

AN INTEGRATED FRAMEWORK FOR IMAGE CLASSIFICATION

*Xuling Luo and Gagan Mirchandani**

Department of Electrical and Computer Engineering
The University of Vermont
Burlington, Vermont 05405

ABSTRACT

This paper presents a novel method for classifying an image into one of predefined classes in a data bank by applying mutual information to a representation of the Fourier amplitude domain. Template and test images are made translation and rotation invariant through the Fourier-Mellin transform. While mutual information could be employed here, we choose instead to apply it to the lower dimension phase spectrum generated by the complex multiresolution wreath product transform of the Fourier-Mellin amplitude spectrum. The phase information of this transform adequately preserves edges even at lower resolutions while permitting at the same time, a reduction in the computational burden. Brodatz textures and ORL faces are used to demonstrate the capability of this algorithm.

1. INTRODUCTION

Image classification has wide applications in many areas and is a basic task in the field of computer vision. Typically, the classification problem requires comparing a given image with a set of reference images which are representative of several classes. Assigning an image to a specific class then permits matching the image with only a subset of the database. Accordingly, classification may be considered a coarse-level matching of images into one of the many prescribed classes. For identification, a finer-level matching within the class may then be employed [3]. Implementing an image classifier requires identifying appropriate features usable for measurement and a suitable metric for measuring the distance between the test and reference images. The class of the reference image, which corresponds to the minimum distance is assigned to the test image. Some examples of this scenario are face identification [8], fingerprint classification [3] and searching through a database [9].

Some key concerns in such applications are robustness, discrimination capability and computation efficiency. Consider for example, the face identification problem. Images of the same person taken at different times may have many differences due to factors such as facial expression, lighting variation and pose (even when frontal poses are assumed). In addition, there may be shifts from one image to another and imaging noise may be present. Hence, robustness and accuracy are important elements in any classification technique. Computation efficiency is critical, especially in real-

time recognition systems where large size data sets may be involved.

In this paper we describe an integrated algorithm for classification that addresses the issues described above. The Fourier-Mellin transform is used to provide a translation and rotation invariant representation of the test image. We propose to use mutual information on the Fourier amplitude as the metric for classification. In order to address the issue of computation efficiency we employ the cyclic group-based wreath product transform (WPT) [2], [6] and specifically make use of its phase characteristics at coarser resolutions. Using the quadtree decomposition, we maintain locality of the phase which also contributes to the classification accuracy. On combining these techniques, we demonstrate its validity with experiments in Brodatz textures discrimination and face identification.

The paper is structured as follows. In Section 2, after a brief discussion of similarity measures between images, we propose mutual information in Fourier magnitude for such a measure. The Fourier-Mellin transform is modified for application to our experiments and that is summarized in Section 3. In Section 4 we describe the multiresolution representation of these invariant transforms with the phase of the WPT. The integrated framework of the algorithm is described in Section 5 where we give experimental results for Brodatz textures and face classification. Discussion and conclusions are given in Sections 6 and 7 respectively.

2. MUTUAL INFORMATION IN FOURIER MAGNITUDE

Correlation coefficients and moment based metrics have been widely adopted in practice for the determination of similarity. Ideally, such measures should be akin to that of the human visual system. Recently, mutual information (MI) has been proposed as a new similarity metric for image registration problems [5], [12]. MI measures the statistical dependency between two data sets. Assume A and B are two random variables, and $p_A(a)$, $p_B(b)$, $p_{AB}(a, b)$ their marginal probability density functions and joint probability density function, respectively. MI between A and B is defined as:

$$MI(A, B) = H(A) + H(B) - H(A, B), \quad (1)$$

where

$$H(A) = - \sum_a p_A(a) \log p_A(a), \quad (2)$$

*Supported by NASA/JOVE Grant NAG8-1281.

$$H(A, B) = - \sum_{a,b} p_{AB}(a, b) \log p_{AB}(a, b). \quad (3)$$

When MI is used to measure the similarity between images, the joint and marginal *pdfs* need to be estimated by normalized joint and marginal image histograms.

The mutual information measure can be considered quite general since it makes very few assumptions about the relationship between the image intensities or the nature of the imaging process. It does not assume a linear or functional dependence but only a statistical one. The earliest work in the use of MI for recognition appears to be that by Viola [12] and Maes *et al.*, [5]. It rests on the idea that the MI of corresponding pixel values is maximal when the two images are geometrically aligned. While MI has been used primarily in the spatial domain, we propose to generalize it to the Fourier magnitude domain.

