

Infrared dynamic hand gesture recognition based on Gabor feature and sparse representation

Lei Shang
School of Information Science
and Engineering
Shandong University

Zhi Liu
School of Information Science
and Engineering
Shandong University
liuzhi@sdu.edu.cn

Haixia Zhang
School of Information Science
and Engineering
Shandong University
haixia.zhang@sdu.edu.cn

ABSTRACT

Dynamic hand gesture recognition is a subject which has been investigated almost from the beginning of using terminals to interact with the computer central unit. We present a method for dynamic hand gesture recognition with image source acquired by a single IR camera. First of all, the hand images are captured by one infrared camera, which are light independent, and not be limited by skin color. Second, due to the dynamic hand gesture sets contain more than one frame, the data dimension is very large and the adjacent frames have limited difference. We use the twofold selections to choose the key frames from the gesture image sets. The Gabor feature is used to describe the locality variation. And the Sparse representation based classification (SRC) is used for recognition. Experimental results on dynamic hand gesture images with variations of rotation and translation demonstrate the good performance of our method.

Keywords

Dynamic hand gesture recognition, infrared image, twofold selection, Gabor feature, Sparse representation classification

1. INTRODUCTION

Human computer interaction is one of the most visible and challenging research topics in computer vision and machine learning [1], [2]. Hand gestures as the most common used communication mode and hand recognition technology have become the focus of study in the field of computer vision and pattern recognition. In passed decades, many research methods are proposed for recognizing various gestures [3].

For splitting up the hand from the image, the most common is using skin color [2],[4]. Compared with the traditional image[5] capture machines such as color camera, instrumented gloves and Kinect sensors, the infrared imaging can acquire the intrinsic temperature information of skin, which is not sensitive to the illumination conditions and disguises. In this paper we use infrared imaging, which makes the hand seg-

mentation more robust than based on skin color information method segment from the background.

In hand recognition section, different methods are proposed such as texture information, skin color, contour information [2] and shape parameters [3].Based on infrared images, exploring a robust feature extraction method is very important. Due to its characteristic of extracting the image within a specific area of multi-scale, multi-direction spatial frequency, Gabor filters [6] can achieve good results in this field. Therefore, Gabor features for recognition have been widely used, such as Gabor feature-based Fisher discriminate [7] and so on. However, the classical Eigen and Fisher algorithms consider only the global scatter of training samples. As we known, Gabor feature image is formed by convolution of hand gesture image with dozens of Gabor filters, whose dimension of feature-space is very large.

Although dynamic hand gesture images have a high dimensionality, they usually lie on a lower dimensional subspace or sub-manifold. The success of manifold learning implies that the high dimensional hand images can be sparsely represented or coded by the representative samples on the manifold. Recently, an interesting work was reported by Wright et al [8], where the sparse representation (SR) technique is employed for robust face recognition. In [8], using the training data as the dictionary book to find the sparse linear combination representation for the testing image via ℓ_1 -norm. The SR based classification (SRC) is defined as evaluating which class of training samples could result in the minimal reconstruction error of the input testing image with the sparse coding coefficients. The dimensionality of the training and testing samples should be reduced beforehand to make the ℓ_1 -norm sparse coding computationally feasible.

In this paper, we use Gabor feature based sparse representation classification (GSRC) [9] to realize dynamic hand gesture recognition.

2. RELATED WORK

2.1 Gabor wavelets

The Gabor wavelets whose kernels are similar to the 2D cortical simple cells, exhibit desirable characteristic of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains. The Gabor wavelets with orientation μ and scale ν are defined as:

$$\Psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}} [e^{ik_{\mu,\nu}z} - e^{-\frac{\sigma^2}{2}}] \quad (1)$$

where, $z = (x, y)$ denotes the pixel, $\|\cdot\|$ denotes the norm operator, and the wave vector $k_{\mu,\nu}$ is defined as follows:

$$k_{\mu,\nu} = k_{\nu} e^{i\phi_{\mu}} \quad (2)$$

$k_{\nu} = \frac{k_{max}}{f\nu}$ and $\phi_{\mu} = \frac{\pi\mu}{8}$ is the maximum frequency and f is the spacing factor between kernels in the frequency domain. In addition, the σ determines the ratio of the Gaussian window width to wavelength.

