

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

**ANSI***net*

Asian Network for Scientific Information  
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

## Facial Expression Recognition Using Improved Support Vector Machine by Modifying Kernels

<sup>1</sup>W. Liejun, <sup>2</sup>Q. Xizhong and <sup>1</sup>Z. Taiyi

<sup>1</sup>Department of Information and Communication Engineering,  
Xi'an Jiaotong University, People's Republic of China

<sup>2</sup>School of Information Science and Engineering, Xinjiang University, People's Republic of China

---

**Abstract:** This study proposes a novel facial expression recognition approach based on improved Support Vector Machine (SVM) by modifying kernels. The idea comes from the work of Amari that enlarging the spatial resolution around the margin by a conformal mapping, such that the separability between classes is increased. Experiments on Japanese Female Facial Expressions (JAFFE) database show that the Classification Accuracy Rate (CAR) is remarkably improved after modifying the Gaussian kernel. Experiments also show that the importance of selecting an appropriate parameter when modifying the kernel.

**Key words:** Support vector machine, facial expression recognition, kernel, riemannian geometry

---

### INTRODUCTION

Facial expressions reflect emotions, mental activities, social interaction and physiological signals. Over the past years, automatic facial expression recognition has become an active research area that finds potential applications in areas such as human emotion analysis, human-computer interfaces and image retrieval (Fasel and Luetttin, 2003).

Support Vector Machine (SVM) have become very popular as methods for learning from examples in science and engineering. Some advances of this subject matter have been provided (David and Sanchez, 2003; Campbell, 2002; Perez-Cruz and Bousquet, 2004). SVM is properly motivated by statistical learning theory. It is the most well known of a class of algorithms which use the idea of kernel substitution. SVM has many benefits, for example, it is no local minima to complicate the learning process since training involves optimization of a convex cost function (Campbell, 2002). SVM has been successfully applied to a number of applications ranging from face recognition and text categorization to signal processing (Perez-Cruz and Bousquet, 2004).

The performance of SVM depends on the kernels. Based on the structure of the Riemannian geometry induced by the kernel function, Amari (1999) proposes a method of modifying a Gaussian kernel to improve the performance of a SVM. The idea is to enlarge the spatial resolution around the margin by a conformal mapping, such that the separability between classes is increased

(Amari, 1999). Due to the encouraging results with modifying kernel, this study proposes a novel facial expression recognition approach based on improved SVM by modifying kernels. We test the approach on JAFFE expressions database. The results show that the recognition performance (measured by CAR) is remarkably improved after modifying the Gaussian kernel function. Experiments also show that the importance of selecting an appropriate parameter when modifying the kernel.

### FACIAL EXPRESSION RECOGNITION: A BRIEF REVIEW

Here, we give a brief review on the facial expression recognition and the details may be found in the references (Fasel and Luetttin, 2003).

The history of facial expression analysis can be tracked back into the nineteenth century. Darwin demonstrated in 1872 the universality of facial expressions and their continuity in man and claimed among other things (Fasel and Luetttin, 2003). Ekman and Friesen (1971) classified expressions into six basic emotions, that is, happy, sad, surprise, fear, disgust and anger. Then they put forward Facial Action Coding System (FACS). Suwa *et al.* (1978) presented a preliminary research on automatic facial expression analysis from an image sequence. In the 1990s, automatic facial expression analysis research gained much inertia starting with the pioneering study of Mase and Pentland (1991).

Feature extraction is an important step toward facial expression recognition. The well-known techniques include Principal Component Analysis (PCA), Eigenface (Turk and Pentland, 1991) and Fisherface (Belhumeur *et al.*, 1997) etc. In general, these techniques can be classified into two categories, i.e., local feature based and global feature based (Zilu *et al.*, 2006). The former recognize expressions by detecting and localizing the geometry structures of face, mouth and eyebrows etc. The later treat the face images as a whole matrix and extract global features.

After feature extraction, a classifier is designed to classify these features into different categories. According to the Fasel and Luetttin (2003), the classifiers can be distinguished as spatio-temporal based and spatial based. The former includes Hidden Markov Models (HMM), Recurrent Neural Networks (RNN) and spatio-temporal motion-energy templates, etc. For instance, several HMM-based classification approaches can be found in the literature (Otsuka and Ohya, 1998) and were mostly employed in conjunction with image motion extraction methods. The spatial based classifiers includes rule-based neural networks and SVM, etc. These classifiers were either applied directly on face images or combined with facial features extraction and representation methods such as PCA or Gabor wavelet filters, etc.

