OPTIMIZED SPACE FREQUENCY KERNEL FOR TEXTURE CLASSIFICATION

Mahdi Sabri [1], *Javad Alirezaie* [1,2]

1. Department of Electrical and Computer Engineering, Ryerson University,Toronto, Canada
2. Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada
msabri@ee.ryerson.ca, jalireza@ee.ryerson.ca

ABSTRACT

The performance of the Support Vector Machine(SVM) algorithm is highly dependent on the choice of the kernel function suited to the problem at hand. In a Support Vector Machine algorithm feature selection is implicitly performed by kernel function. On the other hand, feature selection is the most important stage in any texture classification algorithm. In this work, the performance of SVM is improved by choosing an optimized space-frequency (SFR) kernel function. The proposed method is evaluated in a two-texture and multi-texture problems. The results are compared with the original SVM and other recently published texture classification methods. The comparison shows a significant improvement in error rates. Improvement of more than 40% in compare with original SVM and about 60% in compare with logical operators (LO) and wavelet co-occurrence features (WCOF) are obtained.

1. INTRODUCTION

Texture classification is motivated by the wide range of different industrial and medical applications. In [1] texture features were used for object recognition in industrial environments. A defect detection in textured materials was proposed in [2] by using texture analysis. Texture features and maximum likelihood classifier were used by Horng *et al.* [3] to classify ultrasonic liver images into three category of normal liver, liver hepatitis and cirrhosis. Texture analysis were used in [4] to detect microfocal lesions in ultrasound liver images. Gurcan *et al.* [5] used higher order statistical texture features for detection of microcalcifications in mammograms. A computer aided diagnosis (CAD) system for automatic abnormalities detection in chest radiographs was introduced in [6] using local texture analysis.

Texture classification algorithms are generally include two crucial steps: 1) feature extraction 2) classification. In feature extraction stage, a set of features are sought that can be efficiently computed and represent discriminative information about the textural characteristics. In the next stage, a classification paradigm is constructed to distinguish between texture features correspondent to different texture classes. Bayesian classifier is known as optimal classifier, but calculation of underlying probability distribution of the problem under study is not practically possible. Specially in the absence of adequate number of training samples. In this work, SVM is used and it is proven to outperform other classification methods [7]. Superiority of SVM originates from its ability to generalize in high dimensional spaces, such as the space which is spanned by texture patterns. The generalization ability of SVM is based on its profound relation to the underlying statistical learning theory. In SVM, instead of minimizing an objective function based on the training samples (such as mean square error), it is attempted to minimize a bound on generalization error (i.e., the error made by the learning machine on test data not used during training). Therefore, an SVM tends to perform well when applied to data outside the training set. SVM achieves this advantage by focusing on the training examples that are most difficult to classify. These "borderline" training examples are called support vectors.

Unlike other texture classification methods, SVM based method dose not necessarily incorporate any external feature extraction method. Kim *et al.* [8] showed the effectiveness of SVM in texture classification problem. In fact, in an SVM, feature extraction is implicitly performed by a kernel function, which is defined as the dot product of two mapped patterns. The main focus of this paper is to utilize a new optimized kernel function and investigate the effectiveness of deploying external features.

In section two the SVM is reviewed. In section three the selection of optimized SFR is discussed. In section four and five experimental results and discussion are presented respectively.

2. SUPPORT VECTOR MACHINE

Let vector $\mathbf{x} \in \chi$ denotes a pattern to be classified, and scalar d denotes its class label $d \in \{\pm 1\}$. In addition, let $O_N = \{(\mathbf{x}_i, d_i), i = 1, 2, ..., N\}$ denote a given set of N training examples. The problem is how to construct a classifier [i.e., a decision function $f(x)$] that can correctly classify an input pattern that is not necessarily from the training

0-7803-8554-3/04/$20.00 ©2004 IEEE.

set. On the assumption of linearly separable case, there exist a linear function of the form

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b \tag{1}$$

Such that $f(\mathbf{x}_i) \geq 0$ for $d_i = +1$, and $f(\mathbf{x}_i) < 0$ for $d_i = -1$. In other words, training examples are separated by the hyper-plane $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0$.

For a given training set, while there may exists many hyper-plane that separate the two classes, the SVM classifier is based on the hyper-plane that maximizes the separating margin between the two classes.