The Gabor feature representation of an image is the convolution of the image with the Gabor kernel. Let $i(z)$ indicates the gray-level image, and the Gabor kernel outputs

$$G_{\mu,\nu}(z) = i(z) * \Psi_{\mu,\nu}(z) \quad (3)$$

where $*$ denotes the convolution operator, and $\Psi_{\mu,\nu}(x, y)$ is the Gabor kernel. The Gabor filtering coefficient $G_{\mu,\nu}(z)$ is a complex number, which can be written as $G_{\mu,\nu} = M_{\mu,\nu}(z) e^{i\theta_{\mu,\nu}(z)}$ with $M_{\mu,\nu}$ being the magnitude and $\theta_{\mu,\nu}(z)$ being the phase. It known that magnitude information contains the variation and local energy in the image. In [8], the Gabor feature vector χ is defined as via uniform down-sampling, normalization and concatenation of the Gabor filtering coefficients:

$$\chi = (a_{0,0}^{(\rho)t}, a_{0,1}^{(\rho)t}, \dots, a_{4,7}^{(\rho)t}) \quad (4)$$

t is the transpose operator, $a_{\mu,\nu}^{(\rho)}$ is the concatenated column vector from down-sampled magnitude matrix $M_{\mu,\nu}(z)$ by a factor of ρ .

2.2 Sparse representation based classification

Denote by $A_i = [s_{i,1}, s_{i,2}, \dots, s_{i,n_i}] \in R^{m \times n_i}$ the set of training samples of the i^{th} object class, where $s_{i,j}, j = 1, 2, \dots, n$ is an m -dimensional vector stretched by the j^{th} sample of the i^{th} class.

For a test sample $y_0 \in R^m$ from this class, intuitively, y_0 could be well approximated by the linear combination of the samples within A_i , i.e.

$$y_0 = \sum_{j=1}^{n_i} \alpha_{i,j} s_{i,j} = A_i \alpha_i \quad (5)$$

where $\alpha_i = [\alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n_i}]^T$ are the coefficients. Suppose we have K object classes, and let $A = [A_1, A_2, \dots, A_K]$ be the concatenation of the n training samples from all the K classes, where $n = n_1 + n_2 + \dots + n_K$, then the linear representation of y_0 can be written in terms of all training samples as

$$y_0 = A\alpha \quad (6)$$

$$\alpha = [\alpha_1, \dots, \alpha_i, \dots, \alpha_K] = [0, \dots, 0, \alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n_i}, 0, \dots, 0]^T$$

2.3 Gabor-feature based SRC

The augmented Gabor hand feature vector χ , which is a local feature descriptor, can not only enhance the hand feature but also tolerate to image local deformation to some extent. So we use the Gabor feature instead of the holistic features in the SRC framework, and the Gabor-feature based SR is

$$\begin{aligned} \chi(y_0) &= X(A_1)\alpha_1 + X(A_2)\alpha_2 + \dots + X(A_K)\alpha_K \\ &= X(A)\alpha \end{aligned} \quad (7)$$

$X(A) = [X(A_1)X(A_2) \dots X(A_n)]$ and $X(A_i) = [\chi(s_{i,1}), \dots, \chi(s_{i,n_i})]$. With replacing y_0 and A and $\chi(y_0)$ and $X(A)$, the Gabor-feature based SRC (GSRC) can be achieved.

3. DYNAMIC HAND GESTURE RECOGNITION

The process of dynamic hand gesture recognition is shown as following

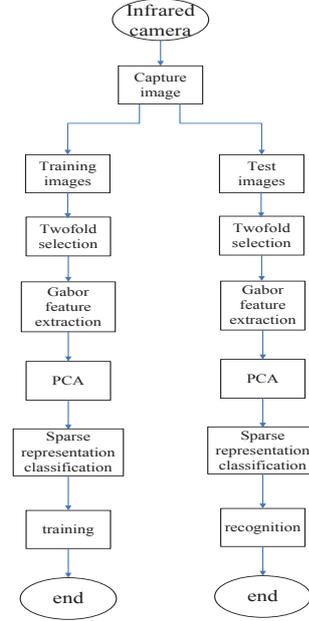


Figure 1: The flowchart of our dynamic hand gesture recognition.

3.1 hand segmentation

The capture image tool is infrared camera, compared with gray and color hand gesture imaging, infrared imaging can acquire the intrinsic temperature information of the skin, which is robust to the impacts of illumination and disguises.

