http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (14): 2963-2968, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.2963.2968

An Improved Hu-moment Algorithm in Gesture Recognition Based on Kinect Sensor

[1]Li Guangsong, [2]Ou Qiaoxin and [1]Luo Jiehong
[1]Department of Information Engineering, Guangdong Polytechnic, Foshan, 528041, Guangdong, People's Republic of China
[2]Synthesis department, Guangdong power gird corporation Foshan power supply bureau, Foshan, 528000, Guangdong, People's Republic of China

Abstract: In order to solve the problem of variability of gesture recognition, The Kinect sensor and the improved Hu-moment is used in this study. Depth image, after image restoration, its feature data is extracted using depth threshold is obtained based on Kinect. Then the scale factor is added to the Hu-moment, so that the invariant moments contain more detail characteristics of gesture and have relationship with geometry but not scaling. For illustration, a dress fitting system example is utilized to interpret the algorithm. Empirical results show that the gesture recognition method is performed correctly in illumination and complex background. So the method of gesture recognition has invariance in image rotation, scaling and translation and has robustness to background interference.

Key words: Hu-moment, gesture recognition, depth image, scale factor

INTRODUCTION

The depth image ,which refers to an image or an image channel, is also known as the distance image, image an image or an image channel and related to information contained in the image and surface distance in the scene, looking from the perspective. In recent years, the depth image is used in the technology of pattern recognition, which is mainly due to lower the cost of the depth map camera. The research status of depth image in the field of pattern recognition and application of human recognition (Li *et al*., 2012) introduces a method of integrating point features and gradient features to identify human body. Although this method can reduce the influence of image in the illumination, posture, shelter and other factors in a certain extent but the accuracy and recognition rate needs to be improved and it can not meet the real-time requirement. The method of calculating the features derived from a similarity of depth histograms (Ikemura and Fujiyoshi, 2011) achieves a detection rate of 95.3% with a false positive rate of 1.0% and can run in real-time but because the depth map from TOF, of only 176×144 resolutions comparing with Kinect of 640×480, so the recognition accuracy is not enough. Using Kinect depth information to reconstruct models (Song *et al*., 2012) has used Kinect's raw depth map, because the depth image of Kinect has large range of missing-values due to occlusion, so the reconstruction is of some defects when under magnification. Hand gesture recognition (Li, 2012) using finger detection can recognize well but is no consideration of scaling. With respect to the depth extraction equipment TOF (Kang and Ho, 2011), Kinect has lower price and can extract depth map with high resolution.

Using the spatial relationship between image objects to recognize image accords with people's habits and the features of spatial relationship is easy to combine with other visual features of image and to realize the integrated retrieval of fusion in multi visual feature and the fusion of retrieval image method is the future direction (Hua and Xia, 2011). In the simulated images, Hu-moment has Characteristics of invariance in the image translation, rotation, scaling but Hu moment does not have scale invariance in digital image, so it is not suitable for the analysis in the application of digital image (Liu *et al*., 2008). At present, most of the Hu-moment transformation mainly concentrated in the combined moment but actually mathematical theory in combined moments is of certain problems since the two formulas $x' = kx$ and $y' = ky$ is not established in digital image scaling, therefore, the use of the derived combined moment from the two equations is not effective if improving the scaling invariance of digital image (Chen, 2011). The method to use the ratio between the distances of moment to remove the scale factor based on Hu-moment (Li *et al*., 2012), although can achieve invariance but this method generates a maximum error relatively large.

Corresponding Author: Li Guangsong, Department of Information Engineering, Guangdong Polytechnic, Foshan, 528041, Guangdong, People's Republic of China Tel: 0757-82862386

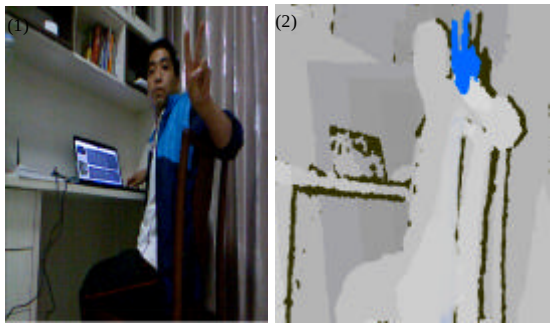

Fig. 1(1-2): Kinect color image (1) and (2) Depth image

An algorithm to repair a Kinect raw depth map, combining the time background fitting-information and spatial information (Wang *et al*., 2012) has achieved good results, although time has been significantly improved but still can not meet the requirement of real-time fitting system. Threshold segmentation method is simple in calculation, high efficiency and fast. Global threshold is effective segmentation for different target of difference gray from background. When the difference gray of image is not obvious or the values of different target overlap, the local threshold or dynamic threshold segmentation method should be used.