We give a simple example here to justify the motivation for using MI with Fourier magnitude. A set of 5 images are synthesized from an image (scale 0) with the *db4* wavelet, using lowpass reconstructions at scales 1 through 5. The original and synthesized images are shown in Figure 1(a). The normalized correlation coefficient (NCC) and MI between the synthesized and the original images are calculated both in the spatial and Fourier magnitude domains. Results are shown in Figure 1(b). We observe that NCC and MI in the spatial domain, as well as MI in the Fourier magnitude domain have the same general characteristics that we might expect of our visual system. NCC in Fourier magnitude shows higher correlation for less correlated images and hence would be inappropriate for detecting similarity.

3. TRANSLATION AND ROTATION INVARIANT DESCRIPTION

As is well known, an image and its translated, rotated and scaled version are closely related after application of the Fourier-Mellin transform [1], [10]. We modify this slightly to get a translation and rotation invariant transform. Let $f_1(x, y)$ be the original image, and $f_2(x, y)$ its translated and rotated version. Then

$$f_2(x, y) = f_1(x \cos \theta_0 + y \sin \theta_0 - x_0, -x \sin \theta_0 + y \cos \theta_0 - y_0), \quad (4)$$

where (x_0, y_0) is the translation and θ_0 the rotation angle. Taking Fourier transforms we get

$$F_2(\xi, \eta) = e^{-j2\pi(\xi x_0 + \eta y_0)} \times F_1(\xi \cos \theta_0 + \eta \sin \theta_0, -\xi \sin \theta_0 + \eta \cos \theta_0). \quad (5)$$

Letting M_1 and M_2 represent the magnitudes of F_1 and F_2 , we have

$$M_2(\xi, \eta) = M_1(\xi \cos \theta_0 + \eta \sin \theta_0, -\xi \sin \theta_0 + \eta \cos \theta_0). \quad (6)$$

Now it is clear that M_2 can be obtained by rotating M_1 counterclockwise, by an angle θ . To decouple this rotation, we first transform the cartesian coordinates (ξ, η) to polar coordinates (ρ, θ) , to get

$$M_2(\rho, \theta) = M_1(\rho, \theta + \theta_0). \quad (7)$$

Therefore rotation is reduced to translation in the magnitude polar domain. By taking the Fourier transform in the second variable, we have

$$P_2(\rho, \mu) = e^{-j2\pi\theta_0} \times P_1(\rho, \mu). \quad (8)$$

Finally, letting MP_1 and MP_2 represent the magnitudes of P_1 and P_2 , we get

$$MP_2(\rho, \mu) = MP_1(\rho, \mu). \quad (9)$$

Hence, the combination of the above transforms leads to a translation and rotation invariant descriptor of the image.

4. MULTIREOLUTION REPRESENTATION

We propose to apply MI to the translation and rotation invariant description generated. However since the computational complexity for searching through large data banks can be high, we consider techniques for reducing the dimensionality of the problem. We choose specifically, the cyclic group-based WPT for application to the Fourier magnitude spectrum. The WPT is a 1-dimensional block transform that provides a multiresolution representation of an image. The quadtree-based complex WPT, extracts *local* edges recursively from the lower scales, thus providing a multi-scale edge representation. In our analysis, we specifically use the phase information at the coarser scales. In the quadtree scanning scheme, the phase at any scale is determined from the complex spectrum generated from scanning 2×2 subblocks. For this subblock consisting of four elements x_0, x_1, x_2, x_3 scanned in clockwise order, the phase is the angle θ associated with the complex spectrum in the 4-point discrete Fourier transform:

$$\theta = \arctan \frac{(x_3 - x_1)}{(x_0 - x_2)}. \quad (10)$$

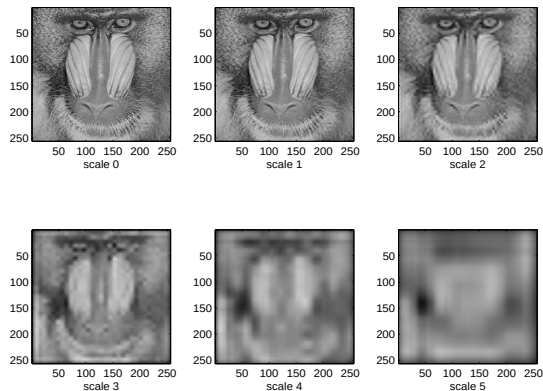
Compared to the spectra generated with other wavelets, WPT phase appears to maintain a good balance between inheriting the approximations and the details of an image, probably due to its locality.

