2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

A Novel Hand Posture Recognition System Based on Sparse Representation Using Color and Depth Images

Dan Xu, Yen-Lun Chen, Xinyu Wu, Wei Feng, Huihuan Qian and Yangsheng Xu

Abstract— **Hand posture is a natural and effective human robot interaction way. In this paper, an user-independent hand posture recognition system using depth and color images captured from an RGB-D camera is presented. To recognize hand posture against complicated background conditions, we propose a novel method for automatic and accurate hand posture segmentation which detects the hand with Chamfer matching, tracks the hand with Kalman filter and segments the hand with region growing algorithm only in the depth space. A new hand posture descriptor invariant to scale, shift and in-plane rotation is constructed with the combination of local contour Fourier descriptor and global Bag-of-Features (BoF) descriptor based on Scale Invariance Feature Transform (SIFT). The sparse representation-based classification (SRC) is applied to perform the hand posture recognition task in the system. Experiments with a self-built large scale hand posture database collected online show the robustness and effectiveness of the proposed system.**

I. Introduction

Human robot interaction (HRI) is an attractive topic in the computer vision and robotics research community. As an effective and natural human-robot interaction interface, vision-based hand posture recognition has been studied by many researchers for years. However, due to the complicated background in practical interaction applications, illumination conditions and the highly deformable structure of the hand, hand posture recognition remains to be a challenging problem.

In general, vision-based hand posture recognition systems can be categorized into 3D model-based and appearance-based approaches [1]. The 3D model-based approaches can provide a wide class of hand postures through building a rich description hand model, which however, is a complicated procedure requiring a huge hand-image database containing all the characteristic hand shapes under various views. Besides, in the feature extraction of 3D model-based approaches, handling the singularities from ambiguous views is still a difficult problem.

In this paper, we focus on the appearance-based hand posture recognition, which generally includes hand detection and segmentation, feature representation of the hand posture, and classification. Many previous research papers implement hand detection and segmentation in color images using skin color information. Skin color-based methods have low computational complexity, but skin-color detection is greatly influenced by the background and illumination [2], which thus can not output robust enough segmentation results for the demands of hand posture recognition in the dynamic environment. Efficient feature representation of the hand posture is crucial for the recognition performance of the system. Local invariant descriptor such as SIFT feature, shape descriptors such as region moments and Zernike moments are popular descriptors used in several hand posture recognition systems [3] [4], but in these systems only a single feature is used without considering the fusion of different levels of features to comprehensively represent the hand posture. In current research work, various classification models such as support vector machine (SVM) [5], Hidden Markov Model (HMM), Hidden Conditional Random Fields (HCRF) [6] are extensively employed in the hand posture recognition. These models all have the need of model selection and parameter training which is critical for classification performance.

In this paper, we present a robust hand posture recognition framework for human robot interaction as illustrated in Fig. 1. Color and depth images taken from an RGB-D camera are used as the input of the system. To avoid the influence of complicated background and illumination conditions, we propose a novel method for hand detection, tracking and segmentation using the depth image only. The extracted hand silhouette in the depth image is mapped as an image mask for background subtraction in the corresponding color image. A local contour Fourier descriptor extracted from the depth posture image and a global Bag-of-Features descriptor with SIFT extracted from the color posture image are fused to represent the hand posture invariant to scale, translation and in-plane rotation. The sparse representation with l_1-norm minimization is applied into the system and obtain promising recognition results.

II. Related Work

There have been several research papers published on the appearance-based hand posture recognition in the literature. Triesch et al. propose a person-independent system using Elastic Graph Matching to recognize hand postures against complex background [7]. The system locates the hand posture region simply relying on skin-color detection and only a small scale dataset used for training the classifier resulting an accuracy rate of 85.8%. Fang et al. use an adaptive hand

The work described in this paper is partially supported by the National Natural Science Foundation of China (61005012), Shenzhen Fundamental Research Program (JC201105190948A), and Guangdong Innovative Research Team Program (201001D0104648280).

