Size, Scale and Rotation Invariant Proposed Feature vectors for Trademark Recognition

Muhammad Faisal Zafar[1] and Dzulkifli Mohamad[2]
Faculty of Computer Science & Information Systems,
University Technology Malaysia, Engineering,
81300 Skudai, Johor, Malaysia
Tel. +60-7-550-3814, Fax.: +60-7-556-5044
e-mail : [1] faisal@gmm.fsksm.utm.my, [2] dzul@fsksm.utm.my

Abstract: The classification and recognition of two-dimensional trademark patterns independently of their position, orientation, size and scale by proposing two feature vectors has been discussed. The paper presents experimentation on two feature vectors showing size-invariance and scale-invariance respectively. Both feature vectors are equally invariant to rotation as well. The feature extraction is based on local as well as global statistics of the image. These feature vectors have appealing mathematical simplicity and are versatile. The results so far have shown the best performance of the developed system based on these unique sets of feature. The goal has been achieved by segmenting the image using connected-component (nearest neighbours) algorithm. Second part of this work considers the possibility of using backpropagation neural networks (BPN) for the learning and matching tasks, by simply feeding the feature vectosr. The effectiveness of the proposed feature vectors is tested with various trademarks, not used in learning phase.

1. Introduction

The trademarks matching task is a very interesting challenge in the area of image processing and pattern recognition as it involves with a complex mixture of graphics and texts. Pattern recognition needs as basis knowledge about the object. Image features that capture the essential traits of an object and are insensitive to different procedural changes are ideal for recognition or matching. In this research work, some unique features are proposed, which involve some simple ratios of the image pixels.

The selection of appropriate features is the key to solve the problem. The success of any such practical system depends critically upon how far a set of appropriate numerical attributes or features can be extracted from the object of interest for purpose of matching or recognition [13]. Invariant features form a compact, intrinsic description of an object, and can be used to design recognition algorithms that are potentially more efficient [12].

There exists a vast amount of literature on shape comparison but unfortunately the majority of the proposed methods are inappropriate for classes of objects as large and complex as *trademarks[1], which are the objects of interest* to this paper. A search of the literature has found very few previous attempts to solve either of the problems of recognizing or matching logos. However, recently attempts have been made to solve the problem of matching or identifying trademarks using 2-D Fast Fourier Transform [13], Invariant moments[3], Hotelling Transform [15], and Eigenvector Modeling technique [4]. But it is argued that the moment invariants are not good image features [1], as they are sensitive to noise and suffer from information suppression, loss, and redundancy. Hotelling Transform is not invariant under image scaling, rotation and translation [15]. And eigenvector modeling technique is invariant to translation and scaling only [4].

As regards the choice of classifier, it is now established that multilayered neural networks are able to match, and often improve upon, the performance of conventional classifiers [5], [8], [9] in a large number of applications, including character recognition [6], [7], and face recognition [14]. It was therefore decided to carry out work to assess the usefulness of BPN for matching and recognizing trademarks.

Three sets 34-image set, 40-image set, and 50-image set of scanned images of trademarks were considered for case study. As an input to the system, sample logos were saved in gray mode in raw format. Then to extract features, every image after thresholding was segmented using connected component algorithm. Three different patterns (regarding rotation, scaling, translation and variation in size) of each logo were use for feature extraction. The normalized feature sets then were fed to the BPN model for training/learning. Finally the efficiency of the system was tested using different samples of logos, not used in training phase.

This paper

a. discusses the extraction of some simple features and scanner based interface developed for supplying hybrid feature vectors to the BPN models.
b. describes the experiments mentioned above and presents the performance statistics of BPN models. The authors have made an attempt to critically analyze the experimental observations and have presented their views.

2. Models And Algorithms

2.1 Image Thresholding

Thresholding is one of the most important approaches to segment an image which contains an object having homogeneous intensity and a background with a different intensity level. As in the case of trademark, the digitized image has the intensity level from 0 to *255*. An image as such, can be segmented into two regions by simple thresholding. To make segmentation more robust, the threshold should be automatically selected by the system [17]. Therefore, **automatic thresholding** has been applied.

It analyzes the gray value distribution in an image, by using a histogram of the gray values, to select the most appropriate threshold. Figure 1b below shows the result iterative threshold selection [17].