This study introduces a scale factor in the Hu-moment, which can effectively reduce maximum error and has good robustness to the background interference, also it has small amount of computation and can satisfy the real-time requirement.

THE DEPTH IMAGE ACQUISITION

The main raw output of Kinect is an image that corresponds to the depth in the scene. Rather than providing the actual depth z, Kinect returns "inverse depth"d (Smisek *et al*., 2011).

When we interact with Kinect, the streamed data can be getting in real time. Using Kinect, through adjusting the angle, we can quickly gain access to the RGB color image and depth map of the scene, as shown in Fig. 1.

DEPTH IMAGE RESTORATION

In the 3D scene, since the background keeps out of foreground and individual part of the foreground is close to Kinect, which cause the part is out of the infrared square, part of the Kinect emission of infrared light cannot be reflected back and lead to the background or foreground cavity.

Ordinary image restoration cannot distinguish between foreground and background image and using simple weighted summation method to fill the blank of foreground and background will lead hybrid. After introducing the depth map, the depth map can be used to distinguish the foreground and background of image. When repairing the background, the depth map can be used to filtering those prospects reference points, so that the blank points of background only are got by the background weighted sum and vice versa.

The restoration formula of the depth image is:

$$f(x,y)=\frac{\sum_{(m,n)\in R}(D(m,n)*W(m,n)*I(m,n))}{\sum_{(m,n)\in R}(D(m,n)*W(m,n))} \quad (1)$$

In which, D (m, n) is the depth weighted factor:

$$D(m,n)=\begin{cases}0, \text{depth}(m,n)-\text{depth}(x,y)>Td,\\ 1, \text{otherwise}\end{cases} \quad (2)$$

The steps of restoration algorithm are shown as the following:

- For each point in depth map
- Get the base gray value of depth map
- Get the relative x coordinate of base point
- Get the relative y coordinate of base point
- Calculate the weight,which is the reciprocal of square of the distance
- Deal with the red channel
- Deal with the green channel
- Deal with the blue channel
- If the point is the last one, then finish. Otherwise go back to Step 2

DEPTH THRESHOLD

Threshold segmentation method has two main steps. The first step is to determine the segmentation threshold values needed. The second step is to divide the pixels by comparison the threshold segmentation values and pixel values. In this study, the selected threshold objects, different from traditional threshold method, is not gray but the depth value of model. This method is so called depth threshold method. According to the depth level, the image depth threshold is designed to divide the set of pixels into different subset get each subset and each subset forms a region corresponding to a realistic scene and each has consistent properties, while the adjacent area does not have this same attribute. We can select one or more threshold from the image depth level to achieve and specific algorithm is shown as Eq. 3 and 4:

$$F(\bar{X},\bar{Y}) = \varphi(\mu_X < \varphi < \sigma_X \text{and} \mu_X < \sigma_X) \quad (3)$$

$$F(\bar{X},\bar{Y}) = \varphi(\varphi \le \mu_X \text{or} \varphi \ge \sigma_X) \quad (4)$$

Given the image $F(\bar{X},\bar{Y})$, for each input image depth value, we determine two depth values (i.e., threshold), μ_X and σ_X. When the pixel depth value φ is greater than φ or less than μ_X, then $F(\bar{X},\bar{Y}) = 0$, else F (x, y) = D.

IMPROVED HU-MOMENT

Concept of the Moment: For the continuous grey function f (x, y), its two-dimensional Order origin moment M_{pq} is defined as:

$$M_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^p y^q f(x,y)dxdy \quad p,q = 0,1,2\ldots \quad (5)$$

The focus of the f (x, y) image is determined by the two first-order moment: M_{01} and M_{10}. The barycentric coordinate is (x′, y′), in which, x′ and y′ are defined as:

$$x' = \frac{M_{10}}{M_{00}}, y' = \frac{M_{01}}{M_{00}} \quad (6)$$

If the focus of the object coincides with the origin of the coordinate system of coincidence, i.e., x′ = 0, y′ = 0.

Suppose that f (x, y) is piecewise and continuous bounded function and its value is a non-zero value in the finite region of X-Y plane. According to the uniqueness theorem, if the continuous image of a two-dimensional function f (x, y) is piecewise and continuous, i.e. as long as there is a non-zero value in a limited region of X-Y plane, then the moments all exist and the moment sequences $\{M_{pq}\}$ are determined by f (x, y) only, on the other hand, $\{M_{pq}\}$ also uniquely determines f (x, y).