In practice, input patterns ($\mathbf{x}$ of dimension m_0) are unlikely to be linearly separable. Thus, slack variables, denoted by ξ_i, are introduced to relax the separability constrains. Also, according to the Cover's theorem [9] on the separability of patterns, a set of non-linear transforms $\{\mathbf{\Phi}_j(\mathbf{x})\}_{j=1}^{m_1}$ is sought to transfer input patterns into feature space of dimension m_1. Then, the hyperplane acting as decision surface is:

$$\sum_{j=1}^{m_1} w_j \mathbf{\Phi}_j(\mathbf{x}) + b = 0 \tag{2}$$

where $\{w_j\}_{j=1}^{m_1}$ denotes a set of linear weights connecting the feature space to output space, and b is the bias. It can be shown that finding optimal hyper-plane in (2), is equal to minimizing the cost function [10]:

$$\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i \tag{3}$$

ξ_i are slack variables and C is user defined. A large C corresponds to assigning a higher penalty to the training errors. Given the training samples as $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$, the constrains which must be satisfied are:

$$d_i(\mathbf{w}^T\mathbf{\Phi}(\mathbf{x}_i) + b) \geq 1 - \xi_i \;\; for \;\; i = 1, 2, ..., N \tag{4}$$

This constrained optimization problem is solved using the Lagrange multiplier. The Lagrangian function is constructed as:

$$\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i[d_i(\mathbf{w}^T\mathbf{\Phi}(\mathbf{x})_i + b) - 1 + \xi_i] \tag{5}$$

where the nonnegative variables α_i are called Lagrange multipliers. Practically Lagrange multipliers are solved from dual form of (3), which is expressed as:

$$\sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j d_j \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) \tag{6}$$

Subject to:

$$0 \leq \alpha_i \leq C \;\; for \;\; i = 1, 2, ..., N, \quad \sum_{i=1}^{N}\alpha_i d_i = 0 \tag{7}$$

Then, the non-linear SVM classifier is obtained as:

$$f(x) = \sum_{i=1}^{N}\alpha_i d_i \mathbf{K}(\mathbf{x_i}, \mathbf{x}) + b \tag{8}$$

where the function $\mathbf{K}(.,.)$ is defined as:

$$\mathbf{K}(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{\Phi}(\mathbf{x_1})^T\mathbf{\Phi}(\mathbf{x_2}) \tag{9}$$

$\mathbf{K}$ is referred as kernel function.

Most of the training examples in a typical problem are correctly classified by the trained classifier ,i.e., only a few training examples will be support vectors (borderline samples and incorrectly classified samples). For simplicity, let $\mathbf{s}_j, \alpha_j^*, j = 1, 2, ..., N_s$, denotes these support vectors and their corresponding nonzero Lagrange multipliers respectively. The decision function in (8) can be simplified as:

$$f(x) = \sum_{j=1}^{N_s}\alpha_j^* d_j \mathbf{K}(\mathbf{s}_j, \mathbf{x}) + b \tag{10}$$

As it can be seen in (6) and (10), The nonlinear mapping $\mathbf{\Phi}(.)$ never appears explicitly in either dual form of SVM training problem or the resulting decision function. The mapping $\mathbf{\Phi}(.)$ only enters the problem implicitly through the kernel function $\mathbf{K}(.,.)$, thus it is only necessary to define $\mathbf{K}(.,.)$ which implicitly defines $\mathbf{\Phi}(.)$. Our proposed kernel function is presented in the next section.

3. SFR KERNEL FOR TEXTURE CLASSIFICATION

The performance of SVM classifier is strictly dependent on the choice of a SVM Kernel $\mathbf{k}(.,.)$ suited to the problem at hand. In this work, we choose $\mathbf{k}(.,.)$ based on an approach similar to the non-stationary signal classification algorithm introduced in [11]. Assume that Cohen's group SFR representation of pattern $\mathbf{x}(n)$ is shown by $C_{\mathbf{x}}^{\phi}(n, \Omega)$ parameterized by its SFR kernel ϕ. Where Ω is discrete frequency. Given two patterns $\mathbf{x}(n)$ and $\mathbf{x}'(n)$ the Gaussian radial basis function kernel of SVM becomes:

$$\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = exp\frac{-1}{2\sigma^2}[\sum_{n=1}^{m_0}\sum_{\Omega=1}^{m_0} \mid C_{x_i}^{\phi}(n, \Omega) - C_{x_j}^{\phi}(n, \Omega) \mid^2] \tag{11}$$