D. Xu, Y.-L. Chen, X. Wu , W. Feng are with Guangdong Provincial Key Laboratory of Robotics and Intelligent System, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China {dan.xu, yl.chen, wei.feng}@siat.ac.cn.

X. Wu is the corresponding author {xy.wu@siat.ac.cn}.

X. Wu, H. Qian and Y. Xu are also with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China {hhqian, ysxu}@mae.cuhk.edu.hk.

978-1-4673-6357-0/13/$31.00 ©2013 IEEE

segmentation scheme with the color and motion cues and recognize the hand posture through a palm-like and finger-like structure description [8]. To avoid the low robustness of hand segmentation in the color images, some researchers carry out the hand posture recognition using the depth images [9] [10]. Because the depth images are not affected by the illumination, in our approach, we detect, track and segment the hand in the depth images.

Scale Invariance Feature Transform (SIFT) features [11] are invariant to scale, orientation and translation changes, which have been used in various applications such as object recognition, tracking and gesture recognition. Wang et al. use Adaboost combined with SIFT features for hand posture recognition under slightly cluttered background [3]. Fourier descriptor is an efficient method for representing shape feature invariant to scale, rotation and translation. Compared with most of other shape descriptors, it has advantages of perceptually meaningful representation and easy normalization [12]. In our system, the shape feature of the hand posture is described by the Fourier descriptor and the local characteristic of the hand posture is described by SIFT features. These two levels of feature are combined to comprehensively represent the hand posture for improving the recognition performance.

The sparse representation has been successfully applied in different fields of pattern recognition in recent years for its effectiveness for representing and compressing high-dimensional signal data. A sparse representation-based classification (SRC) framework is discussed in-depth by Huang et al. [13]. Wright et al. use the sparse representation-based method for the face recognition under varying illumination and occlusions [14] and obtain robust recognition results. Unlike classic classifier such as SVM, the SRC provides a new solution for pattern recognition task, which represents a test sample with the whole training samples rather than choosing subsets of corresponding classes for training the model. We apply the SRC into our system for the hand posture classification.

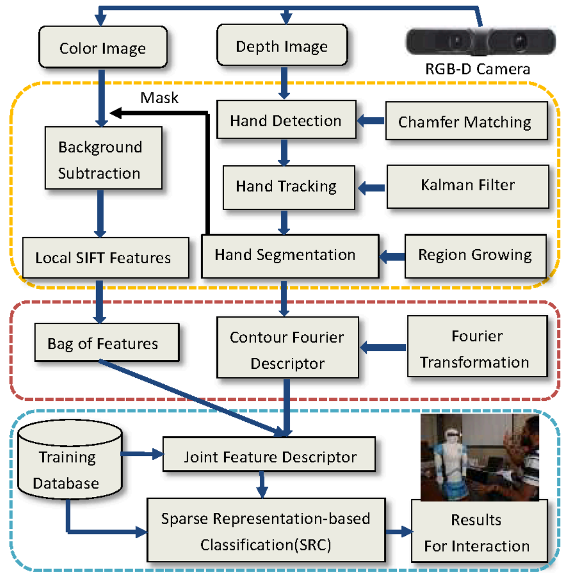


Fig. 1. The flowchart of the proposed hand posture recognition system.

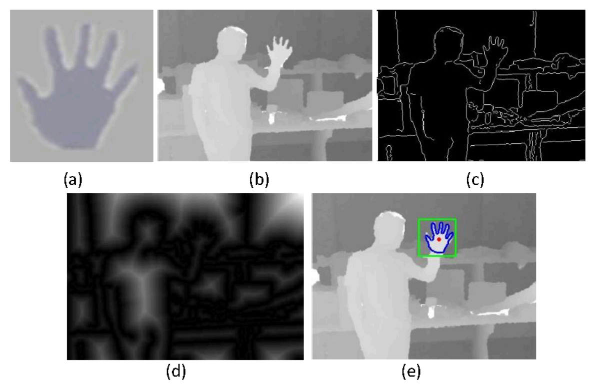


Fig. 2. Experimental examples of hand detection using chamfer distance matching. (a) is an "open" hand posture template. (b) is the current input depth frame after the depth smoothing. (c) is the binary edge image obtained from Canny edge detector. (d) is the distance map from the distance transform of the edge image, and (e) is the hand detection result, in which the region with the highest match score is kept.