5. IMAGE SEARCHING ALGORITHM AND EXAMPLES

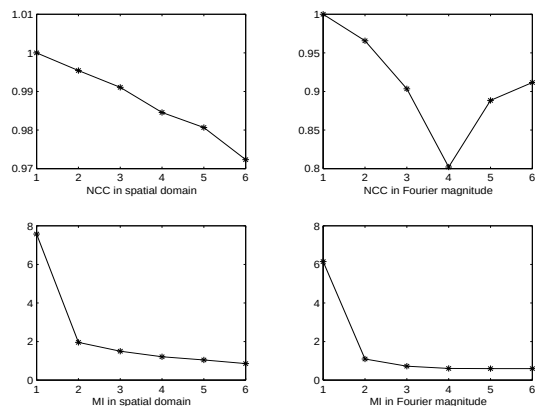
5.1. Integrated matching framework

The integrated framework for classification consists of two steps. In the first, we construct a set of template parameters for each of the classes in a given data bank. That is, a representative image for each class is first transformed to generate an amplitude spectrum which is translation and rotation invariant. After application of the WPT, the phase spectra at one of the coarser scales—chosen here of size 32×32 —is kept as the template parameters. If in a class, several images are selected as representative, then an average of the phase spectrum is assumed to provide a better signal-to-noise ratio.

The second step is the matching process. Phase parameters of the image to be classified are obtained like those for the template images. The MI between the WPT phase of the test image and that of templates corresponding to



(a) Multiscale images of Baboon



(b) Comparison of two distance metrics

Figure 1: MI and NCC

different classes is calculated; the one achieving the largest MI is chosen as the matched template and the image classified accordingly. The capability of the algorithm is tested by application to two classification problems.

5.2. Brodatz textures recognition

In this example, we have a set of 91 Brodatz textures. They are obtained from 13 photographs in the Brodatz album, each photograph having been rotated with 7 different angles and then scanned [13]. Thus the digital texture images of size 512×512 , belong to 13 classes and have some translation and rotation within each class. We use the first three images of each class for construction of the template parameters, and the remaining 4 images to test the algorithm.

In constructing the templates, we first window the $N \times N$ image with a rotationally symmetric Gaussian window with a standard deviation of $N/2 - N/4$ ($N = 512$) in order to minimize boundary effects. The windowed data is Fourier transformed and its magnitude subjected to another Gaussian window in order to reduce the lowpass components. The data is then resampled in polar coordinates, using bicubic interpolation. A Fourier transform is taken along the angle direction, and a Hanning window applied to accentuate the high frequency. The process is repeated for 2 more images and the data averaged. Finally, the WPT is applied to the average and the WPT phase parameters at the scale corresponding to the 32×32 spectrum are used as the template for that class.

The remaining 52 images are used to test the validity of the algorithm. Each of them is processed in the same way as above. Then the MI between the phase of the test image and that of the 13 templates is calculated. The template class achieving the largest MI is assigned to the test image. In our experiments, all the test images were correctly classified. We also used the NCC to test for classification. Here too, all the images were correctly classified. This agreement

is predictable since the data had been made translation and rotation invariant. Other multiresolution representations, such as Haar and *db4* wavelets were also used in the determination of MI. The classification rate with all the 32×32 detail coefficients was low.

The robustness of the algorithm was tested as follows: Addition of white Gaussian noise with zero mean and 100 variance (which corresponded to a peak signal-to-noise ratio of 28dB, and a signal-to-noise ratio of 13dB) to the test images resulted in a correct classification of all the test images. The algorithm proposed here has several advantages over the method used in [9] for the same problem. Using MI as the discriminant function makes the result insensitive to such operation as histogram equalization. Furthermore, the smaller parameter size of the high scale of the WPT greatly reduces the computation burden.

5.3. Face identification

Face identification is a problem of wide interests where many different methods have been proposed, a few of which are [8], [11]. For our work here, we applied our algorithm to the ORL faces data set [11]. This set of 400 images consists of 10 images of each of 40 individuals.

Class templates were constructed as follows. We first choose 4 images for each face, 2 corresponding to direct to frontal shots (most parts of both ears were shown), the other 2 corresponding to side frontal shots (most parts of one ear were not shown). When lacking images in either of these two categories, at least 2 images were used to serve both categories. For these images Fourier transforms were taken along a horizontal direction, since we were concerned about horizontal translation and the desire to fuse 2 side frontal shots. The transformed data was windowed with a Hanning window and the WPT applied. The 32×32 phases of the WPT transform of the 2 categories were averaged and kept as the template. In this way, 80 templates for 40 individuals

were generated.

The other nearly 240 images were used as test samples. In our experiments, close to 98% images were correctly classified when 32×32 phase coefficients were used. In the experiments in [11] which are based on Hidden Markov models, half the total number of images were used to construct the templates; a 95% recognition rate was achieved after parameter optimization.