**CLASSICAL SVM FOR CLASSIFICATION**

**SVM for binary classification:** Here, we briefly describe the underlying concepts of the SVM for binary classification. Given a two-class labeled training samples set  $\{(x_i, t_i)\}$ ,  $i = 1, 2, \dots, N$ , the goal of SVM is to find a hyperplane:

$$f(x) = w \cdot x + b \tag{1}$$

The weights  $w$  is given by a linear combination of the vectors of the training samples

$$w = \sum_{i=1}^N \alpha_i t_i x_i \tag{2}$$

The parameter  $\alpha_i$  is obtained by solving the following constrained quadratic programming problem:

$$\begin{aligned} \min G(\alpha) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j (x_i \cdot x_j) \\ \text{s.t. } \sum_{i=1}^N t_i \alpha_i &= 0, C \geq \alpha_i \geq 0 \end{aligned}$$

where,  $C$  is the regularization factor:

We can extend the SVM to the nonlinear case by mapping each sample of input space  $R$  into a feature space  $F$  through a nonlinear mapping  $\phi$  and then finding a hyperplane in  $F$ . By using the Eq. 2, it is possible to rewrite Eq. 1 to obtain this hyperplane

$$f(x) = \sum_{i=1}^N \alpha_i t_i \phi(x) \cdot \phi(x_i) + b \tag{3}$$

By use of kernel function of the form

$$K(x, x_i) = \phi(x) \cdot \phi(x_i) \tag{4}$$

The Eq. 3 can be rewritten as:

$$f(x) = \sum_{i=1}^N \alpha_i t_i \cdot K(x, x_i) + b$$

Since the kernel function defined in Eq. 4 changes the mapping from the input to the feature space, it is very important to the SVM.

**SVM for multi-class classification:** SVM were originally designed for binary classification. However, facial expression recognition is a problem of multi-class classification. How to effectively extend SVM for multi-class classification is still an ongoing research issue. Currently there are two types of approaches for multi-class SVM. One is by combining several binary classifiers while the other is by directly considering all training samples into one optimization formulation (Hsu and Lin, 2002).

The one-against-all approach, by combining  $c$  binary classifiers (the parameter  $c$  is the No. of classes), is adopted in this study. The  $i$ -th SVM constructs a hyperplane between class  $i$  and the  $c-1$  other classes. A majority vote across the classifiers or some other measure can then be applied to classify a new sample (Weston and Watkins, 1998). The  $i$ -th SVM is trained with all of the examples in the  $i$ -th class with positive labels and all other examples with negative labels. Thus given  $N$  training samples  $(x_1, t_1), \dots, (x_N, t_N)$ , where  $x_i \in R$  and  $t_i \in \{1, 2, \dots, c\}$  is the class of  $x_i$ , the  $i$ -th SVM solves the following problem (Hsu and Lin, 2002):

$$\begin{aligned} \min & \left[ \frac{1}{2} \|w_i\|_2^2 + C \sum_{j=1}^N \xi_j^i \right], \quad i = 1, 2, \dots, c \\ \text{s.t. } & (w_i)^T \phi(x_j) + b_i \geq 1 - \xi_j^i \quad \text{if } t_j = i \\ & (w_i)^T \phi(x_j) + b_i \leq -1 + \xi_j^i \quad \text{if } t_j \neq i \\ & \xi_j^i \geq 0, \quad j = 1, \dots, N \end{aligned} \tag{5}$$

After solving Eq. 5, we obtain  $c$  decision hyperplane. The new sample  $x$  is in the class which has the largest value of the decision function:

$$\arg \max_{i=1, \dots, c} \sum_{j=1}^N \alpha_j^i K(x, x_j) + b_i$$

### IMPROVED SVM ALGORITHM BY MODIFYING KERNELS

**Theoretical analysis:** A nonlinear SVM maps each sample of input space  $R$  into a feature space  $F$  through a nonlinear mapping  $\varphi$ . The mapping  $\varphi$  defines an embedding of  $S$  into  $F$  as a curved submanifold (Fig. 1). Denote  $\varphi(x)$  the mapped samples of  $x$  in the feature space, a small vector  $dx$  is mapped to:

$$\varphi(dx) = \nabla \varphi \cdot dx = \sum_i \frac{\partial}{\partial x^{(i)}} \varphi(x) dx^{(i)}$$