III. Hand Posture Extraction from Depth and Color Images

The depth and color images are captured from an RGB-D camera (The inexpensive Kinect sensor is used in the system). For the depth images, there are some missing points with depth value 0 due to the high refraction of the structure light. We first perform a preprocessing by using a median filter with a 4×4 sliding window to smooth the depth images.

A. *Hand Detection on 2D Depth Image*

Hand detection is an important and difficult task in the hand posture recognition system. In most previous research works, hand detection is performed on color images using different color spaces such as RGB, HSV and YCbCr. In these works, skin color is usually used as a key clue to find hand candidates. However, skin color distribution is greatly influenced by the illumination changes and it is hard to obtain robust skin color segmentation result under complex illumination conditions. In addition, skin color-like object interference and the dress of the operator (short sleeve or long sleeve) are also problems for further confirming the position of the hand.

To avoid the restrictions of hand detection in the color space, we detect the hand only using the depth data. Since the depth images do not have rich local texture compared with color images, they can provide better edge maps for shape matching. In the system, we use the chamfer distance matching to find the hand in the depth images with a pre-specified hand posture. Chamfer distance matching [15] is a popular technology for matching two edge maps, which has low computational overhead and high robustness to clutter. The first step is to obtain the edge maps from both the hand shape template and the query depth image by a canny edge detector. Then a sliding window with the template edge image of different scales is used to match in the query depth edge image. The similarity between these two edge images is measured by the chamfer distance. If we let $U_T(u_i \in U_T, i = 1,2,...,n)$ and $V_Q(v_j \in V_Q, j = 1,2,...,m)$ represent the point set of the template edge image and the sub-window edge image respectively, the chamfer distance between U_T and V_Q can be computed as follows:

$$d_{cham}(U_T, V_Q) = \frac{1}{n} \sum_{u_i \in U_T} \min_{v_j \in V_Q} \|u_i - v_j\|. \tag{1}$$

The value of d_{cham} actually is the mean of distances between each point $u_i \in U_T$ and its nearest edge point in V_Q. To reduce the matching cost, the chamfer distance can be computed effectively by the distance transform (DT), which converts the query binary edge image into gray image by assigning each edge pixel with zero and each non-edge pixel with the distance value to its nearest edge point. Through the sliding window match, we keep several hand candidates according to the chamfer distance score d_{cham}. Since the hand is in front of the background when the gesture interaction starts, we can define the final match score S_{hand} for every hand candidate as follows:

$$S_{hand} = (1 - \frac{v_{can}}{v_{max}}) * d_{cham}. \tag{2}$$

Here, v_{max} represents the maximum depth distance of the depth sensor and v_{can} represents the depth distance of the hand candidate. The value of S_{hand} is used to confirm the best hand match position in the system. Examples of the hand detection based on chamfer matching are shown in Fig. 2.

B. Hand Tracking and Segmentation in 3D Depth Space

To recognize dynamic hand posture, the system needs to track the hand based on the initial hand detection result. Unlike in the color space, hand tracking in the depth space can not use color features, but we have the movement information of the hand in the 3D space. Due to the similarity of the hand velocity between adjacent frames, so we can estimate the hand position in the current frame based on the hand movement of the previous frame:

$$P_{current} = P_{previous} + V_{control} \Delta t. \tag{3}$$

where $P_{current}$ and $P_{previous}$ represent the point cloud coordinates of the hand central points in adjacent frames; $V_{control} = \{v_x, v_y, v_z\}$ is a control vector with the hand velocity values in x, y and z axis; Δt is the interval time between the current frame and previous frame. In our system, the Kalman filter technique [16] is used to track the hand in the depth space.