2.2 Image Segmentation

The goal of image segmentation is to process the data of an acquired image so as to arrive at a partitioning of the image plane. The compartments resulting from this process are called *regions* or *segments*. The purpose of *component labeling* is to assign unique labels – often numbers – to all components The classical way to handle this is to scan the region map row-by-row by a mask. Each black run is assigned an integer number called label and the labels of connected black runs must be the same. Segments are then distinguished by their labels. The component labeling procedure identifies black runs for each row, analyzes the connectivity of black runs between the current row and the previous row, assigns labels to black runs of the current row and/or updates labels of black runs of the previous rows, and defines sements of a binary image in terms of their coordinates. Figure 1c below shows the result.

a b c)

Figure 1: Component labeling; a) A bitmap containing connected components. b) Result of automatic thresholding c) Result of component labeling

2.3 Backpropagation Neural Network

Backpropagation neural network (BPN) has been implied for classification and recognition purpose. In such an architecture, information is processed as follows: the outputs from the processing element (PEs) of the input layer, after multiplying with the corresponding interconnecting weights, serve as inputs to the PEs of the hidden layer. The output from the PEs of the hidden layer, after multiplication with corresponding interconnecting weights, serve as input to the PEs of the output layer. A bias processing element supplies a constant output of +1 to all the PEs of the hidden and the output layer. BPN models with more than one hidden layer process information on the same principle [10].

For a PE, the output is typically a function of the sum of input into it. For BPN models, PEs with linear, sigmoid and hyperbolic tangent transfer function are typically used depending upon the position of the PE in the network architecture and the nature of the problem under consideration. With sigmoid PEs in the hidden and the output layer, the output of a PE in the network's output layer becomes highly non-linear function of the input to the network. For only one PE in the output layer, the network can be thought of as representing a non-linear function of the inputs. For ***n*** PEs in the input layer and ***m*** in the output layer, the network represents ***m*** non-linear functions of ***n*** variable.

3. Experiments

3.1 Sample Preparation

Fifty logos were scanned and digitized into a gray form. The digitized images ranged in size from 100 by 100 pixels to 120 by 120 using scanning resolution 200 dpi. Three patterns of every image were created for different rotations, scales, and sizes. Then these were converted into binary form by applying automatic thresholding. The whole algorithm steps then were coded into C language.

3.2 Hybrid Feature Selection

Connected component algorithm was applied to find out the segments in the image, as stated earlier. The developed software, for the binary image, searched for only "on" and "off" pixels and extracted the following information regarding the image:

- No. of total regions or segments (N).
- No. of pixels in each segment (X_i, $i = 1,2,\ldots,N$).
- Total no. of active image pixels by adding all the region pixels ($T = \Sigma X_i$).
- No. of background pixels (B).

The proposed features which are given below:

1. Ratio of norm of all image segments to total active image pixels i.e
   $\|V\|_2 / T$, where $\|V\|_2 = \sqrt{\Sigma X_i^2}$ $(i = 1,2,..,N)$
2. Ratio of the norm of three biggest image segments to total active image pixels i.e
   $\|V\|_2 / T$, where $\|V\|_2 = \sqrt{\Sigma X^2_{N-i}}$, $(i = 0,1,2)$
3. Ratio of active image area to Sum of differences of all segment area and image area i.e.
   $T / \Sigma (T - X_i)$.
4. Ratio of sum of total active image pixels and total number of segments to greatest segment area i.e
   $(T + N) / X_{N-1}$
5. Ratio of greatest segment area to the total active image area i.e.
   X_N / T
6. Ratio of second greatest segment area to the total active image area i.e.
   X_{N-1} / T
7. Ratio of third greatest segment area to the total active image area i.e.
   X_{N-2} / T
8. Ratio of background pixels to active image area i.e
   (B/T).
9. Ratio of background pixels minus active image area to background pixels plus image area i. e $(B-T) / (B+T)$

All these features are invariant to size rotation, translation. First seven features are also invariant to scaling. That's why experiments were done separately with 9-feature vector (best when scaling is not involved) and with 7-feature vector(making scaling invariant too). These are given below as A and B respectively.

A $\{\|V_a\|_2 / T, \|V_b\|_2 / T, T/W, (T+N)/X_N, X_N/T, X_{N-1}/T, X_{N-2}/T, B/T, (B-T)/(B+T)\}$

B $\{\|V_a\|_2 / T, \|V_b\|_2 / T, T/W, (T+N)/X_N, X_N/T, X_{N-1}/T, X_{N-2}/T\}$

3.3 Feature Normalization

Feature normalization is normally performed to rescale the data to a required range and it is done before matching (training and recognition) process. As Backpropagation Neuraral Network (BPN) has been used as matching technique that uses the unipolar sigmoid function, so it requires the training and the testing data to be in the range of [0,1]. Some times the feature's output is out of range[0, 1], thus the linear scaling to unit variance technique has been used. The formula for this is[16]:

$$x' = (x-l)/(u-l)$$

where l and u are lower and upper bounds respectively for feature component x.