Many parametric SFR kernel (ϕ) have been proposed in literature, and an efficient choice is the family of radially Gaussian kernels, defined in the ambiguity plane as:

$$\phi(\rho, \varphi) = exp(-\frac{\rho^2}{2c(\varphi)^2}) \tag{12}$$

Where ρ and ϕ are the polar coordinates. and the contour function is:

$$c(\varphi) = a_0 + \sum_{p=1}^{P_{max}}[a_p\cos(2p\varphi) + b_p\sin(2p\varphi)] \tag{13}$$

The SFR kernel parameters are then a_0', a_p and b_p, with $p = 1, ..., p_{max}$ and $\theta = [a_0^{'}, a_1, b_1, ..., a_{p_{max}}, b_{p_{max}}, \sigma]$

In optimization procedure, we consider O_L and O_T to be subsets of approximately equal size obtained by randomly partitioning training set O_N in two parts, each containing elements from both classes. There are T_{+1} samples in O_T labelled +1, and T_{-1} samples labelled -1. Also, There are L_{+1} samples in O_L labelled +1, and L_{-1} samples labelled -1. In order to optimally obtain θ, an optimization criterion $P(\theta \mid O_N)$ is introduced which is minimized with respect to θ via a standard optimization procedure. For a given training set O_N and given kernel parameters θ, $P(\theta \mid O_N)$ is calculated as follows:

- Step1) Use the set O_L to train the SVM classifier.
- Step 2) For each element in O_T, compute the decision function $f_{O_L}(.)$ using (11)
- Step3) Compute the empirical mean m_{+1}, m_{-1} and standard deviation s_{+1}, s_{-1} of $f_{+1}^{(i)}$ and $f_{-1}^{(i)}$
- Step 4) Compute the criterion introduced in [11]:

$$P(\theta \mid O_T, O_L) = \frac{1}{2}Q(\frac{m_{+1}}{s_{+1}}) + \frac{1}{2}Q(\frac{m_{-1}}{s_{-1}}) \qquad (14)$$

where

$$Q(v) = \int_{v}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}) \qquad (15)$$

steps one to four are repeated R times with different subsets O_L and O_T, and the final criterion $P(\theta \mid O_N)$ is obtained as the average of $P(\theta \mid O_T, O_L)$ over all the subsets tested. When the optimal $\theta^{'} = \arg\min_{\theta} P(\theta \mid O_N)$ is obtained, the classifier is trained over full learning set O_N using the optimal SFR kernel.

4. RESULTS

4.1. Comparison with Original SVM method

The optimized SVM (OSVM) has been compared with original SVM by classifying texture samples(Fig. 1) selected from Brodatz [12] and VisTex [13] database. Classifiers were trained by randomly selected windowed subimages that were not included in the test images. Approximately 1.7 percent of available input patterns were used during training phase. The original SVM shows the optimal classification rate at window size 17×17. Classification error rates are presented in table 1 for different window sizes of texture samples. The error rate of the proposed method is 44% better than the original SVM on average.

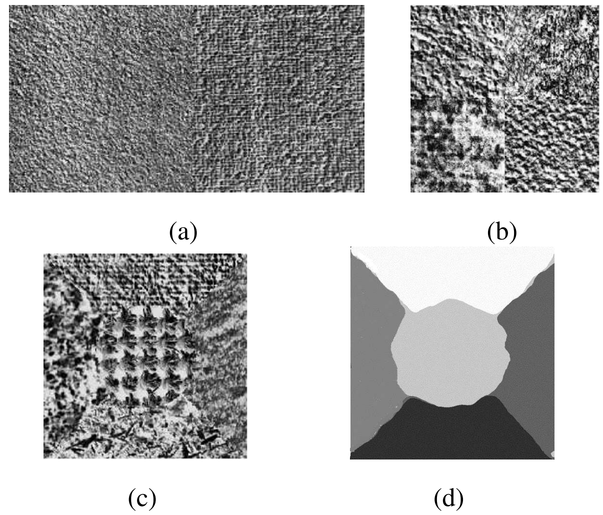

Fig. 1. a) D4,D84 from Brodatz b)D4, D9, D19, D57 from Brodatz c)Fabric07, Fabric09, Leaves03, Misc02, Sand02 from VisTex d) segmentation results for image in (c)

Table 1. Error rates (percent) for multitexture problem.

win	Fig1.a		Fig1.b		Fig1.c	
size	SVM	OSVM	SVM	OSVM	SVM	OSVM
13	9.4	7.6	21.2	11.9	20.0	12.3
17	8.6	5.3	17.9	8.7	16.5	9.2
21	7.3	3.9	16.1	5.8	14.7	6.7

4.2. Comparison with other methods

The proposed method was compared with two recently published texture classification methods, Logical Operators(LO) [14] and wavelet co-occurrence features (WCOF) method [15]. Results are listed in table 2. In the first experiment the proposed OSVM shows an average error rate of 6% which is 65% better than LO method. In the second experiment OSVM is compared with WCOF. The error rate of OSVM(3%) is 62% better than WCOF(8%).