Hand segmentation is a necessary step for the hand posture recognition. The hand centroid point position from the hand segmentation can be used to correct the predicted position in the hand tracking. Since the depth value of the hand surface in the point cloud is continuous, we can use a classic region growing approach [17] to segment the hand region. We select the hand central point obtained from hand tracking as a seed point and define an initial hand region A which is a 8-connectivity region of the seed point. If we let $d(x)$ represent

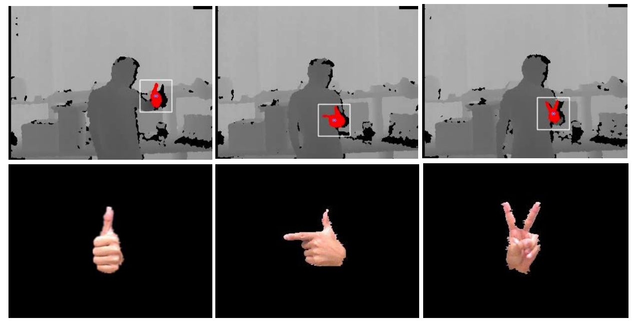


Fig. 3. Experimental examples of hand posture detection. The hand region extracted in the depth image can be mapped as an image mask to segment the hand in the corresponding color image.

the depth value of a point x and let N represent the set of neighbors around A, then we can allocate a point x in N into A through a difference measure δ:

$$\delta(x) = |d(x) - \frac{1}{n} \sum_{y_i \in A, i=1}^{n} [d(y_i)]|, \tag{4}$$

where n is the number of points in the region A. After traversing all boundary points in N, A and N are both updated. If we repeat this operation until there is no neighboring point whose difference value is less than a specified threshold, the process is terminated and A is the final hand region. To avoid part of the wrist is also segmented, we use edge and area of the hand as prior knowledge to improve the accuracy of hand segmentation. Fig. 3 shows examples of hand segmentation results in both depth and color images.

IV. Feature Representation

Once a hand posture is segmented from both the depth and color images, a new feature descriptor which combines local and global feature is proposed for improving the hand posture recognition performance. The joint feature descriptor is constructed by using a local contour feature representation based on Fourier transform and a global bag-of-features (BoF) representation based on scale invariant feature transform (SIFT).

For the hand posture classification, 2D contour is important and discriminative information, which can be deemed as a local shape feature of the hand. Benefit from the effective hand posture segmentation in the depth image of the system, we can extract the contour feature by using an invariant Fourier descriptor. If we let $k(n) = \{x_n, y_n\}, (n = 1, ..., N)$ denote a series of discrete ordered points of a segmented hand posture contour, a Fourier descriptor $F = \{F_k\}$ of the contour can be calculated from the 1-D discrete Fourier transform applied on the sequence $k(n)$:

$$F_k = \frac{1}{N} \sum_{n=0}^{N-1} k(n) e^{-\frac{2\pi jnk}{N}} \quad k = 0, 1, ..., N-1, \tag{5}$$

Based on three important properties of the Fourier descriptor: the information of shape is described through the spectral magnitude; the information of rotation and the start point is described through the phase; the information of translation

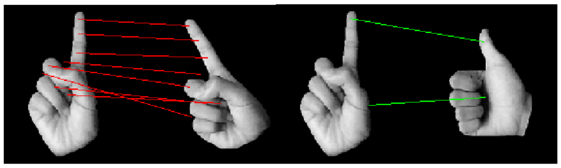


Fig. 4. Local SIFT interest point matching results between two selected postures with similar shape. Obviously, the matching rate between rotated 'One' postures is higher than that between postures 'One' and 'Thumb'.