In addition, it can also define the (p+q) order central moments of f (x, y) as:

$$\mu_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y)dxdy \quad (7)$$

Hu-moment: Hu (1962) used the theory of algebraic invariants for normalized central moments, in 1962 and constructed the following seven invariant moments of translation, rotation and scale. Using Hu invariant moments is an important method of image recognition and image matching and Hu invariants can be expressed by the following 7 formula Hu (1962):

$$M1 = \mu_{20} + \mu_{02} \quad (8)$$

$$M2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \quad (9)$$

$$M3 = (\mu_{30} - \mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \quad (10)$$

$$M4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} - \mu_{03})^2 \quad (11)$$

$$M5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} - \mu_{03})^2] + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \quad (12)$$

$$M6 = (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03})^2 \quad (13)$$

$$M7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{12} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \quad (14)$$

The normalized central moment is defined for:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^r} \quad (15)$$

In which:

$$r = \frac{p+q}{2} + 1, p + q = 2,3,\ldots$$

In the seven moment invariants, the first six have invariance of mirror image. They are suitable for the description of overall target shape, so it has wide application in edge extraction, image matching and object recognition. The calculation of these seven Hu-moments is different and the information content is not the same. The useful information of image generally concentrated in low order moments, which are relatively small computation.

Improved Hu-moment: Through calculation of Hu-moment, it is different in the powers number of times between order moments and the parallel image pixel coordinates, so the dimension of the different order moments obtained in the scale is associated with the order of the moments.

Supposing that $(\bar{x},\bar{y})$ is the value of coordinate after introducing the scaling factor κ (κ>0) into Hu-moment, then, compared with non-introducing scaling factor, its coordinate satisfies the following relations:

$$\bar{x} = \kappa x, \bar{y} = \kappa y$$

The formula is derived as following:

$$\bar{x}' = \frac{M_{10}}{M_{00}} = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\bar{x}' f(x,y)dxdy}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy} = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\kappa x' f(x,y)dxdy}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy} = \kappa x' \tag{16}$$

And the following formula can be got:

$$\bar{x} - \bar{x}' = \bar{x} - \kappa x' = \kappa x - \kappa x' = \kappa(x - x') \tag{17}$$

Similarly the following formula can be got:

$$\bar{y} - \bar{y}' = \kappa(y - y') \tag{18}$$

If put $\bar{x} - \bar{x}'$ and $\bar{y} - \bar{y}'$ into:

$$\mu_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\bar{x})^p(y-\bar{y})^q f(x,y)dxdy \tag{19}$$

The following can be got:

$$\begin{aligned}\bar{\mu}_{pq} &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(\bar{x}-\bar{x}')^p(\bar{y}-\bar{y}')^q f(\bar{x},\bar{y})dxdy \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\kappa^p(x-x')^p\kappa^q(y-y')^q f(x,y)dxdy \\ &= \kappa^{(p+q)}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-x')^p(y-y')^q f(x,y)dxdy \\ &= \kappa^{(p+q)}\mu_{pq}\end{aligned} \tag{20}$$

If put $\bar{\mu}_{pq}$ into:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^r}$$

we can get:

$$\bar{\eta}_{pq} = \frac{\bar{\mu}_{pq}}{\bar{\mu}_{00}^r} = \frac{\kappa^{(p+q)}\mu_{pq}}{\bar{\mu}_{00}^r} = \kappa^{(p+q)}\frac{\mu_{pq}}{\mu_{00}^r} = \kappa^{(p+q)}\eta_{pq} \tag{21}$$

It can be seen from the above equation, effect of scale factor on the central moment, is not only related with scaling factor κ but also related with the order p+q of moments, so when the scaling factor is existing, according to the ratio changes relationship between the normalized central moment, i.e., η_{pq} and $\bar{\eta}_{pq}$ before and after.

Putting the new normalized central moment into the seven Hu invariant moments formula, namely:

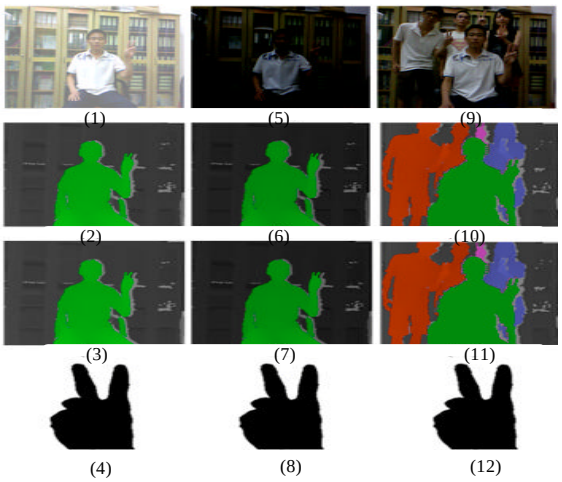

Fig. 2(1-12): Robustness verification

$$\bar{M1}=\kappa^2 M1, \bar{M2}=\kappa^4 M2, \bar{M3}=\kappa^6 M3, \bar{M4}=\kappa^6 M4$$
$$\bar{M5}=\kappa^{12} M5, \bar{M6}=\kappa^8 M6, \bar{M7}=\kappa^{12} M7$$