5. CONCLUSION AND FUTURE WORK

This paper described an improvement on SVM classification method for the purpose of texture classification by introducing a SFR kernel. The proposed kernel creates a feature space with more chance of separability at lower dimension. The excellent performance on different textures where achieved. It was shown that the proposed method outperforms other recently published methods. Since the SVM is originally introduced for the case of two-class problems, it can be an ideal solution for several medical application where detection of abnormal tissues from normal tissues are

Table 2. Comparison of correct classification rate for each individual texture in two different multi-texture classification problem

7 texture from Brodatz			7 texture from VisTex		
Texture	LO	OSVM	Texture	WCOF	OSVM
D15	89	89	Bark.06	92	89
D19	97	100	Clouds.01	94	100
D52	81	100	Fabric.17	97	100
D65	84	100	Grass.01	78	96
D74	73	81	Leaves.12	91	94
D82	86	98	Misc.02	97	100
D84	72	95	Sand.02	96	100

desired. In general, authors are looking into the applications of the proposed method in medical imaging in specifically in breast cancer detection.

6. REFERENCES

[1] J. Marti, J. Batlle, and A. Casals, "Model-based objects recognition in industrial environments for autonomous vehicles control," in *International Conference on Robotics and Automation*, 1997, Proc. IEEE.

[2] A. Kumar and G. Pang, "Defect detection in textured materials using gabor filters," in *Industry Applications Conference*, 2000, Proc. IEEE.

[3] M. H. Horng, Y. N. Sun, and X. Z. Lin, "Texture feature coding method for classification of liver sonography," *Computerized Medical Imaging and Graphics*, vol. 26, pp. 33–42, 2002.

[4] J. S. Bleck, U. Ranft, M. Gebel, H. Hecker, M. Westoff-Bleck, C. Thiesemann, S. Wagner, and M. Manns, "Random field models in the textural analysis of ultrasonic images of the liver," *IEEE Transaction on Medical Imaging*, vol. 15, pp. 796–801, 1996.

[5] M. N. Gurcan, Y. Yardmici, A. E. Cetin, and R. Ansari, "Detection of microcalcification in mammograms using higher order statistics," *IEEE Signal Processing Letters*, vol. 4, pp. 213–216, 1997.

[6] B. V. Ginneken, S. Katsuragawa, B. M. T. H. Romeny, K. Doi, and M. A. Viergever, "Automatic detection of abnormalities in chest radiographs using local texture analysis," *IEEE Transaction on Medical Imaging*, vol. 21, pp. 139–149, 2002.

[7] B. Scholkopf, K. Sung, C.J.C. Burges, F. Girosi, P. Niyogi, T. Pogio, and V. Vapnik, "Comparing support vector machines with gaussian kernels to radial basis function classifiers," *IEEE Transaction on Signal Processing*, vol. 45, pp. 2785–2765, 1997.

[8] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Support vector machine for texture classification," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1542–1550, 2002.

[9] T.M. Cover, "Geometrical and statistical properties of systems of linear inequalities with application in pattern recognition," *IEEE Transaction on Electronic Computers*, vol. 14, pp. 326–334, 1965.

[10] B. Scholkopf, C. Burges, and A. Smola, *Advances in kernel methods: Support Vector Learning*, MIT Press, Cambidge,MA, 1999.

[11] M. Davy, C. Doncarli, and G. F. Boudreaux-Bartels, "Improved optimization of time-frequency based signal classifier," *IEEE Signal Processing Letters*, vol. 8, pp. 52–57, 2001.

[12] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, Newyork, 1966.

[13] MIT Vision and Modeling Group, 1998.

[14] V. Manian, R. Vasquez, and P. Katiyar, "Texture classification using logical operators," *IEEE Transaction on Image Processing*, vol. 9, pp. 1693–1703, 2000.

[15] S. Arivazhagan and L. Ganesan, "Texture classification using wavelet transform," *Pattern Recognition Letters*, vol. 24, pp. 1513–1521, 2003.