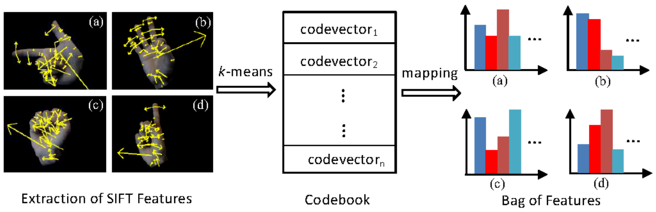


Fig. 5. The schematic diagram of generating global Bag-of-features descriptor with SIFT.

is described through the zero component, we can extract the contour feature invariant to scale, rotation, translation and the start point via discarding the zero component of F and normalizing the spectral magnitude:

$$c = [\frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \ldots, \frac{|F_{N-1}|}{|F_0|}] \tag{6}$$

Since higher dimensional elements of F only focus on the details of the contour, we only keep the first few elements in our final local contour feature descriptor c.

Since the SIFT feature has strong discrimination for similar shapes as shown in Fig. 4, it is used as a supplement of the local contour Fourier feature. We detect local keypoints using the SIFT algorithm in the color hand posture image, and extract a 128-dimensional SIFT feature vector for each keypoint to represent local features of a hand posture. Then a SIFT feature descriptor of the hand posture is constructed with a global Bag-of-Features approach. For all SIFT feature vectors extracted from the whole training database of hand posture images, k-means algorithm is used to cluster them into specified number of clusters whose centers are usually called codevectors of a codebook. The codebook describes local feature patterns in the training images, and is used for quantizing SIFT feature vectors of a hand posture image by mapping each feature vector into the nearest codevector with the Euclidean distance metric. Then the global descriptor of the hand posture (denoted as s) using Bag-of-Features based on SIFT (BoF-SIFT) is represented with a normalized frequency histogram of the codevectors as illustrated in Fig. 5.

V. Sparse Representation for Hand Posture Recognition

In this section, we discuss the sparse representation for the hand posture classification. For each pair of training depth and color images, we can obtain a joint feature vector $v = [c_p, s_q]^{\mathrm{T}} \in \mathbb{R}^m$ $(m = p + q)$ which combines p dimensional local contour Fourier feature and q dimensional bag-of-features. Assume there are k classes of hand postures for classification and we have n_i training samples, namely n_i joint feature vectors for the i-th class. A matrix $H_i = [v_1, v_2, \ldots, v_{n_i}] \in \mathbb{R}^{m \times n_i}$ can be defined for the training samples of the i-th class, and the whole training set of all classes can be expressed with the matrix:

$$H = [H_1, H_2, \ldots, H_k] \in \mathbb{R}^{m \times n} \tag{7}$$

where $n = \sum_{i=1}^{k} n_i$ ($n \gg m$ in the hand posture recognition case) is the number of all training samples. For a new test sample $h \in \mathbb{R}^m$, its linear representation with a combination of the training samples can be given by

$$h = Ht, \tag{8}$$

where t is a coefficient vector with only a few entries is nonzero. The problem of finding the sparse solution of the underdetermined system described with equation (8) can be treated as a ℓ^1-minimization problem:

$$\hat{t}_1 = \arg\min_t \|t\|_1 \quad \text{subject to} \quad \|Ht - h\|_2 \le \varepsilon, \tag{9}$$

where $\varepsilon > 0$ is a tolerance error threshold. The ℓ_1-norm minimization is actually a convex optimization problem and can be effectively solved with linear programming algorithms [18]. For the hand posture classification based on sparse representation, the hand posture class label L of the input test sample h can be determined by minimizing the residual as follows:

$$L(h) = \arg \min_{i=1,2,\ldots,k} \|h - H\delta_i(\hat{t}_1)\|_2. \tag{10}$$

Here, the function $\delta_i(\hat{t}_1)$ is used to generate a coefficient vector whose elements are the coefficients only associated with the i-th class from the sparse coefficient vector $\hat{t}_1$.

VI. Experimental Evaluation

The hand posture recognition system is developed as an interaction interface for a humanoid robot designed in our lab as shown in Fig. 6. There is an RGB-D camera (a Microsoft Kinect sensor) in the eye position of the robot, which captures 640×480 color images and 320 × 240 depth images as the input of the system with 30 frames per second in average. The basic processing platform of the robot is Intel Core i5 2.53 GHz CPU with 2 GHz memory.