3.4 Learning / Training

In the BPN, sigmoid PEs were used in the hidden and the output layer. Thirty-four and forty output layer PEs corresponded to Thirty-four logos and forty logos to be recognized. For example, a 'high' output value on the first PE in the output layer and 'low' on the others would mean that the network classifies the input as a 'logo1'. As an other example, a network output vector of [0.00 0.01 0.16 0.02 0.09 0.96 0.13 0.00 --- 0.15 0.04] would be translated as the model classifying the input image as 'logo6'.

Experiments were started with a strict convergence criterion: training would be stopped only when the network classified all the training samples correctly. While checking the network's performance during training, an output layer PE's output value of 0.9 was translated as 'high'. Thus, for this criterion , an out put vector [0.00 0.03 0.06 0.12 0.09 0.16 0.03 0.91 0.17 --- 0.15 0.14] for logo 8 in the training sample set would be termed as proper classification; a 0.85 instead of 0.91 in the out put vector would render the training input as not properly recognizable till that stage in the training process.

4. Performance of BPN Model

BPN model was evaluated on samples, which were not used in the initial process of setting up the training data set. This was done keeping in view the eventual aim of developing an efficient recognition model, the quality of which related to translate images irrespective of size, translation, rotation etc. Three different sets of images were being experimented; one with 34 images, 40 images, 50 images having three different samples each. For developed *BPN model*, the debate on a valid high PE output in the output layer was resolved by evaluating the performance for different decision making criteria. Model was tested with high thresholds of 0.90 and 0.5, using the PE with the highest value above the threshold for input classification. Another criterion used in translating the BPN model's outputs was to eliminate the concept of threshold and simply use the highest value. Note that the first two criteria will always have the possibility of a **recognition failure**: a network decision of not attributing any logo to the input image. The last criteria will eliminate this somewhat desirable feature in the decision making process. Tables 1 and 2 below present the statistics. CRs, MRs, FRs, and RFs are abbreviation for Correct Recognitions, Multiple Recognition, False Recognitions, and Recognition Failures respectively.

Table 1: Performance Of Bpn Model with diffirent criteria of interpreting model output for 9-Feature set

9-Feature									
	34-image set			40-image set			50-image set		
	'threshold': 0.9	'threshold': 0.5	'threshold': NONE	'threshold': 0.9	'threshold': 0.5	'threshold': NONE	'threshold': 0.9	'threshold': 0.5	'threshold': NONE
CRs	91.2 %	94%	94%	87.5%	87.5%	90%	86%	86%	88%
MRs	0%	1.6%	1.6%	1.25%	1.25%	2.5%	0%	1%	3%
FRs	4.4%	4.4%	4.4%	7.5%	8.5%	7.5%	10%	11%	9%
RFs	4.4%	0%	0%	3.75%	2.75%	0%	4%	2%	0%

Table 2: Performance Of Bpn Model with diffirent criteria of interpreting model output for 7-Feature set

7-Feature									
	34-image set			40-image set			50-image set		
	'threshold': 0.9	'threshold': 0.5	'threshold': NONE	'threshold': 0.9	'threshold': 0.5	'threshold': NONE	'threshold': 0.9	'threshold': 0.5	'threshold': NONE
CRs	65 %	71%	74%	67.5%	74%	78%	62%	76%	78%
MRs	0%	0%	3%	4%	4%	10%	2%	3%	6%
FRs	15%	20%	20%	7.5%	10%	12%	13%	16%	16%
RFs	20%	10%	0%	21%	12%	0%	23%	4%	0%

5. Analysis of Performance

Although BPN was trained on images having sizes 100x100 and 120x120 and orientations 0, 90cw and 180 degree but it was tested for sizes 80x80 and 110x110 with orientations 0 and 90ccw. In this way about 400 images were tested with three set of images. Exactly Correct recognitions (CRs) for all the sets are from 94% to 86% under different criteria of interpretation for 9-Feature category. 34-image set shows even greater efficiency as it has relatively smaller domain to compete. In case of 7-feature category, the performance for CRs was recorded from 78% to 62% under different criteria of interpretation. It means reduction of features has significant effect on the performance of the model along with scaling factor. It is important to note that the sigmoid functions in the output layer behave as 'smoothed' bipolar switches; the inputs to these bipolar switches are values of the decision functions ([11] provides a detailed explanation of the decision function concept). These decision functions, or decision surfaces, have positive values for a PE's output greater than 0.5 and negative values for PE outputs less than 0.5. Multiple recognitions (MRs) mean that there are more than one outputs for a test image as some images have very close and quite similar features but these are not in a big number. False recognitions (FRs) show that output results in a quite different image or wrong image. The chances for MRs and FRs increase with increase in samples; i. e. 40 and 50 image sets, has been observed in case of 9-Feature category and it is quite understandable. However, FRs were minimum under 7-feature category and performance of the model for CRs was also improved with increase in samples. This shows if number features are reduced then performance increases with increase in samples or patterns.

