{\omega=\omega_0}^{\omega_1}\sum_{j=0}^{m-1}(x_0(\omega,\Psi_j+\theta_0)-\widehat{x}_0)^2 \tag{10}$$

$$\widehat{\sigma}_y^2 = \frac{1}{\Omega}\sum_{\omega=\omega_0}^{\omega_1}\sum_{j=0}^{m-1}(y_0(\omega,\Psi_j+\theta_0)-\widehat{y}_0)^2 \tag{11}$$

$$\widehat{x}_0 = \frac{1}{\Omega}\sum_{\omega=\omega_0}^{\omega_1}\sum_{j=0}^{m-1}x_0(\omega,\Psi_j+\theta_0) \tag{12}$$

$$\widehat{y}_0 = \frac{1}{\Omega}\sum_{\omega=\omega_0}^{\omega_1}\sum_{j=0}^{m-1}y_0(\omega,\Psi_j+\theta_0) \tag{13}$$

where $\Omega = m(\omega_1-\omega_0+1)$ and ω_0 and ω_1 define the radii of the filter annular bandpass region in the spectral domain which define the filter cutoff frequencies.

The algorithm compares the value of $\widehat{\lambda}$ obtained from equation (9) with an empirical threshold value to decide whether $s(x,y)$ and $r(x,y)$ contain the same pattern. Notice that if the images $s(x,y)$ and $r(x,y)$ are not correlated[1], the use of equations (6) and (7) to estimate the translation between them will produce a large value for $\widehat{\sigma}_x^2$ and $\widehat{\sigma}_y^2$ and a low value for $\widehat{\lambda}$. Hence the algorithm will correctly conclude that the object in $s(x,y)$ is not the same as the reference object in $r(x,y)$. Notice also that if the shape of the objects are a perfect match then $\widehat{\lambda} = 0$ dB.

	a) $\alpha_0 = 0.40$ and $\theta_0 = 58.3^0$		b) $\alpha_0 = 0.32$ and $\theta_0 = 61.1^0$	
	$\widehat{\lambda} = -4.1\times10^{-3}$ dB	$\widehat{\lambda}_r = -0.61$ dB	$\widehat{\lambda} = -0.56$ dB	$\widehat{\lambda} = -0.19$ dB
$\widehat{x}_0$	40.5 pixels	48.7 pixels	36.2 pixels	41.8 pixels
$\widehat{y}_0$	29.7 pixels	32.0 pixels	26.4 pixels	26.9 pixels

Table 1: Algorithm estimates when the input and reference images are respectively a) Fig. 2b and Fig. 2a and b) Fig. 2c and Fig. 2a.

The filter computes the translation estimates $\widehat{x}_0$ and $\widehat{y}_0$ using samples along the HSEDs that are inside the bandpass ring defined by $[\omega_0,\omega_1]$. Choosing ω_0 and ω_1 such that the highest values of $|R_O(\omega,\Psi)|$ along its HSEDs are kept inside the filter bandpass, we minimize the uncertainties in $\widehat{x}_0$ and $\widehat{y}_0$. Because the length of a circumferential arc is proportional to its radius for a given arc angle, the finite resolution noise of the SAD is higher for the samples farther from the center of the spectrum . This led us to adopt $\omega_0 = 0$ and $\omega_1 = \omega_c$, where ω_c is the radius of the circle in the spectral domain for which $|R_0(\omega,\Psi)| < \frac{1}{2}\max\{|R_0(\omega,\Psi)|\}$. This bandpass threshold was determined experimentally aiming to increase the precision of the estimators $\widehat{x}_0$ and $\widehat{y}_0$.

[1]Remember that at this point, in spite of the estimates obtained for θ_0 and α_0, we have not decided whether the image under analysis $s(x,y)$ and the reference image $r(x,y)$ have any similarity.

