Real-Time 3D Hand Gesture Recognition from Depth Image

Lin Song[1,a], Ruimin Hu[1,b], Yulian Xiao[2,c] and Liyu Gong[2,d]
Correspondence author: Ruimin Hu

[1]National Engineering Research Center for Multimedia Software,
Computer School, Wuhan University, China

[2]Eedoo Inc., China

[a]lin_song511@163.com, [b]hrm1964@163.com, [c]xiaoyl@eedoo.cn, [d]gongliyu@gmail.com

Keywords: Hand Gesture Recognition; 3D Image Processing; Depth Images; Kinect; Human Computer Interaction

Abstract. In this paper, we propose a novel real-time 3D hand gesture recognition algorithm based on depth information. We segment out the hand region from depth image and convert it to a point cloud. Then, 3D moment invariant features are computed at the point cloud. Finally, support vector machine (SVM) is employed to classify the shape of hand into different categories. We collect a benchmark dataset using Microsoft Kinect for Xbox and test the propose algorithm on it. Experimental results prove the robustness of our proposed algorithm.

Introduction

Human computer interaction (HCI) is one of the most important research topics in information technology area. Human hand movement is a natural way to interact with computer, especially for gaming. Automatically hand gesture recognition is difficult to solve. Traditional hand gesture recognition algorithms are based on RGB color information [1-7] which captures 2D appearance cues only. However, hand gesture information is mostly described by the 3D shape of human hand. Obtaining 3D information using 2D cameras are not robust and computing intensive. Traditional 3D scanner are slow and expensive thus not suitable for ordinary applications. Some wearable devices [9-10] can capture 3D information effectively but its applications are restricted. Recently, 3D cameras are becoming cheaper and faster. Many companies (e.g. Microsoft, PrimeSense [11] etc.) have launched their 3D cameras which can capture depth information as well as RGB color information. These affordable and fast cameras open a door for researchers to design new algorithms for hand gesture based HCI.

In this paper, we propose a depth image based 3D hand gesture recognition algorithm for HCI. We segment hand regions from the depth images and convert them into 3D point clouds. 3D moment invariants are then computed as feature descriptors. The features encoded the shape information of human hand. The gesture categories are then recognized by classification on the feature descriptors using support vector machine (SVM). Results on a dataset collected using Microsoft Kinect for Xbox prove the effectiveness of our proposed algorithm.

Proposed method

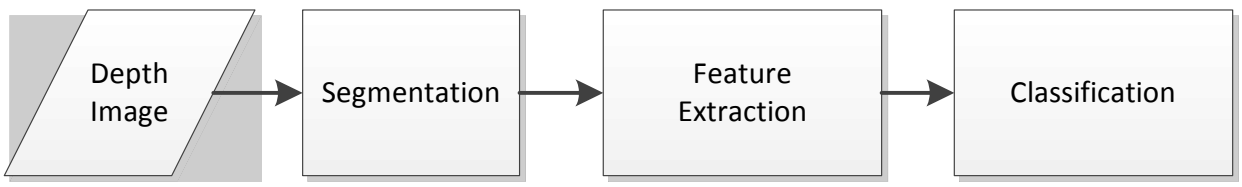

Figure 1. The processing chain of our propose hand gesture recognition algorithm.

Published by Atlantis Press, Paris, France.

