# Fusion hand gesture segmentation and extraction based on CMOS sensor and 3D sensor

Disi Chen, Gongfa Li*, Ying Sun, Guozhang Jiang, Jianyi Kong and Jiahan Li

College of Machinery and Automation,
Wuhan University of Science and Technology,
Wuhan, China
Email: 554693623@qq.com
Email: ligongfa@wust.edu.cn
Email: 493530316@qq.com
Email: whjgz@wust.edu.cn
Email: 15697188659@wo.com.cn
Email: jiahan.li@foxmail.com
*Corresponding author

Honghai Liu

School of Mechanical Engineering,
Shanghai Jiao Tong University,
Shanghai, China
and
School of Computing,
University of Portsmouth,
Portsmouth, PO1 3HE, UK
Email: honghai.liu@sjtu.edu.cn
Email: honghai.liu@port.ac.uk

**Abstract:** Gesture recognition is one of the most promising subjects in the field of computer vision and artificial intelligence; the development of it will have a profound influence on the research of robot control and Human–Machine Interface (HMI) and image segmentation is a key step in image recognition. This paper based on Kinect sensor developed by Microsoft gives an introduction on its hardware structure and the measuring method of depth camera. At the same time we will collect the colour hand gesture images as samples, to introduce the traditional skin colour image segmentation method in HSV and YCbCr colour space. Finally, a new 3D image hand gesture segmentation method based on depth information will be proposed.

**Keywords:** Kinect; CMOS sensor; skin colour segmentation; depth image.

**Reference** to this paper should be made as follows: Chen, D., Li, G., Sun, Y., Jiang, G., Kong, J., Li, J. and Liu, H. (2017) 'Fusion hand gesture segmentation and extraction based on CMOS sensor and 3D sensor', *Int. J. Wireless and Mobile Computing*, Vol. 12, No. 3, pp.305–312.

**Biographical notes:** Disi Chen received BS degree in Mechanical Engineering and Automation from Wuhan Textile University, Wuhan, China. He is currently occupied in his MS degree in Mechanical Design and Theory at Wuhan University of Science and Technology. His current research interests include mechanical CAD/CAE, signal analysis and processing.

Gongfa Li received the PhD degree in Wuhan University of Science and Technology, Wuhan, China. He is currently an Associate Professor in Wuhan University of Science and Technology. His major research interests are computer-aided engineering, mechanical CAD/CAE, modelling and optimal control of complex industrial process.

Ying Sun is currently an Associate Professor in Wuhan University of Science and Technology. Her major research focuses on teaching research in mechanical engineering.

Guozhang Jiang received the PhD degree in Wuhan University of Science and Technology, China. He is currently a Professor in Wuhan University of Science and Technology. His research interests are computer-aided engineering, mechanical CAD/CAE and industrial engineering and management system.

Jianyi Kong received the PhD degree in Helmut Schmidt Universitat, Germany. He is currently a Professor in Wuhan University of Science and Technology. His research interests are intelligent machine and controlled mechanism, mechanical and dynamic design and fault diagnosis of electrical system, mechanical CAD/CAE, intelligent design and control.

Jiahan Li received BS degree in Mechanical Engineering and Automation in Mechatronic Engineering from Hubei Polytechnic University, Huangshi, China. He is currently occupied in his MS degree in mechanical design and theory at Wuhan University of Science and Technology. His current research interests include mechanical CAD/CAE, signal analysis and processing.

Honghai Liu received the PhD degree in Intelligent Robotics from Kings College, University of London, London, UK. He is currently a Professor of Intelligent Systems in Portsmouth University, Portsmouth, UK. His research interests are approximate computation, pattern recognition, multi-sensor-based information fusion and analytics, human–machine systems, advanced control, intelligent robotics and their practical applications.

*This paper is a revised and expanded version of a paper entitled 'Fusion Hand Gesture Segmentation and Extraction Based on CMOS Sensor and 3D Sensor' presented at the '7th International Workshop on Swarm Intelligent Systems (IWSIS2016) & 5th International Workshop on Mobile and Wireless Computing (IWMWC2016)', Hangzhou, China.*

## 1 Introduction

With the development of intelligent robot industry, now the tasks for robots are becoming complex, especially humanoid robots. Because of their changeable using scenes and the diversity of using population, the design of human–machine interface of robot controlling is getting higher. In order to reduce the learning cost of people using a robot, variety of human–robot interaction methods have been proposed. Among them, the most popular way of human–robot interaction is to apply the computer graphics technologies into human intention acquiring. On the one hand, by using computer vision technology, the learning costs and the difficulty of people using robots can be minimised, on the other hand, with the advancement of CMOS (Wan et al., 2012) technology in image sensors and the parity of high pixel sensors, using machine vision to acquire gestures in turn to control robots becomes more and more popular.

