Hand Posture Recognition via Sparse Representation

Chuqing Cao∗ and Ying Sun†

∗ Harbin Institute of Technology, 150001 Harbin, P.R.China
E-mail: caochuqing@gmail.com Tel: +86-451-86418285
† National University of Singapore, 117576 Singapore
E-mail: elesuny@nus.edu.sg Tel: +65-65167839

Abstract—We propose a new method for hand gesture recognition via sparse representation. Initially, we present the region of the hand is detected based on skin color segmentation in the YCbCr color space and image normalization. Then, the recognition of hand posture is casted as the sparse representation of a test image with a set of the database images. The $l_1$-minimization is applied to accurately and efficiently calculate the sparse representation so as to classify different postures under a variety of conditions. Experimental results demonstrate the effective and robust performance of the proposed method.

I. INTRODUCTION

Hand posture recognition is one of the most active research areas in computer vision, mainly for the purpose of sign language recognition and Human-computer Interaction (HCI). Hand based interface is the most effective, general-purpose interaction tool. Due to its dexterous functionality in communication and manipulation, it can provide natural and friendly interface between computer and untrained users[1]. There are a lot of important applications, and the field has many applications for how we may build advanced systems, such as robots [2] and editing system [3].

Vision-based hand posture recognition is a complex problem that has to be solved in many different ways. Contour and shape are the commonly used features in hand posture recognition. Stergioupolou and Papamarkos [4] described shape of different hand postures by neural network. Rokade et al. [5] proposed a method for hand gesture recognition which is based on thinning of segmented image. Chang [6] used CSS features for hand posture recognition. In practice, it is rather hard to extract continuous contours and integrated shape of the hand from the image with complex background. Some researches focus on extracting more features. A posture-based human-robot interaction system based on elastic graph matching was proposed in previous studies [7]. Yin and Xie [8] presented a posture recognition system by finger identification. According to the method proposed by Flasinski and Myslinski [9], Polish sign language recognition is performed based on graph description. A real-time hand gesture recognition system was built based on Haar wavelet representation [10].

This paper presents a novel and robust approach which recognizes a set of 10 hand postures via sparse representation. Skin-color segmentation is used to pick up hand regions from the images. Then, hand images need to be normalized. After image process, the sparse representation algorithm is introduced to classify the different hand postures. The result is quite stable even if hand with rotation and partial occlusion.

II. IMAGE PROCESSING

Due to hand posture images are always with different background, simple and fast processing of the input image is necessary for hand posture recognition. Image processing is applied to both test images and database images. The framework of image processing is shown in Fig.2.

A. Hand Region Detection

The first step image processing is the detection and extraction of the hand region from the background. The proposed hand region detection technique is applied in the YCbCr color space. Equation(1) is used to transform RGB to YCbCr color space:

\[
\begin{bmatrix}
Y \\
Cb \\
Cr
\end{bmatrix} = \begin{bmatrix}
16 & 128 \\
128 & 128 \\
128 & 128
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix} + \begin{bmatrix}
65.48 & 128.55 & 24.97 \\
-37.8 & -74.2 & 112 \\
112 & -93.79 & -18.22
\end{bmatrix}
\]

From training of 50 images, Chai and Ngan [11] created a map of the chrominance components of skin color. It was found that the ranges of Cb and Cr values, which were narrowly and consistently distributed, are $R_{cb} = [80,105]$ and $R_{cr} = [130,165]$, respectively. Hand region image is translated from RGB into YCbCr color space and picked out the pixels belong to the skin color region. The advantage of this approach
Hand region detection

Resize the hand region

Reject the background

Get the center point

Adjust the position of hand region

Down-sample filter

Gray-level transform

Normalized images

Input images

Fig. 2. The framework of image processing.

is invariant to rotation and scaling as well as morphologic variations of the hand.

B. Image Normalization

Hand regions \((h_i)\) are the different sizes in input images \((I_i)\). It is necessary to resize hand regions to the size \(200 \times 165\). The hand region is converted to binary image, where pixels are only black and white. The pixel value of hand region is 1 and background area is 0. Then the hand region is eroded to obtain the center point \((C_0)\) of the palm. Hand region was moved to guarantee that center point \((C_0)\) matches the reference point \((R_0)\), as shown in Fig.3. Furthermore, we down-sample the new images from the size \(200 \times 165\) pixels to the size \(12 \times 10\) pixels and transform them to gray level images.

III. RECOGNITION BASED ON SPARSE REPRESENTATION

The main problem of hand posture recognition is using labeled database images from \(k\) different hand postures to correctly determine the class to which test images belongs. If the test posture image indeed belongs to one of the class in the database, this linear combination will only involve the images of that class, so the representation is naturally sparse. The recognition problem is cast as how to represent the test image by the database images of each class.

The hand posture images from \(k = 10\) different postures that have been properly normalized. Each image is of the size \(w \times h\) and can be viewed as a point in the space \(R^m\) with \(m = w \times h\).

