Pattern Recognition 42 (2009) 437 – 444

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

View independent face detection based on horizontal rectangular features and accuracy improvement using combination kernel of various sizes

Kazuhiro Hotta*

The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan

ARTICLE INFO

Article history:
Received 9 August 2007
Received in revised form 6 August 2008
Accepted 10 August 2008

Keywords:
Combination kernel
Summation kernel
Local kernel
View independence
Horizontal rectangular features
Face detection
Support vector machine

ABSTRACT

This paper proposes a view independent face detection method based on horizontal rectangular features, and accuracy improvement by combining kernels of various sizes. Since the view changes of faces induce large variation in appearance in the horizontal direction, local kernels are applied to horizontal rectangular regions to model such appearance changes. Local kernels are integrated by summation, and then used as a summation kernel for support vector machine (SVM). View independence is shown to be realized by the integration of local horizontal rectangular kernels. However, in general, local kernels (features) of various sizes have different similarity measures, such as detailed and rough similarity, and thus their error patterns are different. If the local and global kernels are combined well, the generalization ability is improved. This research demonstrates the comparative effectiveness of combining the global kernel and local kernels of various sizes as a summation kernel for SVM against use of only the global kernel, only the combination of local kernels and Adaboost with SVMs with a kind of local kernel.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

The realization of view independent face detection continues to present challenges [1–3] since the view of faces is changed dynamically in real environments and applications. These view changes induce large variations in appearance, making it difficult to create a method robust enough to view changes. In particular, the appearance changes of faces in the horizontal direction are much larger than those in the vertical direction, and thus the modeling of horizontal features becomes an important issue in achieving view independent face detection. In addition, view changes of the face induce large non-linear variations in feature space; an eigenspace under view changes is shown in Fig. 1, where each line shows the view changes of a subject. Therefore, any robust method must also cope with non-linear variations induced.

Kernel-based methods can represent non-linear variations easily, and therefore such methods have been used to cope with view changes [4,5]. In particular, support vector machine (SVM) is typically used in object detection tasks since they are binary-classification problems. In general, a kernel function in SVM is applied to all features extracted from a sample; however, the appearance changes in the horizontal direction are not modeled well by this strategy. In recent years, SVM with local kernels has been proposed in order to use local features effectively [6–8]. This method is useful for modeling both the appearance changes in the horizontal direction and non-linear variation. To model the appearance changes in the horizontal direction, a face region is divided into horizontal rectangular regions and local kernels are applied to these regions. Each local kernel measures the similarity between certain horizontal regions of samples, like a montage picture. A summation kernel is then used to combine the local kernels [6].

However, the combination of local horizontal rectangular features of the same size is not sufficient to achieve the high generalization ability shown by humans. Various similarity measures are required to improve the accuracy further. For example, a local kernel measures detailed similarity whereas the global kernel measures rough similarity, and as such the error patterns are different in local features and global features. If kernels of various feature sizes are combined well, a good detector using various similarities will be developed [9]. To combine the kernels of various feature sizes, a summation of them is used. It is known that the summation of classifiers gives good performance [10,11]. Of course, the summation of kernels of various feature sizes satisfies Mercer's theorem [12,13]. Since a normalized polynomial kernel whose output is between 0 and 1 is used as a kernel function, the output of each kernel is combined fairly.

To evaluate the proposed method based on horizontal rectangular kernels, a large number of face images of various views and non-face images are gathered from databases and the WWW [14–17], and the generalization ability is evaluated by using the receiver

* Tel.: +81 424 43 5232; fax: +81 424 43 5291.
E-mail address: hotta@ice.uec.ac.jp
URL: http://www.htlab.ice.uec.ac.jp/~hotta/.

0031-3203/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.08.013

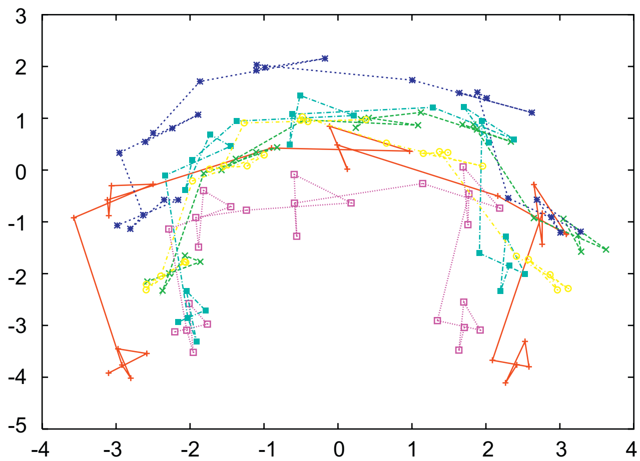

Fig. 1. Eigenspace under view changes. The horizontal and vertical axes show the first and second principal component axes. This space is constructed from 125 face images of five subjects (125 images = 5 images × 5 views × 5 subjects). Each line represents view changes of an individual subject.

operating characteristic (ROC) curve. First, a comparison of the global kernel with the summation of only local horizontal rectangular kernels reveals that the combination of horizontal rectangular features outperforms the global kernel. Likewise, the combination of horizontal rectangular kernels is shown to be more effective than that of local vertical rectangular kernels. Next, local horizontal rectangular kernels of various sizes and the global kernel are combined by summation in order to use both detailed and rough information simultaneously. It is confirmed that such combination improves the generalization ability. Finally, a comparison of the proposed combination kernel with two approaches based on Adaboost [18] shows that SVM with the proposed combination kernel outperforms the Adaboost with SVMs with a kind of local kernel.