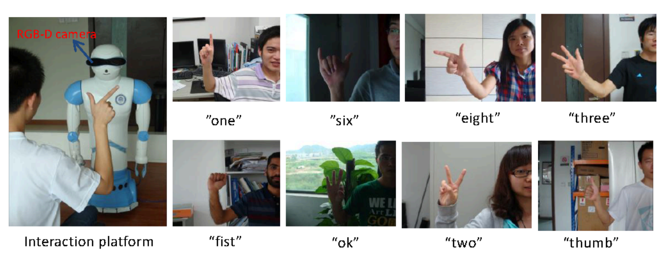
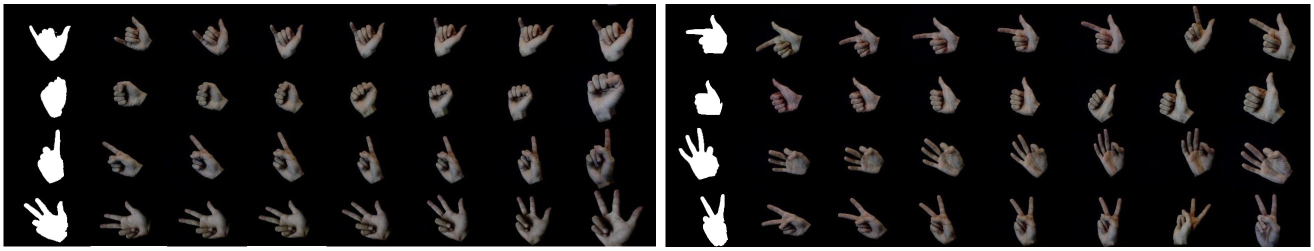


Fig. 6. User-independent human robot interaction with different types of postures against complex background and varying illumination.

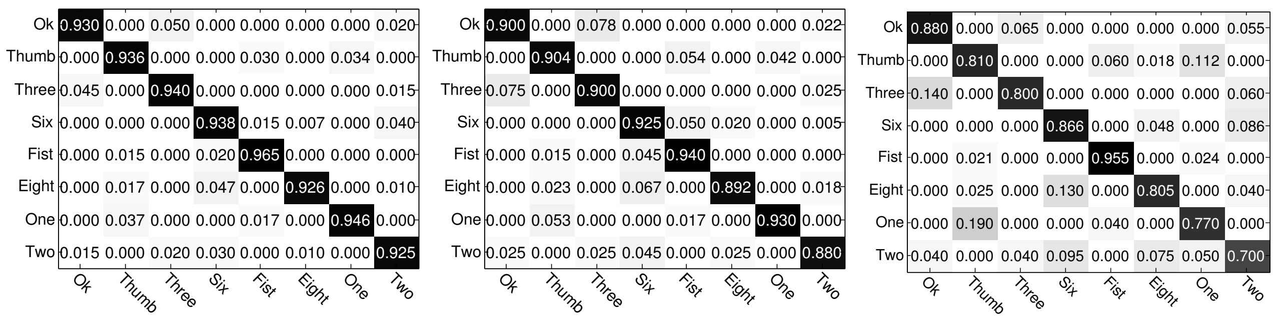
To evaluate the performance of the proposed hand posture recognition system, we establish a hand posture database



(a) Hand posture classes: 'six','fist','one','three'

(b) Hand posture classes: 'eight','thumb','ok','two'

Fig. 7. Examples of hand posture segmentation results using hand posture extraction module of the system. The first column is the binary silhouettes of hand postures obtained from depth stream. The other columns are hand posture samples segmented from color stream.