6. Conclusions

The calculation of descriptors and feature extraction may be considered as two individual processes, but it is important to take into account the fact that they are at the heart of the system. In other words, if the system is successful in representing the input image with well defined features, the most difficult part of the recognition problem may be solved. In this paper, connected component algorithm is used to segment the images and then using different ratios some simple features are introduced, which are invariant to image scaling, rotation and translations. This process could tremendously reduces the complexity task in extracting the features like in other approaches, for example contour based description. There is no difficulty in calculating geometrical properties and the ratios among many numerical pixel values of the segmented image. And if there occurs some noise, loss or redundancy; which is quite possible while scanning and processing the image, even then the system would work properly. In this case, the inherent capability of BPN plays a vital role. Experimental results show the agreement of this method to a level of confidence.

References

[1] Sheng, (1994). "Orthogonal Fourier – Mellin moments for invariant Pattern Recognition", *J. Opt. Soc. Am.* A/Vol. 11, No. 6, 1748–1757.

[2] Cortelazo, (1994). "Trademark Shapes Description by String-Matching Technique", *Pattern Recognition*, Vol. 27, No. 8, 1005–1018.

[3] D. Mohamad and G. Sulong, S S Ipson, (1995). "Trademark Matching Using Moment Invariants", *Second Asian Conference on Computer Vision, Singapore*, 439-444.

[4] D. Mohamad and G. Sulong, (1996). "Trademark Identification Using Eigenvector Modeling Technique" *Proceeding of the IASTED International Conference, Florida, U. S. A.*, 69–73.

[5] A.K. Jain, R.P.W Duin, and J Mao, (Jan. 2000). "Statistical Pattern Recognition: A Review", *IEEE transactions on pattern analysis and machine intelligence*, vol. 22 No. 1, 04-37.

[6] Ahmed, S.M.,et.al, (Nov.1995). "Experiments in character recognition to develop tools for an optical character recognition system", *IEEE Inc. 1st National Multi Topic Conf. proc. NUST, Rawalpindi, Pakistan*, 61-67.

[7] Ho J. Kim, H S. Yang, (1994). "A Neural network capable of learning and inference for visual pattern recognition", *Pattern Recognition*, Vol. 27 No. 10, 1291–1302.

[8] H.C.Yau and M.T.Manry, (1991). "Iterative Improvement of a Nearest Neighbor Classifier", *Neural Networks*, Vol. 4, No. 4, 517-524.

[9] Guyon, L., Poujaud, I., Personnaz, L.,Dreyfus, G., Denker, J., and Le Cun, Y., (June 1990). "Comparing different neural network architectures for classifying hand written digits", *Proc. of the Int. Joint Conf. on Neural Networks, II, IEEE Press, New York*, 127-132.

[10]Freeman, J. A., and Skapura, D. M., (1991). "Neural Networks: Algrithms, Applications and Programming Tecniques", *Addison-Wesley Publishing Company.*

[11] Tou, J. T., and Gonzalez, R. C., (1974). "Pattern Recognition Principles", *Addison-Wesley Publishing Company.*

[12]R. Alferez, and Y. Wang, (June 1999). " Geometric and Illumination Invariants for Object Recognition", *IEEE transactions on pattern analysis and machine intelligence*, vol. 21, No. 6, 505-535.

[13]D. Mohamad (1997). "The Identification of Trademark Symbols Based on Modeling Techniques". *PhD Thesis, FSKSM, University Technology, Malaysia.*

[14]R. Ferand et al, (Jan. 2001). "A Fast and Accurate Face Detection Based on Neural Networks." *IEEE transactions on Pattern Analysis and Machine Intelligence vol. 23 no. 1.*

[15]D. Mohamad and G. Sulong, (1995). "Trademark Matching Using Hotelling Transform", *First International Conference of Engineering Computation and Computer Simulation, Changsa, China.*

*[16]*P. Saad, D. Safaai, S. Mariyam and D. Mohamad, (Oct. 2001). " A comparision of Feature Normalization Techniques on Complex Image Recognition." *Proc. Of The 2nd Conf. On Information Technology in Asia, Kuching, Sarawak, Malaysia.*

*[17]*Jain, Kasturi, Schunck, (1995). " Machine Vision." *McGraw-Hill Series in Computer Science, Singapore.*