Traditional CMOS sensors are only capable with 2D image acquisition ability, although the resolution is constantly improving, but the image only contains the colour information of each pixel, which could not convey the instruction information contained by images. Therefore, RGB bitmaps captured by sensors need a complex processing to extract the underlying instruction information. Traditional way of image processing is mainly using complex algorithms to process the image colour information and find the edges of the different objects in images, by referring to the contrasts of colour bitmap data. The most widely used algorithms are the threshold segmentation method, the edge detection method, region splitting and merging method, watershed method, graph cut, grab cut and the random walker method (Li, 2012a). Although the implementation method and the mathematical theory foundation of these algorithms are totally different, their same purposes are to tell the foreground and background apart. Since these algorithms are all based on the transition of colour in images, when the contrast of foreground and the background in images are not obvious, these algorithms cannot finish the image segmentation effectively.

Along with the advance of 3D technology, Microsoft has designed a sensor with multiple cameras to introduce a new way in image segmentation; this sensor is able to detect depth information in the target space of every single pixel along with acquiring the RGB colour information (Wang et al., 2012). The fusion of two kinds of image data can generate a new RGB-D image containing both colours and depth information. Using the depth information in depth image can improve the robustness of gesture segmentation and provide a gesture image segmentation method adapt to various environmental conditions.

## 2 The acquisition of depth image in Kinect

### *2.1 Introduction of Kinect*

Kinect is a peripheral of Xbox, Microsoft's video game console (Luo et al., 2012), for providing a new motion-sensing gaming experience. With more and more developers, the potential function of the device is gradually excavated; at the same time, the Windows version and related developer tools have also been introduced. As shown in Figure 1 is the latest 2.0 edition of Kinect published by Microsoft in 2014, it contains a full HD CMOS sensor and an infrared sensor (Biswas and Basu, 2011), and on top of providing higher definition and wider view angles, it can also remove the ambient light in images through infrared technology, thereby to restore the real structure information of the object.

Compared with the previous generation, Kinect 1.0, the 2.0 version is equipped with higher resolution sensor, which can record videos at 30 frames per second in 1920*1080 full HD resolution (Garcia and Zalevsky, 2008). The 16 bits depth sensor can capture the depth image in the resolution of 512*424 pixels and its theoretical depth detecting range is boosted up to from 0.5 m to 8 m. Besides, the more perfect and useful Kinect for Windows SDK 2.0 developer tools published by Microsoft aims at providing a convenient way to make full advantage of the data generated by Kinect

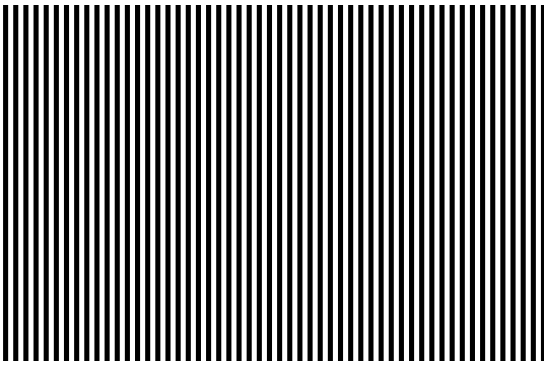
using C++, C# and other programming language (Paris et al., 2006).The libraries inside developer tools provided, such as 3D reconstruct of human gestures, real-time image processing, hand gestures segmentation, etc.

**Figure 1** Kinect 2.0 sensor

## *2.2 Ranging principle based on structured light*

In commercial computer vision sensors, the reasonable production cost, attractive appearance design and low power consumption are as important as the accuracy of the sensor. Thus, light ranging method, as a cheap and high precision ranging principle, was applied in Kinect to obtain the depth image (Zaharescu et al., 2007). In detail, Kinect ranging principle is a practical application of Lightcoding technology, which was patented by the company PrimeSense (Lhuillier and Quan, 2005). Lightcoding is an optical range method; to obtain the distance information, the object space would be marked with a specific pattern of light, which is called Lightcoding. For example a series of parallel white lines, as shown in Figure 2, is a specific pattern of structured light and when it projected to the surface of target object, namely the space which the object situated in has been optical encoded, as shown in Figure 3.