(a) SRC with joint feature descriptor(Average accuracy = 93.8%)

(b) SRC with global BoF-SIFT feature (Average accuracy = 90.1%)

(c) SRC with local contour Fourier feature (Average accuracy = 83%)

Fig. 8. The confusion matrixes of recognition using SRC with single and joint feature descriptors (60-dimension).

with the robot platform. The database currently contains 8 different types of hand postures which are collected under complex scene conditions such as cluttered background and varying illumination as shown in Fig. 6. To guarantee the user independence of recognition, each type of hand posture is performed by 10 different people and a pair of 2-minute video clips from calibrated depth and color stream are captured by each people for it. Each frame of the video is a hand posture sample. Besides, we also make the posture samples with scale, viewpoint and in-plane rotation (0-90°) changes for evaluating the robustness of recognition. The hand posture extraction module allows the system to segment hand posture online in the depth and color spaces for further feature extraction. Fig. 7 shows examples of hand posture segmentation results from the original depth and color video in the database.

The hand posture database is equally divided into two parts: one part is used for the dictionary construction of sparse representation; the other part is used for the test. Each part contains all types of hand postures. The joint feature vector is extracted from a pair of depth and color hand posture sample with the procedure described in Section IV. Before the test experiments, the length of joint feature vector (i.e. m) and the ratio of the dimension of global BOF-SIFT feature in joint feature vector (i.e. q/m) is preliminarily set to 60 and 0.6, respectively. To improve the computational performance, we also optimize the training hand posture database with a dictionary learning algorithm K-SVD [19] to remove highly similar samples in the training database, then replace the large and dense dictionary with a small and sparse one. Fig. 8 shows the test results of using sparse representation with single and joint feature descriptors respectively. It can be noted that using joint feature representation the system yields an average accuracy of 93.8%, which outperforms that of using single global BOF-SIFT feature (90.1%) or local contour Fourier feature (83%).

To evaluate how the performance is affected by the length of the joint feature vector and the ratio of the dimension of global BoF-SIFT feature in joint feature, we carry out a series of experiments with the test samples. Fig. 9 shows the results of average accuracy of recognition with varying lengths (from 20 to 120) and ratios (from 0.1 to 1). We can note that when the length of joint feature is less than 60 the performance can obtain significant improvement with the increase of the feature length, while the performance only presents slight difference when the length is larger than 60. This is because for BoF-SIFT if the codebook can represent the hand posture samples compactly, increasing the size of the codebook further pays little effect on the final classification performance. For the contour Fourier feature, increasing the dimension can only bring more details of the contour but not the most useful shape information for classification. Larger dimension of the joint feature will also generate bigger computational overhead. Under all different lengths of the joint feature, the accuracy of recognition is the best when the ratio is between 0.6 and 0.7. This implies that the global BoF-SIFT component is more important than local contour Fourier component for the classification, and it can help to improve the discrimination for the hand postures with similar shape which is a major limitation for the contour

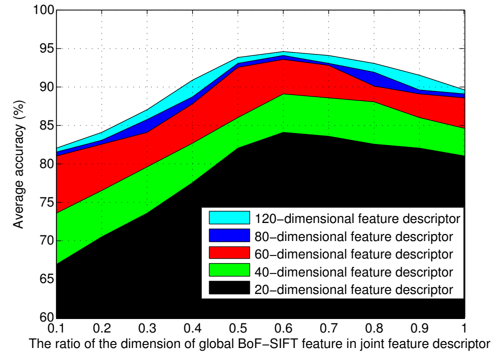


Fig. 9. The performance of recognition with varying lengths of the joint feature descriptor and varying ratios of the dimension of global BoF-SIFT in the joint feature.

TABLE I

PERFORMANCE COMPARISON ON THE SELF-BUILT HAND POSTURE DATABASE.

Method	Average Accuracy
(1)Nearest Neighbor + Joint Feature	83%
(2)Nearest Subspace + Joint Feature	87%
(3)Linear SVM + BoF-SIFT	89%
(4)Linear SVM + Joint Feature	91%
(5)HCRF + Distance + SIFT [6]	90%
(6)Adaboost + SIFT [3]	88%
Our Method (SRC + Joint Feature)	93.8%

Fourier feature.