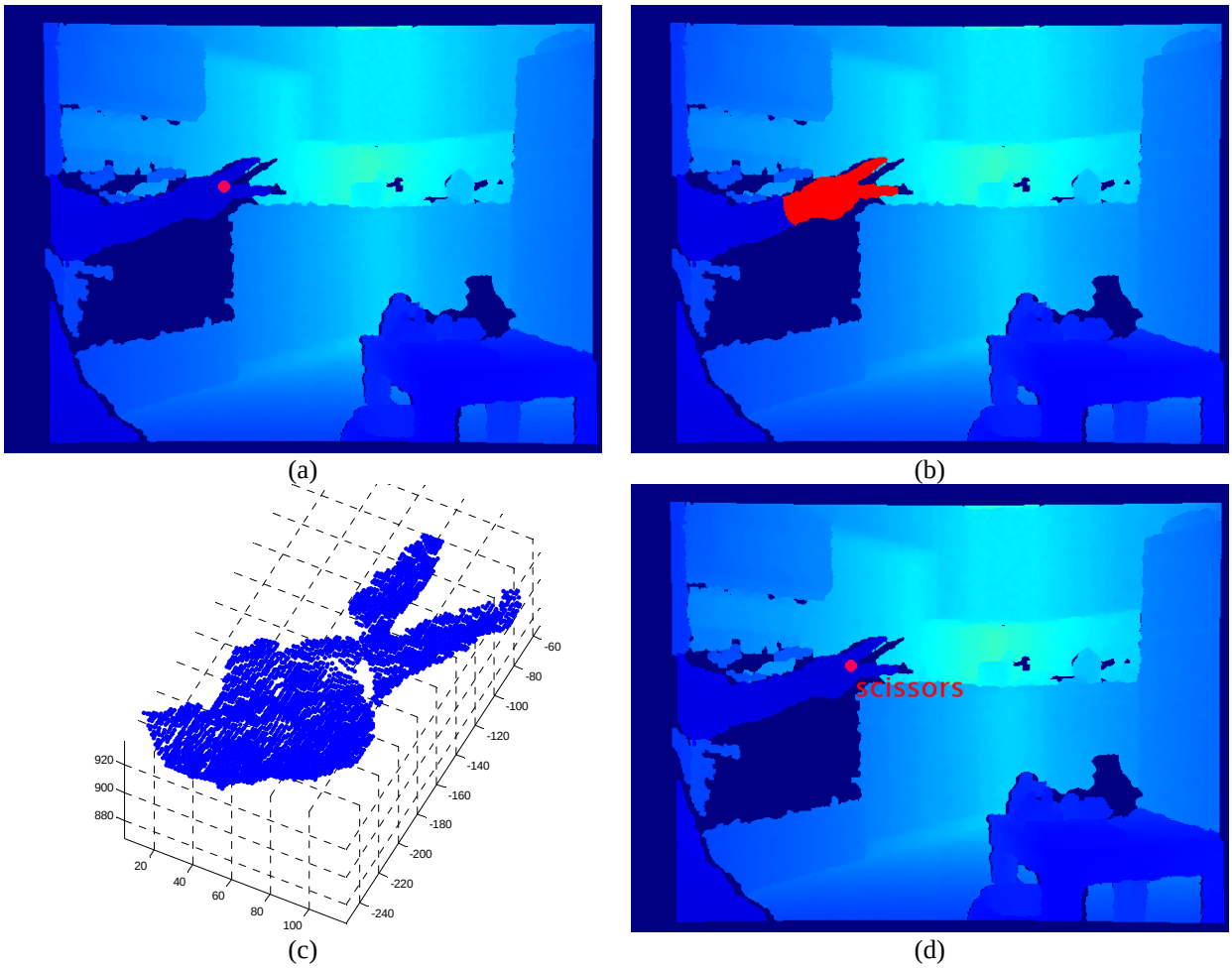

Figure 2. Illustration of our proposed hand gesture recognition algorithm. (a): Input depth image with a hand point; (b): Segmented hand region; (c): Point cloud converted from the depth information of the segmented hand region. Shape features are then extracted from the point cloud; (d): Classification result.

Figure 1 and Figure 2 give an overview and illustration of our proposed hand gesture recognition method. Firstly, the hand region are segmented from input depth image using a given hand position. The hand position can be obtained using a 3D hand tracking algorithm. Then, we convert the segmented hand region into point cloud and extract 3D shape feature of the point cloud. Finally, the features are classified as different gesture categories using machine learning algorithms.

Segmentation. Given a depth image and a 3D point as hand position, we segment out hand region by choose the points within a sphere around the hand position in 3D space. However, if the hand is close (but not connected) to a background object (e.g. table), the region segmented using a sphere with R center at the hand point may contains some pixels from the background. We use a 1-d connect component selection algorithm which is described below to eliminate these background pixels:

Initial: depth value z of hand position; a scalar value r specify the maximum length within which two points are considered as connected; a range R=[z_1 , z_2] with z_1 = z-r and z_2 = z+r; a collection of points P whos depth values fall in the range [z1, z2]
While (any pixel falls in range [z_1-r, z_1))
    Add the pixels in [z_1-r, z_1) to P and set z_1=z_1-r;
End
While (any pixel falls in range (z_2, z_2+r])
    Add the pixels in (z_2, z_2+r] to P and set z_2=z_2+r;
End
Return P

Feature extraction. To extract features of hand shape, we convert the segmented hand region into a point cloud first. Then we compute 3D moment invariants [11] on the point cloud.