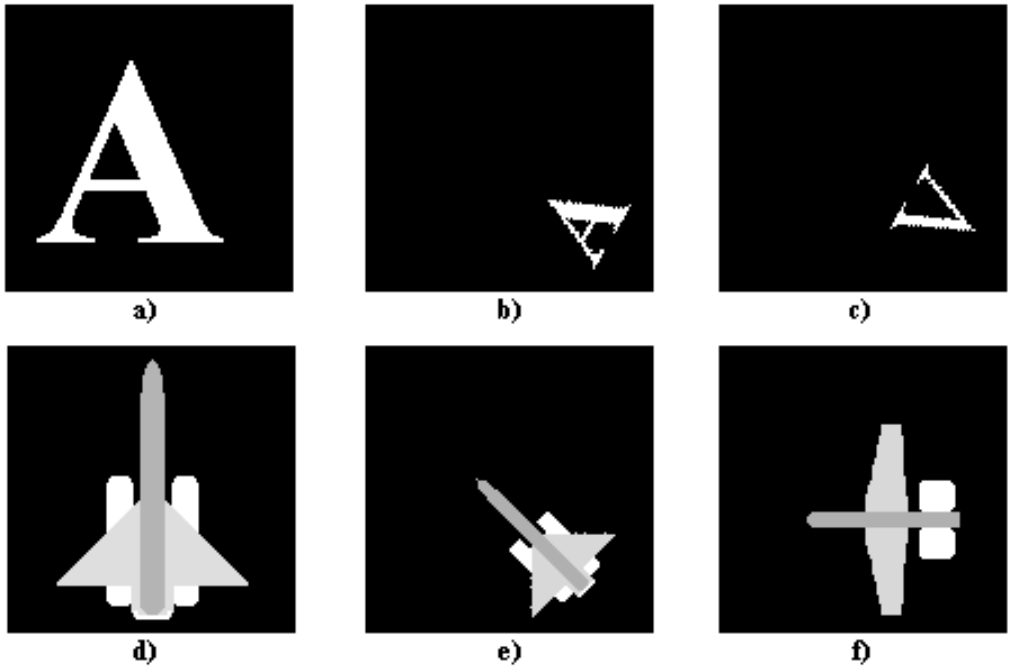

Figure 2: All images are 128×128 pixels of size. The alphabetical characters are *Times New Roman* . Fig. 2b is the image of Fig. 2a scaled by 0.4, rotated by 60^0 and translated by (40,30) pixels. Fig. 2e is the image of Fig. 2d scaled by 0.6, rotated by 45^0 and translated by (10,20) pixels.

	a) $\alpha_0 = 0.59$ and $\theta_0 = 44.8^0$		b) $\alpha_0 = 0.81$ and $\theta_0 = 0.00^0$	
	$\widehat{\lambda} = -2.3 \times 10^{-2}$ dB	$\widehat{\lambda}_r = -4.31$ dB	$\widehat{\lambda} = -31.9$ dB	$\widehat{\lambda} = -19.5$ dB
$\widehat{x}_0$	9.8 pixels	21.3 pixels	6.5 pixels	6.5 pixels
$\widehat{y}_0$	19.9 pixels	31.6 pixels	0.9 pixels	25.8 pixels

Table 2: Algorithm estimates when the input and reference images are respectively a) Fig. 2e and Fig. 2d and b) Fig. 2f and Fig. 2d.

Due to the spectrum symmetry [7], the estimated θ_0 belongs to the interval $[0, 180^0]$, whereas the actual rotation angle may be either θ_0 or $\theta_0 + 180^0$. To solve this ambiguity we apply the spectral relation (14) to equation (5) [4]:

$$RR_0(u,v) = R_0^*(u,v)e^{-i\frac{4\pi}{N}(ux_c+vy_c)} \tag{14}$$

where $RR_0(u,v)$ is the spectrum of $rr_0(x,y)$, the image $r_0(x,y)$ rotated by 180^0, and (x_c, y_c) is the geometric center of the object in $r_0(x,y)$. We apply both $R_0(u,v)$ and $RR_0(u,v)$ to equation (5) to obtain D and D_r. Substituting these two values in the chain of equations (6), (7),(12), (13), (10), (11) we obtain from (9) two likeness ratios $\widehat{\lambda}$ and $\widehat{\lambda}_r$. If $\widehat{\lambda} < \widehat{\lambda}_r$ the relative rotation angle is $\theta_0 + 180^0$ and not θ_0.