The first step is hand segmentation. The raw image changes into gray image. After threshold processing the gray image is turned to binary image. For every hand gesture may in different illumination, the threshold is set adaptively. The threshold value depends on the mean gray value of the first image of the dynamic hand image sets. After the graying and smoothing the original frames, these binary difference-frames are created by computing differences between subsequent frames. In order to reduce the small change frames, a certain threshold is necessary to determine whether difference-frames are important enough to use. Both the first frame and the other $K - 1$ frame which corresponding to the $K - 1$ largest binary difference-frame are selected as the K key frame.

There are twofold selections during the image set processing. The first threshold is used to determine if the pixel belong to hand region or background. The dynamic hand gesture sequence is a delicate procedure, which resulted in videos containing only the gesture itself, but not being aligned in time. For instance, one video of the gesture '1' could have the finger moving at frame 10, whereas another video had the finger moving at frame 15. So the second selection is wiping the small variation frames.



Figure 2: The original image set of gesture '1'.



Figure 3: 15 key frames of image set showed in Figure.2.

3.2 Gabor feature extraction

In this paper, Gabor filters with 5 different scales and 8 different directions are used, so the total number of the filters is 40.

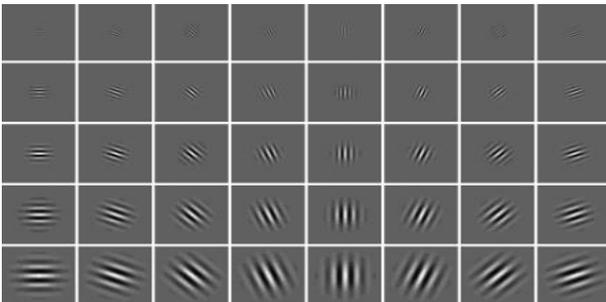


Figure 4: Gabor filter in 5 scales and 8 directions

In our experiment, 15 key frames are chosen from the hand gesture image set. We translate the first frame of the key frame set as the first column, in chronological order; the image set is changed in to a matrix.

3.3 Gabor feature based SRC

The hand gesture matrix get through the Gabor filters, we get the Gabor feature matrix. The dimension of the augmented Gabor feature is very high. Principal Component Analysis(PCA)[10] is then applied to reduce their dimension for classification in SRC.

4. EXPERIMENTAL RESULTS

The infrared database used in this paper was captured by an infrared camera (Point Grey FFMV-03M2C-CS) with a

black board as background. The infrared camera with resolution of 1024x1280 recording at a speed of 30 frames per gesture was used. The training database contained 8 different gesture was created, which had different horizontal rotating movement, wrist rotation, and delicate illumination. Illustrations of each of these hand gestures are shown in Figure 5.

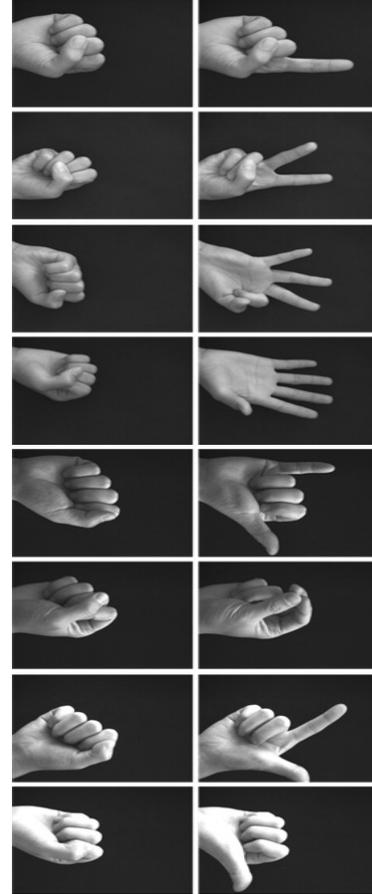


Figure 5: Two frames of the gestures in descending order:1,2,3,5,6,7,8,and thumb.

Five different persons were asked to perform fourteen times of the gestures in Figure 5 presented each, to ensure the GSRC method is largely invariant to different of shapes of hands. Each person was shown the gesture and asked to perform this gesture in front of the infrared camera. Each person imitated one gesture for 14 times with different horizontal rotating movement and wrist rotation. In Figure 6, we can see some gestures with different horizontal rotating movement and wrist rotation. The horizontal rotating angle changed from -45° to 45° , and and the wrist angle approximately changed from -30° to 30° . All experiments were finished on an pc with Intel core i5-4570 , 3.20Ghz CPU and 8 GB of RAM. Software is Win7 opening system, MATLAB R2012b.