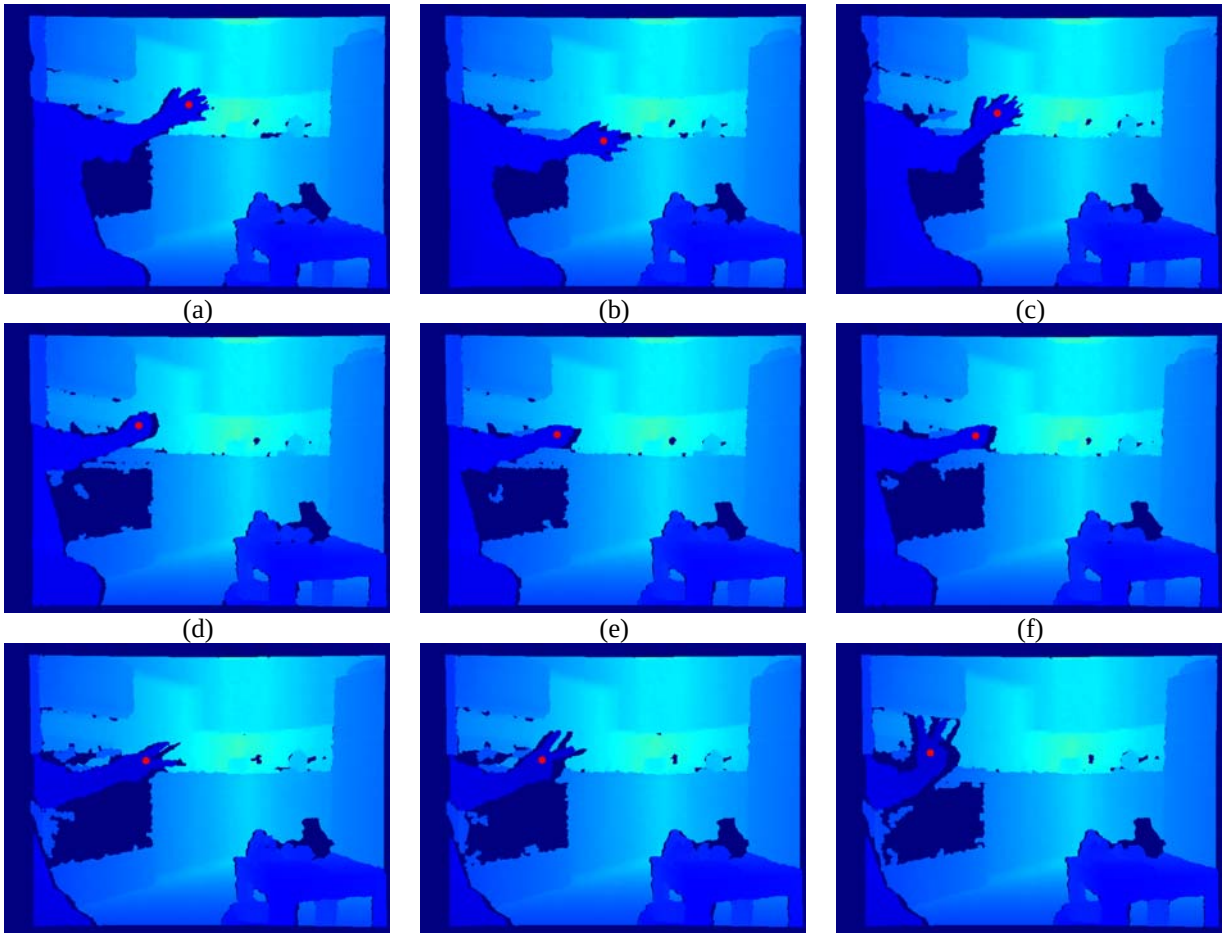

Figure 3. Examples of the benchmark dataset. The first, second and third rows are examples from "paper", "rock", "scissors" catogories respectively. Each sample contains a depth image and a 3D point indicates the postion of hand.