**Figure 2** Structured light pattern

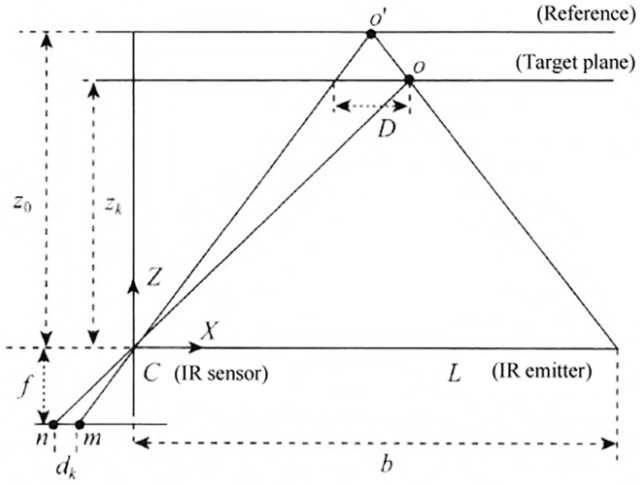


Kinect contains three infrared laser emitters, an infrared receiver-based CMOS technology and a colour CMOS sensor, the depth measurement is based on laser triangulation of distance (Bradley et al., 2008). Laser source emits a beam of infrared laser, then the laser will be divided into multiple leaser light through specific optical grating. After that, the leaser lights are reflected by surface in different depth in the space. Then the infrared CMOS sensor receives the reflected pattern, by comparing the reflected pattern with the reference pattern (Merrell et al., 2007) stored in Kinect's memory to finish the measurement of depth. Finally, the depth information will be represented in a grey-scale image.

**Figure 3** The principle of Lightcoding

As shown in Figure 4, the target $o$ is on the target plane and the distance from the sensor to the target plane $z_k$. The laser pattern reflected by target $o$ is captured by infrared camera; its position on the sensor's film is the point $n$.

**Figure 4** The optical ranging principle of Kinect




When the depth of the target changed from $o$ to $o'$, the position of reflected pattern from $o$ will in turn move from $n$ to $m$ along the $X$-axis on inferred sensor's film plane. According to the similar triangles, the length of corresponding edges can be calculated as follows:

$$\frac{D}{b} = \frac{z_0 - z_k}{z_0} \tag{1}$$

Similarly,

$$\frac{d_k}{f} = \frac{D}{z_k} \tag{2}$$

In the above formula, $z_k$ is the distance from the target plane to the sensor's film plane, which equal to depth value. Then, $b$ is the length of baseline, $f$ is the focal length of infrared camera lens. $D$ is the visual distance of point $k$ on the target plane, $d_k$ is the distance between $m$ and $n$ on the sensor's film plane. From formulas (1) and (2), we can know:

$$z_k = \frac{z_0}{1 + \frac{z_0}{fb} d_k} \tag{3}$$

Since $z_0$ has been defined by the sensor itself, the values of $f$ and $b$ can be obtained through calibration. From formula (3), the depth information can be deduced by the parameters observed.

## 3 Gesture image segmentation based on different colour space

Hand gestures can be real-time detected by Kinect, the RGB images (Tylecek and Sara, 2009) obtaining is the basis of hand gesture image segmentation. In hand gesture recognition system, the most popular gesture segmentation method is human skin colour segmentation and the edge features segmentation. Human skin colour segmentation method is higher in recognition rate; its process is also relatively complicated. In this paper, the segmentation method based on threshold is taken to extract hand gestures (Fuhrmann and Goesele, 2011) and use the colour space transformation to tell the foreground, where hand gestures exist and background apart in a single image.