The squared length of  $\varphi(dx)$  is written as:

$$ds^2 = |\varphi(dx)|^2 = \sum_{i,j} g_{ij}(x) dx^{(i)} dx^{(j)} \quad (6)$$

Where:

$$g_{ij}(x) = \left( \frac{\partial}{\partial x^{(i)}} \varphi(x) \right) \cdot \left( \frac{\partial}{\partial x^{(j)}} \varphi(x) \right) = \frac{\partial}{\partial x^{(i)}} \frac{\partial}{\partial x^{(j)}} K(x, x') \Big|_{x'=x} \quad (7)$$

is the Riemannian metric tensor induced in  $S$ .

In the feature space  $F$ , we can increase the margin (or the distances  $ds$ ) between classes to improve the performance of the SVM. Taking the Eq. 6 into account, this leads us to increase the Riemannian metric tensor  $g_{ij}(x)$  around the boundary and to reduce it around other samples. In view of Eq. 7, we can modify the kernel  $K$  such that  $g_{ij}(x)$  is enlarged around the boundary.

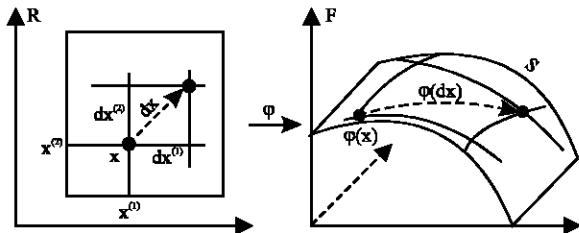


Fig. 1: The mapping  $\varphi$  define an embedding of  $S$  into  $F$  as a curved submanifold

**Modifying kernel based on the structure of the Riemannian geometry:** Assume the kernel can be modified as:

$$\tilde{K}(x, x_i) = p(x)p(x_i)K(x, x_i) \quad (8)$$

Specifically, assume the kernel functions used in SVM is Gaussian kernel, that is:

$$K(x, x_i) = \exp(-\|x - x_i\|^2 / \sigma^2) \quad (9)$$

where, the parameter  $\sigma$  is kernel width. It is proved that the corresponding Riemannian metric tensor is (Amari, 1999).

$$g_{ij}(x) = \frac{1}{\sigma^2} \delta_{ij}$$

After modifying the kernel, the Riemannian metric tensor is changed into:

$$\tilde{g}_{ij}(x) = p_i(x)p_j(x) + p^2(x)g_{ij}(x)$$

To ensure that  $p(x)$  have large value around the Support Vector (SV), it can be constructed in a sample dependent way as:

$$p(x) = \sum_{i \in SV} \alpha_i \exp(-\|x - x_i\|^2 / 2\tau^2) \quad (10)$$

where,  $\tau$  is a free parameter and summation runs over all the support vectors.

### EXPERIMENTS

Here, a problem for facial expression recognition based on the JAFFE database is tested. The database contains 10 Japanese young female with 7 basic facial expressions (Fig. 2). Each expression has 3 images, thus there are 210 facial expression images in the database. The size of each image is  $256 \times 256$ . We select 70 images as the training set and the total 210 images are used as the testing set. This implies that there are 30 test images for each expression.

Some expression image preprocessing has been done before the recognition. These preprocessing includes the sub-regions segmentation, sub-sampling and the normalization of the facial expression images. Feature extraction is an important step toward recognition. In this study, the feature extraction is based on the PCA. After



Fig. 2: Examples of seven facial expression images from JAFFE database

Table 1: Classification Accuracy Ratio (CAR) with SVM on JAFFE database

Expression	Before modifying (%)	After modifying (%)
Angry	93.3	100.0
Disgust	93.3	100.0
Fear	90.0	86.7
Happy	96.7	90.0
Neutral	100.0	100.0
Sad	93.3	93.3
Surprise	96.7	96.6
Avg. CAR	94.8	95.7

Table 2: Comparing the training results for two different values of  $\tau$  after modifying kernel

Expression	$\tau = 0.18$ (%)	$\tau = 2.98$ (%)
Angry	100.0	86.6
Disgust	100.0	90.0
Fear	86.7	83.3
Happy	90.0	96.6
Neutral	100.0	90.0
Sad	93.3	83.3
Surprise	96.6	100.0
Avg. CAR	95.7	90.0

the feature extraction, the samples are represented by the  $d$  dimensional feature vectors. Selecting an appropriate  $d$  is very critical for the recognition. The value of  $d$  is to be set for 28 in the experiments based on a 3-fold cross-validation.