ANALYSIS OF EXPERIMENTAL RESULTS

Experimental environment: To test the validity of the algorithm presented in this study, we developed the program by using Microsoft visual studio 2010 C++ language with the computer environment of Intel® Core™ i5-2450M CPU@2.50GHZ, 4.00GB memory, Windows 7. We get depth image data from view shot by Kienct cameras with random position in accord with the principle of distributing sparsely in space. After image restoration using matlab, we input these depth data to test recognizing program and output gesture segmentation images. The results are shown in Fig. 2: 1, 5 and 9 are the original color images shot at Kinect RGB camera, 2, 6 and 10 are the images of depth shot at Kinect depth camera; 3, 7 and 11 are the images of image restoration; 4, 8 and 12 are the images of segmentation.

Robustness verification: As shown in Fig. 2, 1 ~ 4 are the gesture segmentation effects of the glare conditions and 5 ~ 8 are the gesture segmentation effects of the low light conditions gesture segmentation effects and 9 ~ 12 are the gesture segmentation effects of background interference with a lot of people. From the graphs as can be seen, the gesture segmentation method to the change of illumination and background interference has good robustness.

Comparing the Fig. 2 (1) to Fig. 4 (5) and Fig. 2 (2) to (6), it shows that the depth image of Kinect is not

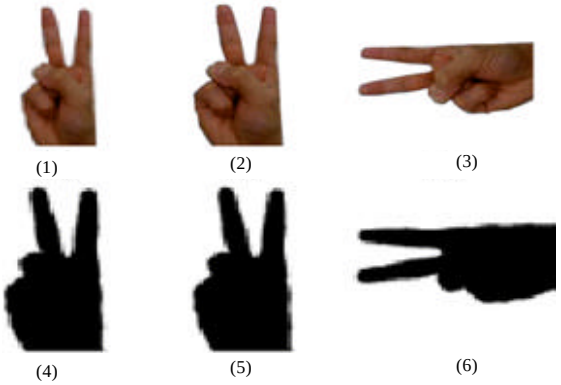

Fig. 3(1-6): Triple invariant patterns

influenced by the effects of light. From Fig. 2 (9), it shows the recognition method has anti jamming. As shown in Fig. 2 (4), (8) and (12), the gesture recognition method is performed correctly.

Triple Invariant Patterns: As shown in Fig. 3 (1) ~ (3) are the gesture of real-time acquisition from the background and Fig. 3 (4) ~ (6) are the marked gesture templates. Comparing the Fig. 3 (1) to (4), it shows that when the hand location of real-time acquisition to the Kinect is far from template acquisition position, the gesture recognition system can complete the identification. Comparing the Fig. 3 (2) to (5), it shows that when the hand level position of real-time acquisition is changed, the gesture recognition system can also complete the identification. Similarly, from (3) and (6) can be seen, when the hand of real-time acquisition is rotated 90 degrees, the gesture recognition system can also complete the identification. From the three groups of pictures above, it can be seen, the gesture recognition system has image scaling, translation, rotation of triple invariant patterns.

Real-time validation: In order to verify the gesture recognition algorithm effective and real-time, the somatosensory fitting system is used. The dress fitting system flow chart and effect image are shown in Fig. 4. In the testing, every gesture recognition from the start to end of dress fitting costs 0.18 sec average. The experimental results show that the costing time is in our range of tolerance and the algorithm can meet the real-time requirement.

Reduce maximum error: In order to verify the effectiveness of improved algorithms, the statistics of the two algorithms error range, are shown in the following Table 1.