### *3.1 Gesture image segmentation based on HSV colour space*

The camera film mode of colour CMOS sensor built in Kinect is Bayer array, namely each individual pixel includes one red, one blue and two green subpixels (Peter et al., 2012), as shown in Figure 5. So the output of the pictures is all in RGBG mode, after processing, colour images based on RGB space can be obtained. The RGB colour space model is easy to understand and most widely used in cameras and colour displayers, but its three colour components, R, G and B, have no connection with the colours judgement of human's vision. The correlations between R, G and B components are weak, which make it difficult to do further processing. Therefore, RGB colour space is not always used to process images.

HSV is another common colour space, which was proposed by A.R. Smit in 1978 (Nierstrasz et al., 2002), known as Hexcone Model as well. The parameter H indicates the colour information, that is, the place of the colour in spectrum. This parameter describes the human skin colour effectively, simplifies the gesture segmentation process in the following steps. The first step of the method is to transform the images output by Kinect from RGB colour space to HSV colour space, since the clustering distribution of human skin colour in HSV colour space (Engelhard et al., 2011). In the second step, noise reduction method based on skin colour model is applied in dealing with the original images and then generate a binary image to represent the hand gesture area after segmentation. The specific steps are as follows:

1 The following formula is used to transform the original image $z_0$ from RGB colour space into HSV colour space, the transformed result is $z_1$.

$$V = \max(R, G, B) \tag{4}$$

$$S = \begin{cases} \dfrac{V - \min(R,G,B)}{V} & if\ V \neq 0 \\ 0 & otherwise \end{cases} \tag{5}$$

$$H = \begin{cases} 60(G-B)/(V - \min(R,G,B)) & if\ V = R \\ 120 + 60(G-R)/(V - \min(R,G,B)) & if\ V = G \\ 240 + 60(R-G)/(V - \min(R,G,B)) & if\ V = B \end{cases} \tag{6}$$

By using formula (6), we can obtain the value of component $H$ in image $z_1$, labelled as $z_H$. Then it can be expressed in grey scale, as shown in Figure 6. If $H < 0$, then $H = H + 360$.

**Figure 5** Bayer array in CMOS sensor (see online version for colours)

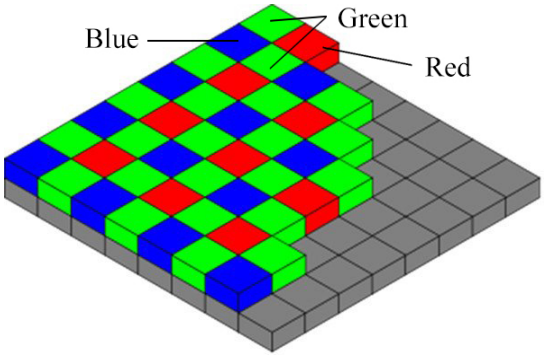
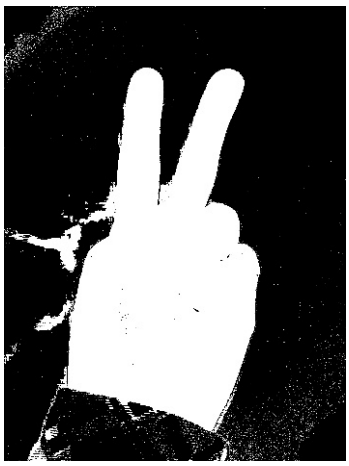



**Figure 6** The transformation from RGB colour space to HSV colour space


2 Set the threshold of $H$ component as $T_0$, then apply this threshold in the value judgement of $z_H$. If any pixel in $z_H$ has the value within the range, the corresponding pixel in a binary image $z_b$ will be assigned as 1, otherwise 0. The range of $T_0$ is usually from 0.03 to 0.128. After noise reduction (Izadi et al., 2011b), the hand gesture area $z$ can be obtained, the binary figure result using the above methods is shown in Figure 7.

**Figure 7** Binary figure results



### *3.2 Gesture image segmentation based on YCbCr colour space*

YCbCr colour space (Izadi et al., 2011a) is another commonly used colour space, compared with the HSV colour space mentioned in advance, YCbCr colour space has a better human skin colour compactness in low saturation range, because the brightness component Y is separated from the chrominance component Cb and Cr; human skin colour distribution in this colour space is less affected by light factors. According to the skin colour feature points, which Anil K. Jain et al. extracted from Heinrich-Hertz-Institute (HHI) image library (Liu et al., 2012), cluster in YCbCr space in the shape of spindle with two pointy ends, that is to say, where component Y is too large or too small, the clustering distribution will shrink. On the one hand, YCbCr colour space is good at restraining the distribution of skin colour, with little environmental impacts, so it is suitable for human gesture segmentation. On the other hand, image in RGB colour format can be directly transformed into YCbCr colour format through linear transformation (Yang et al., 2012); the transformation formula is as follows:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \cdot \begin{bmatrix} 65.738 & 129.057 & 25.06 \\ -37.945 & -74.494 & 112.43 \\ 112.439 & -94.154 & -18.28 \end{bmatrix} \cdot \begin{bmatrix} r \\ g \\ b \end{bmatrix} \quad (7)$$