In the experiment, the Gaussian kernel is modified based on the Eq. 8. We set the kernel width  $\sigma$  to be equal to the optimal one,  $\sigma = 1.0$ . According to the analysis of Amari (1999) the optimal value for the parameter  $\tau$  is around  $\sigma/\sqrt{d}$ . Therefore, we set the value of  $\tau$  to be equal to equal to 0.18. The regularization factor of SVM is set for  $C = 10^5$ . From Table 1, we see that the Classification Accuracy Rate (CAR) is remarkably improved after modifying the kernel.

To show the importance of selecting an appropriate value of  $\tau$ , simulation results for two different values of  $\tau$  are compared in Table 2. It shows that when  $\tau$  takes an appropriate value (0.18 in this example), the performance (measured by CAR) is improved. However, the performance is bad when  $\tau$  is too large (as shown in this example) or small.

## CONCLUSIONS

Motivated by the study of Amari (1999), this study proposes a novel facial expression recognition approach based on improved support vector machine by modifying Gaussian kernels. Experiments on Japanese female facial expressions database show that the Classification Accuracy Rate (CAR) is remarkably improved after modifying the kernel. Experiments also show that the importance of selecting an appropriate parameter when modifying the kernel. In the future study we will extend the Gaussian kernel to other different kernel cases.

## ACKNOWLEDGMENT

The authors would like to express their cordial thanks to Dr. Pan Guofeng and Li Wen for their valuable advice.

## REFERENCES

- Amari, S., 1999. Improving support vector machine classifiers by modifying kernel functions. *Neural Networks*, 12: 783-792.
- Belhumeur, P.N., J.P. Hespanha and D.J. Kriegman, 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19: 711-720.
- Campbell, C., 2002. Kernel methods: A survey of current techniques. *Neurocomputing*, 48: 63-84.
- David, V. and A. Sanchez, 2003. Advanced support vector machines and kernel methods. *Neurocomputing*, 55: 5-20.
- Ekman, P. and W.V. Friesen, 1971. Constants across cultures in the face and emotion. *Personality Soc. Psychol.*, 17: 124-129.
- Fasel, B. and J. Luetin, 2003. Automatic facial expression analysis. *Pattern Recog.*, 36: 259-275.
- Hsu, C.W. and C.J. Lin, 2002. A comparison of methods for multi-class support vector machines. *IEEE Trans. Neural Networks*, 13: 415-425.

- Mase, K. and A. Pentland, 1991. Recognition of facial expression from optical flow. *IEICE Trans.*, 74: 3474-3483.
- Otsuka, T. and J. Ohya, 1998. Spotting segments displaying facial expression from image sequences using HMM. *IEEE Proceeding of the 2nd International Conference on Automatic Face and Gesture Recognition*, April 14-16, Nara, Japan, pp: 442-447.
- Perez-Cruz, F. and O. Bousquet, 2004. Kernel methods and their potential use in signal processing. *IEEE Signal Process. Mag.*, 21: 57-65.
- Suwa, M., N. Sugie and K. Fujimora, 1978. A preliminary note on pattern recognition of human emotional expression. *Proceeding of the 4th International Joint Conference on Pattern Recognition, (IJCPR'78)*, Kyoto, Japan, pp: 408-410.
- Turk, M.A. and A.P. Pentland, 1991. Face recognition using eigenfaces. *IEEE Proceeding of the Conference on Computer Vision and Pattern Recognition*, June 3-6, Maui, HI, USA., pp: 586-591.
- Weston, J. and C. Watkins, 1998. *Multi-Class Support Vector Machines*, CSD-TR-98-04, Royal Holloway. 1st Edn., Department of Computer Science, University of London, London.
- Zilu, Y., L. Jingwen and Zh. Youwei, 2006. Facial expression recognition based on classifier Combinations. *IEEE Proceeding of 8th International Conference Signal Processing (ICSP)*, Nov. 16-20, Guilin, China, pp: 367-372.