Let $(\bar{x}, \bar{y}, \bar{z})$ be the centroid of the hand point cloud. Define the central moment μ_{pqr} as:

$$\mu_{pqr} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^p (y_i - \bar{y})^q (z - \bar{z})^r \quad (1)$$

Where N is the number of points in the point cloud. With this, we calculate three moment invariants:

$$J_1 = \mu_{200} + \mu_{020} + \mu_{002} \quad (2)$$

$$J_2 = \mu_{200}\mu_{020} + \mu_{200}\mu_{002} + \mu_{020}\mu_{002} - \mu_{110}^2 - \mu_{101}^2 - \mu_{011}^2 \quad (3)$$

$$J_3 = \mu_{200}\mu_{020}\mu_{002} + 2\mu_{110}\mu_{101}\mu_{011} - \mu_{002}\mu_{110}^2 - \mu_{020}\mu_{101}^2 - \mu_{200}\mu_{011}^2 \quad (4)$$

We also compute the three eigen values $\lambda_1, \lambda_2, \lambda_3$. Then we concatenate the six features into a 6-d feature vector. Before concatenating, we normalize the 3D moment invariants as $\hat{J}_i = \log(J_i)/3$ and the three eigen values as $\hat{\lambda}_i = \log(\lambda_i)/2$. In our experiments, such a normalization can improve the performance significantly.

Classification. We employ support vector machine (SVM) as our classifier. In our experiments, a linear SVM with soft margin (C=300) is used.

Experimental results

To test our proposed algorithm, we collect a benchmark dataset using Microsoft Kinect for Xbox. Three gesture categories (paper, rock and scissors) are recorded. Each category contains about 400 samples. Each sample contains a 640x480 size depth image and a 3D point represents the hand position. Hand positions are captured by a 3D hand tracking algorithm during data collection. Figure 3 present some example depth images of our dataset.

Published by Atlantis Press, Paris, France.

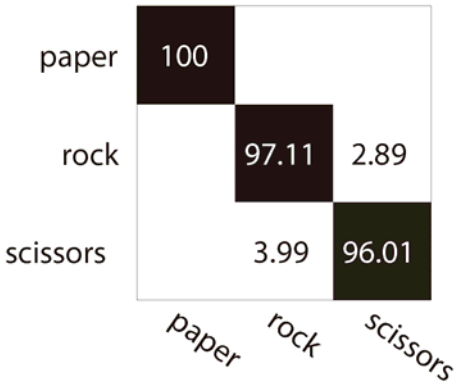

Figure 4. Confusion matrix of the proposed hand gesture recognition on the benchmark dataset

We use the linear SVM algorithm implemented in the Libsvm software package [12] as our classifier. The classifier are tested by a 10 fold cross validation on the collected dataset. We observe an average classification accuracy as 97.7%, which proves the effectiveness of the proposed algorithm. Figure 3 gives the detail confusion matrix of the result. From figure 3, we can see that the major confusion is between rock and scissors gestures. For paper gesture, we get perfect 100% accuracy.

References

[1] V. I. Pavlovic, R. Sharma and T. S. Huang, Visual interpretation of hand gestures for human-computer interaction: a review. IEEE Transaction on Pattern Analysis and Machine Intelligence, 19(7), 677-695, 1997.

[2] A. Just and S. Marcel, A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition. Computer Vision and Image Understanding, 113(4):532-543, 2009/

[3] B. Ionescu, D. Coquin, P. Lambert and V. Buzuloiu, Dynamic hand gesture recognition using the skeleton of the hand. EURASIP Journal of Applied Signal Processing, 2005:2101-2109, 2005.

[4] R. Kjeldsen and J. Kender, Toward the use of gesture in traditional user interfaces, in Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, Killington, VT, USA, 1996.

[5] K. Imagawa, S. Lu and S. Igi, Color-based hands tracking system for sign language recognition, in IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998.

[6] C. Manresa, J. Varona, R. Mas and F. Perales, Hand tracking and gesture recognition for human-computer interaction. Electron Letter of Computer Vision and Image Analysis, 5(3):96-104, 2005.

[7] T.B. Moeslund, A. Hilton and V. Kruger, A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding, 104(2-3):90-126, 2006.

[8] D. J. Sturman and D. Zelter, A survey of glove-based Input. IEEE Computer Graphics and Applications. 14(1):30-39, 1994.

[9] R.Y. Wang and J. Popovic, Real-time hand-tracking with a color glove. ACM Transaction on Graphics. 28(3):1-8, 2009.

[10] PrimeSensor http://www.primesense.com

[11] F. A. Sadjadi and E. L Hall. Three-dimensional moment invariants. IEEE Trans. on Pattern Analysis and Machine Intelligence. 2(2), pp 127-136, March 1980.

[12] C.-C. Chang and C.-J. Lin. LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.