We can know from the formula above, the luminance component Y is not completely independent of chrominance information, so skin colour clustering distribution is non-linear along with different value of Y. Therefore, skin colour gestures segmentation needs to calculate the probability of each human skin colour pixel, and work out the skin colour likelihood map of the image (Huang, 2012). Then use the adaptive threshold method (Zhang, 2013) to convert the colour likelihood map into a binary image. Specific steps are as follow:

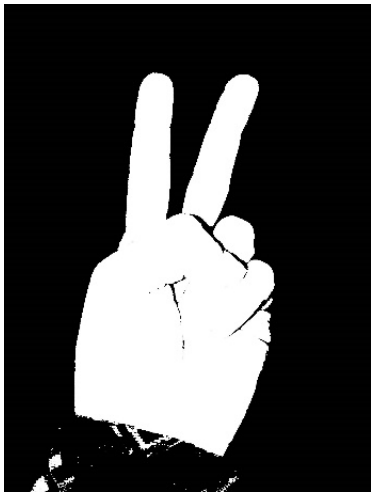
1 Set the initial threshold value, and decrease the value (Chen, 2013) into minimum step by step in equal intervals.

2 Calculate the area of each hand gesture segment after every decreasing step. Then compare the area with the previous stage and take minimum threshold between two adjacent phases as the optimal threshold for expect.

3 Finally, convert the skin colour likelihood probability map (Li, 2012b) into binary image, according to the optimal threshold. Then assign each pixel of the gesture area as 0 and assign background pixel as 1. The final gesture segmentation result is as shown in Figure 8.

Compared with the binary map in the HSV colour space, the processing in YCbCr colour space can obtain smoother edges and eliminate the influence of ambient light in the result. At the same time, the binary map is clearer with less noise (Akio and Koide, 1991) and the availability of gesture segmentation is higher.

**Figure 8** Hand gesture segmentation in YCbCr colour space



## 4 The optimisation of gesture image segmentation using depth information

Depth information is the innovation of Kinect (Gortler et al., 1996); this paper mainly studied the hand detection, which can completely overcome the interference factors such as illumination change and complex background. By using the skeleton model in Kinect API, the hand joint positioning in space can be even faster and more accurate.

### *4.1 Kinect's sensor coordinates system alignment*

Kinect contains two different sensors, they are colour image sensor and infrared sensor, and their data sources include colour RGB image, infrared image (Mitra and Acharya, 2007) and depth image. The depth image is obtained by structure light measurement (Wang et al., 2006), by using the infrared sensor. The positions of each sensors in space are not the same, so the images output by different sensors is not fully reflected the same scene. Therefore, we have to calibrate the two sensors before image acquisition and finish the stereo matching between two cameras (Alon et al., 2009).

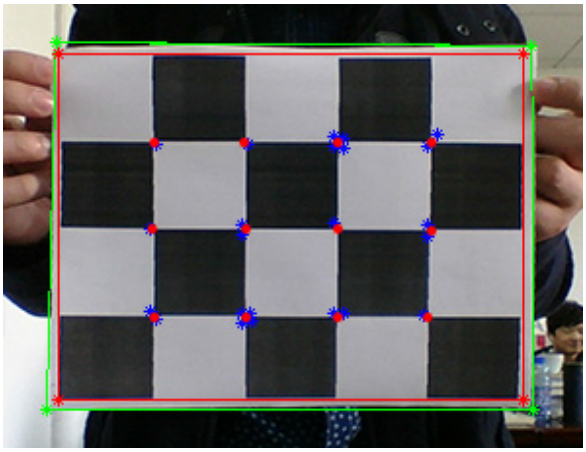
For the convenience of camera calibration, the coordinate system parameters of two sensors and corresponding hand gesture space are shown in Table 1.