A. Recognition as a Sparse Linear Combination

All the normalized images of \(k = 10\) classes in the database are collected as column vectors of one matrix \(A\):

\[
A = [A_1, A_2 \ldots A_s].
\] (2)

A normalized test hand image, stacked as a vector \(y\), can be represented as a sparse linear combination \(Ax\) of the hand images in the database:

\[
y = Ax.
\] (3)

Real images are noisy, so it is impossible to accuracy represent the test image by the images in the database. To address this problem, an error tolerance with quadratic constraint \(\|e\|_2 < \epsilon\) is included:

\[
y = Ax + e.
\] (4)

The space dimension of the image is \(d = 12 \times 10 = 120\) and the number of images in the database is \(s = 1800\), so Equation(4) is under-determined \((d \ll s)\) and its solution is not unique. In general minimum \(l_0\)-norm is selected to address this.

\[
\min \|x\|_0 \quad \text{subject to} \quad \|y - Ax\|_2 \leq \epsilon.
\] (5)

\(\| \cdot \|_0\) denotes the \(l_0\)-norm, which simply counts the number of nonzero elements in the vector. As stated in [12], the minimum \(l_0\)-norm solution of (5) is NP-hard and even difficult to approximate by polynomial-time algorithms. In comparison
with sparse representation approach, it is obvious that exhaust-
ing all subsets of the entries for $x$ is less efficient.

**B. Sparse Solution via $l_1$-Minimization**

Recent research of sparse representation and compressed sensing [13] shows that if the coefficient $x_0$ is sparse enough, the solution of the $l_0$-minimization problem (5) is equal to the following $l_1$-minimization problem:

$$\min ||x||_1 \text{ subject to } ||y - Ax||_2 \leq \epsilon. \quad (6)$$

$|| \cdot ||_1$ denotes the $l_1$-norm. This problem can be solved in polynomial time by standard linear programming or quadratic programming methods [14] when the solution is known to be very sparse. For test image that can be represented by few images of one class in the database, its coefficient vector $x$ is sparse enough.

**C. Classification from Coefficients**

In addition to normalizing the database images, the test images also need to be normalized. Without the normalization, the algorithm may fail to find sparse linear combination of test images. The complete classification procedure is summarized in Algorithm 1.

```plaintext
Algorithm 1: Classification for hand recognition

Input: Normalized database images $A_1, A_2, ..., A_k \in \mathbb{R}^{m \times n}$ for $k$ classes, a test image $y \in \mathbb{R}^m$, an error tolerance $\epsilon$ and threshold $\theta$.

Solve the $l_1$-minimization problem:

$$\min ||x||_1 \text{ subject to } ||y - Ax||_2 \leq \epsilon. \quad (7)$$

if $x_i < \theta$ then
| Set $x_i = 0$;
end

Compute coefficients $\delta(k) = \text{sum}(x_i)$ for each class $k$.

Output: identity $(k) = \max \delta(k)$.
```

Since valid images have a sparse representation in terms of the larger coefficients, we can reject the invalid coefficients which are not above the threshold $\theta$. It is advisable to set the threshold to be 0.2 and the error tolerance $\epsilon$ to be 0.05 through experiments. High coefficient means test image is able to be well represented by the images of this class, so the test image can be classified.

**IV. EXPERIMENT RESULT**

To make sure the test images can be represented by images in the database, more hand posture images should be selected to build the database, including images with different rotation of the posture. We select 15 different images of each posture from male and female with the right hand. From the experiment, the hand posture is well represented by the same posture image with 5 degree rotation. In the database, each posture image was rotated from 60 to 120 degree with a 5 degree interval, as showed in Fig.4. The image number of each posture is 180 and over 1800 images are used to build the database. Running on a 2.8GHz CPU with 4Gb RAM, image processing takes 0.15 second and the average computation time required for a few seconds per test image. To access the performance of our approach, a total of 250 images, which were mostly selected from the public database and web site, were recorded in experiments and the recognition rate is 91.2%. The mistakes are due to false extraction the hand region from the background. The test image which is obtained from [2] is tested by our approach, as showed in Fig.5. As we can see, the largest coefficient is associated with the images from class 5 and other coefficients related to the images from other classes are all less than 0.2.

**A. Scaling and Rotation**

The scaling and rotation of the hand region images were selected to test the performance of our approach in practice.
Our approach is not sensitive to scaling of the hand region, because region resizing is included in the normalization. Database is built with the rotation hand postures, thus the test images with rotation hand posture are also able to be well represented. The test result is showed in Figure 6. The two largest coefficients are both associated with the images from class 3.

B. Robustness to Partial Occlusion

Most of previous approaches were not able to recognize the hand postures with partial occlusion, especially in extracting the shape [15] and contour features [16]. In Fig.7, we cover partially fingers region of the image so as to illustrate proposed approach is robust to partial occlusion, such as fingers and palm were partial covered. As we see, nevertheless three fingers were partially covered in the test, the image related to the largest coefficient belongs to class 5, with the correct classification. Through testing more than 50 partial occlusion images, the proposed approach still achieves better recognition rate unless majority of finger or palm regions are covered.

V. CONCLUSIONS

In this study, a novel and robust technique is proposed to recognize a set of hand posture from images. The results show that this method can achieve satisfactory performance of 91.2%. Robustness to rotation and occlusion allows the approach to tolerate small variation or misalignment of the finger and palm in the images. It is easy to extent the database for more hand postures recognition. However, the number of database images required to represent the hand posture images under various poses may be excessively large and more time will be taken for computing. The more sparse the coefficient vector $\vec{x}$ is, the easier it will be to accurately classify the test image. How to obtain better result with smaller database is an important direction for the future research.

REFERENCES