Based on the self-built hand posture database, we conduct comparison experiments with other popular methods and classifiers. Table I shows the results of the performance comparison. In the experiments, methods (1) (2) (3) (4) replace the sparse representation-based classification (SRC) approach with Nearest Neighbor, Nearest Subspace and Linear SVM and utilize the proposed feature representation and hand posture extraction modules. It is clear that the SRC produce better recognition accuracy than these supervised classifiers who need the model selection which is critical for the classification performance. Methods (5) (6) use HCRF and Adaboost with SIFT using only color hand posture samples and the features are extracted with the background. The joint feature representation with color and depth information and effective hand posture extraction scheme make our method obtain a better performance. For the computational speed aspect, the computation of a pair of color and depth frames requires 160 ms in average, namely about 6 frames per second running on our robot platform. This is a near real-time processing for hand posture interaction applications.

VII. CONCLUSIONS

In this paper, a novel user-independent hand posture recognition system using the sparse representation method with the combination of different-level features of the hand posture is proposed. A reliable hand posture extraction approach in the color and depth images from an RGB-D camera allows the system to work against complicated background and varying illumination conditions. The experiments on a self-built database containing 8 types of hand postures collected by 10 different individuals show that the system can recognize hand postures with scale, viewpoint and in-plane rotation (0-90°) changes and an average recognition rate of 93.8% is presented. The experimental results also demonstrate that the fusion of local contour Fourier descriptor and global BoF-SIFT descriptor can significantly improve the recognition performance. In future, we plan to add more posture types into the database and develop interesting interaction applications on the robot platform based on the proposed system.

REFERENCES

[1] H. Zhou and T. Huang, "Tracking articulated hand motion with eigen dynamics analysis," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2003, pp. 1102–1109.

[2] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in *Proc. Graphicon*, vol. 3. Moscow, Russia, 2003.

[3] C. Wang and K. Wang, "Hand posture recognition using adaboost with sift for human robot interaction," *Recent Progress in Robotics: Viable Robotic Service to Human*, pp. 317–329, 2008.

[4] L. Gu and J. Su, "Natural hand posture recognition based on zernike moments and hierarchical classifier," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2008, pp. 3088–3093.

[5] N. Dardas, Q. Chen, N. Georganas, and E. Petriu, "Hand gesture recognition using bag-of-features and multi-class support vector machine," in *IEEE International Symposium on Haptic Audio-Visual Environments and Games (HAVE)*. IEEE, 2010, pp. 1–5.

[6] T. Liu, K. Wang, A. Tsai, and C. Wang, "Hand posture recognition using hidden conditional random fields," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*. IEEE, 2009, pp. 1828–1833.

[7] J. Triesch and C. Von Der Malsburg, "A system for person-independent hand posture recognition against complex backgrounds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1449–1453, 2001.

[8] Y. Fang, K. Wang, J. Cheng, and H. Lu, "A real-time hand gesture recognition method," in *IEEE International Conference on Multimedia and Expo*. IEEE, 2007, pp. 995–998.

[9] Z. Mo and U. Neumann, "Real-time hand pose recognition using low-resolution depth images," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2. Ieee, 2006, pp. 1499–1505.

[10] K. Fujimura and X. Liu, "Sign recognition using depth image streams," in *International Conference on Automatic Face and Gesture Recognition*. IEEE, 2006, pp. 381–386.

[11] D. Lowe, "Object recognition from local scale-invariant features," in *IEEE International Conference on Computer Vision (CVPR)*, vol. 2. IEEE, 1999, pp. 1150–1157.

[12] D. Zhang and G. Lu, "A comparative study of fourier descriptors for shape representation and retrieval," in *Proceedings of the Fifth Asian Conf. on Computer Vision*. Citeseer, 2002, pp. 646–651.

[13] K. Huang and S. Aviyente, "Sparse representation for signal classification," *Advances in neural information processing systems*, vol. 19, p. 609, 2007.

[14] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[15] M. Liu, O. Tuzel, A. Veeraraghavan, and R. Chellappa, "Fast directional chamfer matching," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 1696–1703.

[16] R. Kalman *et al.*, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[17] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, 1994.

[18] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, pp. 129–159, 2001.

[19] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.