**Table 1** Three coordinate system parameters in Kinect

| *Name* | *Dimensions* | *Units* | *Range* | *Origin* |
|---|---|---|---|---|
| ColorSpacePoint | 2 | Pixels | 1920*1080 | Top left corner |
| DepthSpacePoint | 2 | Pixels | 512*424 | Top left corner |
| CameraSpacePoint | 3 | Metres | 0–8 | Infrared/ depth camera |

First of all, we have to calibrate the colour camera with the inferred camera by calibration technology (Miguel et al., 2011). Since the depth camera (inferred camera) cannot distinguish the colour on standard Chessboard Pattern, designing a calibration board with depth information is relatively complex as well. We manually selected the intersection (Zhang et al., 2012) on the calibration plate, as shown in Figure 9.

**Figure 9** Depth camera calibration Chessboard Pattern (see online version for colours)



After the calibration of colour and depth camera has been completed, we need to apply two cameras' internal parameters into the corresponding RGB image and depth image coordinate systems. Then calibrated colour image and depth image can be output (Shotton et al., 2011). On top of that, you need to connect two camera coordinate systems with the corresponding hand gesture space, namely calibrate the depth camera and colour camera in three dimensions. After the corresponding rotation matrix and transfer matrix are obtained, the relationship between depth information and RGB colour image can be established correspondingly (Bhanu and Zhou, 2004). The specific steps are as follows:

1 Map the depth information into the three-dimensional coordinate system of human hand gesture, according to the parameters shown in Table 1. The right-handed coordinate system is used in gesture space (Lichtenauer et al., 2008); the pixel $(x_d, y_d)$ in depth image coordinate and the point $(x, y, z)$ in gesture coordinate have the following relationships:

$$\begin{aligned} x &= \frac{d(x_d - a_3)}{a_1} \\ y &= \frac{d(x_d - a_3)}{a_1} \\ z &= d \end{aligned} \tag{8}$$

Among them, $a_1$, $a_2$, $a_3$, and $a_4$ are the intrinsic parameters of depth camera, $d$ is the depth value of pixel $(x_d, y_d)$.

2 Mark a certain pixel in gesture coordinate as $P = [x\, y\, z]^T$, and then map it to the RGB colour image, thus three-dimensional coordinates $P = [x\, y\, z]^T$ have the following relations with the colour one:

$$\begin{aligned} P' &= RP + T \\ x_c &= \frac{x' b_1}{z'} + b_2 \\ y_c &= \frac{y' b_3}{z'} + b_4 \end{aligned} \tag{9}$$

Above $R$ is a three-order rotation matrix, $T$ is a 3*1 translation matrix. These two parameters $T$ and $R$ can be obtained by calibration. $P' = [x'\, y'\, z']^T$ is a temporary variable, $b_1$, $b_2$, $b_3$ and $b_4$ are the intrinsic parameters of colour CMOS camera (Claus and Burkhardt, 2004).

### *4.2 Gesture segmentation based on depth image*

The depth information of hand gesture has been already contained in the depth image obtained by Kinect, as well as the background depth information. Through these data, a sheet 3D model (Bartolini et al., 2005) of hand gesture can be constructed. The goal of the palm area segmentation is to extract the palm area from the background in depth image, and map the extracted hand area into a two-dimensional space. The hand position coordinates can be easily got by using the function of bone and joint point detection, which is built in the Kinect SDK.

After getting hand joint coordinates, the central point $o$ of palms can be calculated. Gestures extraction regards $o$ as the geometric centre of extraction rectangle $A$. Set the threshold $T$ of depth as 1000, which is the unit that is generally used inside of Kinect. Then take $g(x, y)$ as the depth difference value of one point $(x, y)$ with palm centre point o in area $A$. Finally, establish a binary image $f(x, y)$ in the two-dimensional space in a corresponding size with the rectangle area $A$, and use the following formula can achieve the mapping from depth image $g(x, y)$ to binary figure $f(x, y)$.

$$f(x,y) = \begin{cases} 1 & g(x,y) \le T \\ 0 & g(x,y) > T \end{cases} \tag{10}$$

The binary map can be obtained by formula above. After relevant operation between binary map and the original RGB image in calibrated coordinates, gesture segmentation can be got; the result is shown in Figure 10.