Section 2 explains a SVM with summation kernel of rectangular features for view independent face detection. Section 3 demonstrates the effectiveness of the proposed method by comparison with SVM of a global kernel and Adaboost with SVMs. Conclusions and future work are described in Section 4.

2. Proposed method

In this paper, local appearance features are used in order to utilize local kernels effectively. For this purpose, Gabor features are used since it is known that they are robust to illumination changes and are effective for object detection and recognition [19–21]. Section 2.1 describes the properties of Gabor filters and Section 2.2 describes SVM based on the combination of local and global kernels.

2.1. Gabor filter

In the mammalian visual cortex, there are many neurons that are characterized as localized and orientation selective. It is known that Gabor filters are well fitted to the receptive field profiles of the simple cells of the cat visual cortex [22].

The outputs of Gabor filter are regarded as sparse coding, because Gabor-like receptive fields are obtained by using the constraint which maximizes the sparseness of the response to natural images [23]. It is also reported that Gabor-like filters are obtained by independent components analysis of natural images [24].

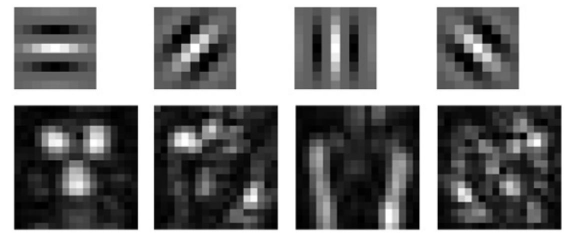

Fig. 2. Gabor filters and Gabor features.

Gabor filters are defined by

$$\psi_{\boldsymbol{k}}(\boldsymbol{x}) = \frac{\boldsymbol{k}^2}{\sigma^2} \exp\left(\frac{-\boldsymbol{k}^2\boldsymbol{x}^2}{2\sigma^2}\right)[\exp(i\boldsymbol{k}\boldsymbol{x}) - \exp(-\sigma^2/2)], \tag{1}$$

where $\boldsymbol{x}=(x,y)^{\mathrm{T}}$, $\boldsymbol{k}=k_\nu \exp(i\phi)$, $k_\nu = k_{max}/f^\nu$, $\phi=\mu\cdot\pi/4$, and $f=\sqrt{2}$. Gabor features of only one frequency level have been found to give good detection performance [19]. Therefore, in the following experiments, Gabor filters of four different orientations ($\mu=0,1,2,3$) with one frequency level ($\nu=0$) are used in order to speed up recognition.

The summation of values in a Gabor filter with upper parameters is nearly 0 when the size of the filter is more than 9×9 pixels. Accordingly, the size is set to 9×9 pixels here. Since the summation of values in the filter is nearly 0, Gabor output at the non-textured region is nearly 0. Fig. 2 shows the Gabor filters of four different orientations and the Gabor features of a frontal face image. In this paper, the norm of the Gabor outputs of the real and imaginary parts at each position is used as Gabor features. Gabor features in Fig. 2 are also the norm of real and imaginary parts. Gabor outputs at many positions of face images are small and only specific positions give large values. This indicates the sparseness of Gabor features.

2.2. SVM based on combination of local and global kernels

To serve as a brief introduction to SVM [12,25], it determines the optimal hyperplane which maximizes the margin, where the margin is the distance between the hyperplane and the nearest sample from it. When the training set (sample and its label) is denoted as $S=((\boldsymbol{x}_i,y_i),\ldots,(\boldsymbol{x}_L,y_L))$, the optimal hyperplane is defined by

$$f(\boldsymbol{x}) = \sum_{i\in SV} \alpha_i y_i \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x} + b, \tag{2}$$

where SV is a set of support vectors, b is the threshold, and α is the solutions of the quadratic programming problem. The training samples with non-zero α are called support vectors.

The upper explanation assumes the linearly separable case. In the linearly non-separable case, the non-linear transform $\Phi(\boldsymbol{x})$ can be used. The training samples are mapped into high dimensional space by $\Phi(\boldsymbol{x})$. By maximizing the margin in high dimensional space, non-linear classification can be carried out. If the inner product $\Phi(\boldsymbol{x})^{\mathrm{T}}\Phi(\boldsymbol{y})$ in high dimensional space is computed by kernel $K(\boldsymbol{x},\boldsymbol{y})$, then training and classification can be done without mapping into high dimensional space. The optimal hyperplane using the kernel is defined by

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{i\in SV} \alpha_i y_i \Phi(\boldsymbol{x}_i)^{\mathrm{T}}\Phi(\boldsymbol{x}) + b \\ &= \sum_{i\in SV} \alpha_i y_i K(\boldsymbol{x}_i,\boldsymbol{x}) + b. \end{aligned} \tag{3}$$

Mercer's theorem gives whether $K(\boldsymbol{x},\boldsymbol{y})$ is the inner product in high dimensional space. The necessary and sufficient conditions are symmetry $K(\boldsymbol{x},\boldsymbol{y})=K(\boldsymbol{y},\boldsymbol{x})$ and positive semi-definiteness of the kernel matrix $\boldsymbol{K}=(K(\boldsymbol{x}_i,\boldsymbol{x}_j))_{i,j=1}^{L}$. If $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\beta}\geqslant 0$ where $\beta\in\Re$ is satisfied, $\boldsymbol{K}$

Download English Version:

https://daneshyari.com/en/article/533692

Download Persian Version:

https://daneshyari.com/article/533692

Daneshyari.com