4 Experimental Results

This section describes some experimental results of the algorithm applied to alphabetical character recognition and gray-scale object identification.

Figures 2a and 2b present the reference and the input images. Table 1a shows the algorithm estimate for the scaling factor α_0, the rotation angle θ_0, and the translation (x_0, y_0). Since $\widehat{\lambda} > \widehat{\lambda}_r$, the actual estimated rotation angle is $\theta_0 = 58.3^0$.

To demonstrate the discriminating power of this algorithm, consider the input image in Figure 2c and the reference image in Figure 2a. Table 1b shows the algorithm performance. Although the character "V" is quite similar to the character"A" both $\widehat{\lambda}$ and $\widehat{\lambda}_r$ resulted in much higher values than the $\widehat{\lambda}$ value in Table 1a. Note that $\widehat{\lambda} < \widehat{\lambda}_r$, meaning that the algorithm interpreted the character "V" as a rough approximation of the pattern "A" rotated by $180^0 + 61.1^0 = 241.1^0$.

Tables 2a and 2b show the algorithm performance on aircraft identification. Notice the low value of λ when the objects are different, as shown in Table 2b.

5 Conclusion

A difficult problem in two-dimensional pattern recognition is to attain invariance to rotation, scaling and translation. We use spectral features representation and the generalization capability of ANNs to develop a method with such properties. We also use spectral properties to develop a direct procedure to determine the translation and the degree of similarity between two images. Future research efforts will add noise to the images to investigate the robustness of the method.

References

[1] Y.S. Abu-Mostafa and D. Psaltis. Recognition aspects of moment invariants. *IEEE Trans. Pattern Anal. Mach. Intel.*, 16(12):1156–1168, December 1984.

[2] A.V.Oppenheim. *Applications of Digital Signal Processing.* Prentice Hall, 1978.

[3] F. C. C. De Castro, J. N. Amaral, and P. R. G. Franco. Artificial neural networks for frequency-domain pattern recognition. In *International Conference on Information Systems Analysis and Synthesis*, pages 108–115, Orlando, USA, July 1996.

[4] F.C.C. De Castro. Reconhecimento e localização de padrões em imagens utilizando redes neurais artificiais como estimador de correlação espectral. Master's thesis, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre - RS - BRAZIL, December 1995.

[5] A.C.C. Coolen and F.W. Kuijk. A learning mechanism for invariant pattern recognition in neural networks. *Neural Networks*, 2:495–506, 1989.

[6] P.M Cooley and J.W. Tukey. An algorithm for the machine computation of complex Fourier series. *Mathematics of Computation*, 19:297–302, April 1965.

[7] R. Gonzalez and P. Wintz. *Digital Image Processing.* Addison Wesley, 1987.

[8] A. Goshtasby. Template matching in rotated images. *IEEE Trans. Pattern Anal. Mach. Intel.*, 7(3):338–344, 1985.

[9] M. Defrise Q. Chen and F. Deconinck. Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition. *IEEE Trans. Pattern Anal. Mach. Intel.*, 16(12):1156–1168, December 1994.

[10] B.S. Reddy and B.N. Chatterji. An FFT-based technique for translation, rotation and scale-invariant image registration. *IEEE Trans. Image Proc.*, 5(8):1266–1271, August 1996.

[11] S.Haykin. *Neural Networks.* Macmillan College, 1992.

[12] Y.Sheng and H.H. Arsenault. Experiments on pattern recognition using invariant Fourier-Mellin descriptors. *J. of Opt. Soc. Am. A.*, 3(6):771–776, 1986.