**Figure 10** Threshold segmentation based on depth image

## 5 Conclusion and future work

Gesture is one of the most intuitive ways in human and machine communication; to complete the human–computer interface based on hand gestures, the most important part is to make the machine 'understand' the meaning of human gestures. Therefore, segmentation and extraction of human hand gestures by using computer graphic technology is the key step for understanding the meaning of the human gestures. This paper introduced the hardware structure of the second generation Kinect, and illustrated the imaging principle of the structure light measurement of depth sensor. This method mainly took two-dimensional RGB colour and grey-scale depth image as the data sources. We also compared the result of the skin colour segmentation strategies based on HSV and YCbCr colour space, applied the unique depth information output of Kinect to help segment the gestures region in space. It is evident that all sorts of gesture segmentation methods have advantages and disadvantages compared with each other, while the segmentation method based on depth information is easier in applications, highly interference-free and greatly simplifies the algorithm in image segmentation. In the near future, human hand gesture segmentation will be further developed in the direction of multi-sensor fusion, especially the fusion of 2D colour images and 3D depth images, in turn, increases the robustness of image segmentation methods.

## Acknowledgements

This work was supported by grants of National Natural Science Foundation of China (Grant Nos. 51575407, 51575338, 51575412, 61273106).

## References

Akio, D. and Koide, A. (1991) 'An efficient method of triangulating equi-valued surfaces by using tetrahedral cells', *IEICE Transactions on Information and Systems*, Vol. 74, No. 1, pp.214–224.

Alon, J., Athitsos, V. and Yuan, Q. (2009) 'A unified framework for gesture recognition and spatiotemporal gesture segmentation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 9, pp.1685–1699.

Bartolini, I., Ciaccia, P. and Patella, M. (2005) 'Warp: accurate retrieval of shapes using phase of Fourier descriptors and time warping distance', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 1, pp.142–147.

Bhanu, B. and Zhou, X. (2004) 'Face recognition from face profile using dynamic time warping', *International Journal of Pharmaceutics*, Vol. 4, Nos. 1–2, pp.87–92.

Biswas, K.K. and Basu, S.K. (2011) 'Gesture recognition using Microsoft Kinect', *IEEE International Conference on Automation, Robotics and Applications (ICARA)*, IEEE, Wellington, New Zealand, pp.100–103.

Bradley, D., Boubekeur, T. and Heidrich, W. (2008) 'Accurate multiview reconstruction using robust binocular stereo and surface meshing', *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Anchorage, Alaska, pp.1–8.

Chen, X. (2013) *Research of 3D Reconstruction and Filtering Algorithm based on Depth Information of Kinect*, Shanghai Jiao Tong University, Shanghai, pp.53–54.

Claus, B. and Burkhardt, H. (2004) 'The writer independent online handwriting recognition system frog on hand and cluster generative statistical dynamic time warping', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 3, pp.299–310.

Engelhard, N., Endres, F. and Hess, J. (2011) 'Real-time 3D visual SLAM with a hand-held RGB-D camera', *24th Annual ACM Symposium on User Interface Software and Technology*, ACM, Vasteras, Sweden, pp.559–568.

Fuhrmann, S. and Goesele, M. (2011) 'Fusion of depth maps with multiple scales', *ACM Transactions on Graphics*, Vol. 30, No. 6, pp.61–64.

Garcia, J. and Zalevsky, Z. (2008) *Range Mapping Using Speckle Decorrelation*, United States Patent, US 7433024 B2. Available online at: http://www.google.co.in/patents/US7433024

Gortler, S.J., Grzeszczuk, R. and Szeliski, R. (1996) 'The lumigraph', *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, IEEE, New Orleans, LA, USA, pp.43–54.

Huang, Y. (2012) *Vision Based Mapping of a Mobile Robot in Unknown Environment*, Nanjing University, Nanjing, pp.56–57.

Izadi, S., Kim, D. and Hilliges, O. (2011a) 'KinectFusion: realtime 3D reconstruction and interaction using a moving depth camera', *24th Annual ACM Symposium on User Interface Software and Technology*, Santa Barbara, CA, USA, pp.559–568.

Izadi, S., Newcomber, A. and Kim, D. (2011b) 'KinectFusion: real-time dynamic 3D surface reconstruction and interaction', *ACM SIGGRAPH 2011 Talk*, IEEE, Vancouver, pp.449–458.