Altogether, the number of sample is $5person * 8gestures * 14attempts = 560$. By applying a ratio of 50% for training set and 50% for the testing set. After the pre-processing, normalization is performed for the frames of the hand ges-

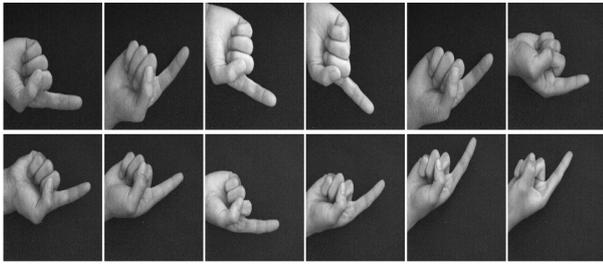


Figure 6: Some samples with different horizontal movement and rotation.

Table 1: the recognition of key frames and original image set without twofold selection (feature dimensionality is 80)

	Original image set with just first selection	Key frame
Recognition	90.71%	93.75%
Consume time	1.2036s	0.8731s

ture images set. The image size is changed from $1024 * 1280$ to $64 * 80$. After twofold selection of the image set, 15 key frames are chosen from the original gesture image set. PCA is then applied to reduce the dimension of the outputs of Gabor filters. Different dimensions are chosen for the feature extraction. Figure 7 is the recognition rate by SRC versus different PCA dimension. Here the consume time means the feature extraction and recognition time.

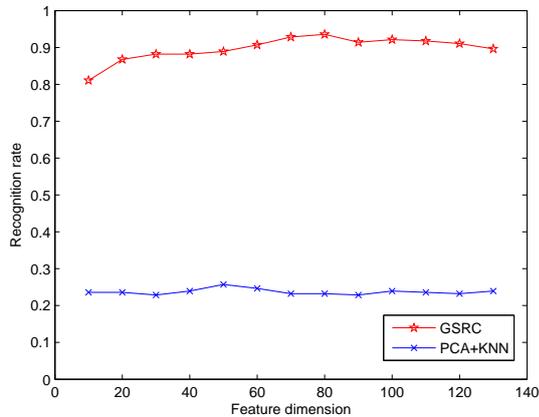


Figure 7: Recognition rate by GSRC vs PCA+KNN versus different feature dimension.

Here, the $\sigma = 0.001$. Compare with PCA dimensionality reduction and k-Nearest Neighbor.

5. CONCLUSIONS

In this paper, Gabor feature based sparse representation classification method is proposed for hand gesture recognition. All hand gestures images in the evaluation database are captured by an infrared camera. The experimental results demonstrate that our method can achieve good performance. The correct recognition rate is about 93.75%.

6. REFERENCES

- [1] Chengjun Liu. Gabor-based kernel pca with fractional power polynomial models for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):572–581, 2004.
- [2] Yu Ren and Fengming Zhang. Hand gesture recognition based on meb-svm. In *International Conference on Embedded Software and Systems, 2009. ICESS'09*, pages 344–349. IEEE, 2009.
- [3] Meenakshi Panwar. Hand gesture recognition based on shape parameters. In *2012 International Conference on Computing, Communication and Applications (ICCCA)*, pages 1–6. IEEE, 2012.
- [4] Hsiang Yueh Lai and Han Jheng Lai. Real-time dynamic hand gesture recognition. In *2014 International Symposium on Computer, Consumer and Control (IS3C)*, pages 658–661. IEEE, 2014.
- [5] Zhihua Xie. Infrared face recognition based on lbp co-occurrence matrix. In *Control Conference (CCC), 2014 33rd Chinese*, pages 4817–4820. IEEE, 2014.
- [6] Yuanbo Yang and Jinguang Sun. Face recognition based on gabor feature extraction and fractal coding. In *2010 Third International Symposium on Electronic Commerce and Security (ISECS)*, pages 302–306. IEEE, 2010.
- [7] Chengjun Liu and Harry Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *11(4):467–476*, 2002.
- [8] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma. Robust face recognition via sparse representation. *31(2):210–227*, 2009.
- [9] Meng Yang and Lei Zhang. Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. In *Computer Vision—ECCV 2010*, pages 448–461. Springer, 2010.
- [10] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.