6. DISCUSSION

Robustness and discrimination capability, while seemingly opposing requirements at first glance, may not be so entirely. In the face identification example, robustness requires an insensitivity to minor variations in some parts, while the discrimination property is satisfied by a high sensitivity to different profiles. Mutual information seems to achieve both characteristics quite well, in both spatial and Fourier amplitude domains.

To this capability, we add the phase information from the WPT and the corresponding reduction in data provided by the WPT multiresolution. While the attributes of phase and phase-only filters has been well recognized [7], [1], our experiments confirm the phase discrimination power and robustness. As demonstrated, the multiresolution phase maintains the discrimination capability quite well upto the 32×32 size scale. At the same time, it also preserves the local edges well enough, thereby making the algorithm robust to minor variations. This is different to other wavelet transforms. For the *db4* and even the Haar wavelets, the detail coefficients (horizontal, vertical and diagonal) are sensitive to variations in the image. Therefore, two images of the same person are difficult to fuse. On the other hand, were approximation coefficients used, that would make the discrimination more difficult for dissimilar images.

Finally, a few words about the numerical robustness of the algorithm. For natural images, its Fourier magnitude has a large dynamic range, usually extending to the 5-6th order. To compress the dynamic range of the Fourier magnitude, a Hanning window or log-magnitude spectra is typically employed [1], [10]. In our experiments, the Hanning window gave a better performance. The reason being that the logarithm eliminates most middle to high frequency components of the image, which correspond to image details, while the Hanning window reduces just the low frequencies, keeping most of the details in the image.

To achieve translation and rotation invariance, we used the Fourier magnitude of the image thereby losing the information contained in the phase. While this may seem counter productive, we argue that for natural images the probability of two images having the same magnitude but different phase would be low; that is why the image phase retrieval problem is approximately solvable [4].

7. CONCLUSION

A new integrated framework for image classification is described. This method employs the Fourier-Mellin transform to achieve translation and rotation invariance. MI is applied to the phase information of the WPT of the Fourier-Mellin magnitude. This is seen to achieve both discrimina-

tion and robustness, even at the coarser scales, which then contributes to computation efficiency. The validity of the algorithm is demonstrated by experiments in classification with Brodatz textures and faces.

8. REFERENCES

- [1] Q.S.Chen, M.Defrise and F.Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition", IEEE Trans. Pattern Anal. Machine Intell., vol.16, no.12, pp.1156-1168, 1994.
- [2] R.Foote, G.Mirchandani, D.Rockmore, D.Healy, and T.Olson, "A wreath product group approach to signal and image processing: part I — multiresolution analysis", to appear in IEEE Trans. on Signal Processing, 2000.
- [3] A.K.Jain, S.Prabhakar and L.Hong, "A multichannel approach to fingerprint classification", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.21, no.4, pp.348-359, 1999.
- [4] X.Luo, M.Zou and Z.Chai, "Phase retrieval using an approximate blind deconvolution algorithm", Proceedings of International Conference on Neural Networks and Signal Processing, pp.805-810, Guangzhou, 1993
- [5] F. Maes, A.Collignon, D.Vandermeulen, G.Marchal and P.Suetens, "Multimodality image registration by maximization of mutual information", IEEE Trans. Med. Imag., vol.16, no.2, pp.187-198, 1997.
- [6] G.Mirchandani, R.Foote, D.Rockmore, D.Healy and T.Olson, "A wreath product group approach to signal and image processing: part II-convolution, correlation, and applications", to appear in IEEE Trans. on Signal Processing, 2000.
- [7] A.V.Oppenheim and J.S.Lim, "The importance of phase in signals", IEEE Proc., vol.69, no.5, pp.529-541, 1981.
- [8] P.J.Phillips, "Matching pursuit filters applied to face identification", IEEE Trans. Image Processing, vol.7, no.8, pp.1150-1164, 1998.
- [9] R.W.Picard and T.Kabir, "Finding similar patterns in large image databases", Proceedings of ICASSP, vol.5, pp.161-164, 1993.
- [10] B.S.Reddy and B.N.Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration", IEEE Trans. Image Processing, vol.5, no.8, pp.1266-1271, 1996.
- [11] F.Samaria and A.Harter, "Parameterization of a stochastic model for human face identification", 2nd IEEE Workshop on Applications of Computer Vision, Sarasota, 1994.
- [12] P.Viola, Alignment by Maximization of Mutual Information, Ph.D thesis, M.I.T. Artificial Intelligence Laboratory, 1995.
- [13] A.G.Weber, "The USC-SIPI image database: version 5", USC-SIPI Report No.315, 1997.