Lhuillier, M. and Quan, L. (2005) 'A quasi-dense approach to surface reconstruction from uncalibrated images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 3, pp.418–433.

Li, G. (2012a) *Research and Implementation of Kinect based 3D Reconstruction*, Beijing Jiaotong University, Beijing, p.47.

Li, Y. (2012b) 'Hand gesture recognition using Kinect', *IEEE International Conference on Software Engineering and Service Science (ICSESS)*, IEEE, Beijing, pp.196–199.

Lichtenauer, J.F., Hendriks, E.A. and Reinders, M.J. (2008) 'Sign language recognition by combining statistical DTW and independent classification', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 11, pp.2040–2046.

Liu, X., Xu, H. and Hu, Z. (2012) 'GPU based fast 3D-object modeling with Kinect', *Acta Automatica Sinica*, Vol. 38, No. 8, pp.1288–1297.

Luo, Y., Xie, Y. and Zhang, Y. (2012) 'Design and implementation of a gesture-driven system for intelligent wheelchairs based on the Kinect sensor', *Robot*, Vol. 34, No. 1, pp.110–114.

Merrell, P., Akbarzadeh, A. and Wang, L. (2007) 'Real-time visibility-based fusion of depth maps', *IEEE 11th International Conference on Computer Vision*, IEEE, Brazil, pp.1–8.

Miguel, R., Gabriel, D. and Sergio, E. (2011) 'Feature weighting in dynamic time warping for gesture recognition in depth data', *IEEE International Conference on Computer Vision Workshops*, IEEE, Barcelona, pp.1182–1188.

Mitra, S. and Acharya, T. (2007) 'Gesture recognition: a survey', *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, Vol. 37, No. 3, pp.311–324.

Nierstrasz, O., Artvalo, G. and Ducasse, S. (2002) 'A component model for field devices', *Lecture Notes in Computer Science*, Vol. 2370, pp.200–209.

Paris, S., Sillion, F.X. and Quan, L. (2006) 'A surface reconstruction method using global graph cut optimization', *International Journal of Computer Vision*, Vol. 66, No. 2, pp.141–161.

Peter, H., Michael, K. and Evan, H. (2012) 'RGB-D mapping: using Kinect-style depth cameras for dense 3D modeling of indoor environments', *International Journal of Robotics Research*, Vol. 31, No. 5, pp.647–663.

Shotton, J., Fitzgibbon, A. and Cook, M. (2011) 'Real-time human pose recognition in parts from single depth images', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, ACM, New York, pp.1297–1304.

Tylecek, R. and Sara, R. (2009) 'Depth map fusion with camera position refinement', *Computer Vision Winter Workshop*, Vienna University of Technology, Austria, pp.59–66.

Wan, T., Wang, Y. and Li, J. (2012) 'Hand gesture recognition system using depth data', *IEEE 2012 2nd International Conference on Consumer Electronics Communications and Networks (CECNet)*, 21–23 April, IEEE, Yichang, China, pp.1063–1066.

Wang, S.B., Quattoni, A. and Morency, L.P. (2006) 'Hidden conditional random fields for gesture recognition', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Washington, DC, pp.1521–1527.

Wang, Y., Yang, C. and Wu, X. (2012) 'Kinect based dynamic hand gesture recognition algorithm research', *IEEE International Conference on Intelligent Human–Machine Systems and Cybernetics (IHMSC)*, IEEE, Nanchang, pp.274–279.

Yang, D., Wang, S. and Liu, H. (2012) 'Scene modeling and autonomous navigation for robots based on Kinect system', *Robot*, Vol. 34, No. 5, pp.581–586.

Zaharescu, A., Boyer, E. and Horaud, R. (2007) 'TransforMesh: a topology-adaptive mesh-based approach to surface evolution', *Asian Conference on Computer Vision*, Vol. 4844, pp.166–175.

Zhang, C. (2013) *3D Indoor Scene Reconstruction with Kinect Depth Camera*, Dalian University of Technology, Dalian, p.48.

Zhang, Y., Zhang, S., Luo, Y. and Xu, X. (2012) 'Gesture track recognition based on Kinect depth image information and its applications', *Application Research of Computers*, Vol. 29, No. 9, pp.3547–3550.