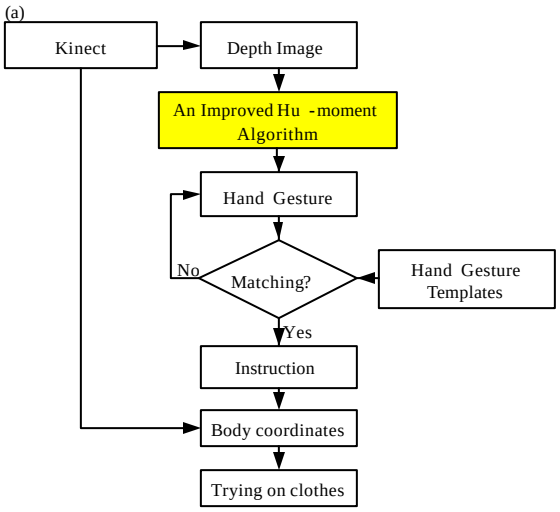

Fig. 4(1-2): Dress fitting system flow chart(1) and effect image(2)

Table 1: Error range statistics

	Error range statistics						
Algorithm	1	2	3	4	5	6	7
Remove scale factor	0.08327	0.00099	0.06528	0.03364	0.05845	0.06358	0.00091
Introduce scale factor	0.00035	0.00006	0.00023	0.00041	0.00017	0.00021	0.00012

From the Table 1, it can be found, the two improved algorithms all can greatly reduce the gap between the seven moments and meet the scaling invariance. The maximum error range of the first algorithm is between 0.00091 and 0.08327 but the maximum error range of second algorithm is between 0.00006 and 0.00041, so the second kind of improved algorithm has less maximum error than the first one, that is to say, the second algorithm is more effective in scaling invariance.

CONCLUSION

The Kinect camera has been a great success in gaming but more importantly it has opened up new fields for researchers. The data streamed from the Kinect in real time requires new processing algorithms. By developing a novel Hu-moment improving algorithm, it can be shown feasibility for using the Kinect in gesture recognition applications such as fitting dress system.

The gesture segmentation of depth image by improved Hu-moment can effectively avoid monitoring gesture interference of light and complex background, at the same time, because of the target recognition algorithm based on improved Hu-moment having translation invariant, rotation invariant and scale invariant characteristic, therefore the operator can conveniently control dress fitting system.

At the same time, through improving the Hu-moment algorithm by inducing a scale factor and it make it less of maximum error than remove a scale factor in improving in Hu-moment algorithm.

ACKNOWLEDGMENT

This study is supported by the Guangdong Province Education Information Technology Special Topic (12JXN048). I would like to thank all the students of Game112 in Guangdong Polytechnic and especially acknowledge the contributions made by Prof. Zhang Jian and as well as contributions made by my students.

REFERENCES

Chen, R.S., 2011. The scale invariability analysis of Hu moments for digital image recognition (in Chinese). Microcomput. ITS Appli., 30: 29-31.

Hu, M.K., 1962. Visual pattern recognition by moment invariants. IRE Trans. Inform. Theo., 8: 179-187.

Hua, B. and L.N. Xia, 2011. Sign language recognition based on Hu invariant moments and euclidean distance (In Chinese). Comput. Eng. Des., 32: 615-618.

Ikemura, S. and H. Fujiyoshi, 2011. Real-Time Human Detection Using Relational Depth Similarity Features. In: Computer Vision-ACCV 2010, 10th Asian Conference on Computer Vision, Queenstown, New Zealand, November 8-12, 2010, Kimmel, R., R. Klette and A. Sugimoto (Eds.). Vol. 6495, Springer Berlin Heidelberg, USA., ISBN: 13-978-3-642-19281-4, pp: 25-38.

Kang, Y.S. and Y.S. Ho, 2011. Disparity map generation for color image using TOF depth camera. Proceedings of the 3DTV-Conference: The True Vision-Capture, Transmission and display of 3D video, May 16-18, 2011, Antalya, Turkey, pp: 1-4.

Li, G.S., J.H. Luo and J. Zhang, 2012. Design and realization of somatosensory dress fitting system based on kinect (In Chinese). Tsinghua Sci. Technol., 3: 48-50.

Li, H., L. Ding and G. Ran, 2012. Analysis of human identification based on kinect depth image. Digital Commun., 4: 21-26.

Li, Y., 2012. Hand gesture recognition using kinect. Proceedings of 3rd International Conference on Software Engineering and Service Science, June 22-24, 2012, Beijing, pp: 196-199.

Liu, J., Y. Liu and C. Yan, 2008. Feature extraction technique based on the perceptive invariability. Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery, October 18-20, 2008, Shandong, pp: 551-554.

Smisek, J., M. Jancosek and T. Pajdla, 2011. 3D with Kinect. Proceedings of the IEEE International Conference on Computer Vision Workshops, November 6-13, 2011, Barcelona, Spain, pp: 1154-1160.

Song, S.C., S.P. Yu and W.J. Xu, 2012. Study on 3D body scanning,reconstruction and measurement techniques based on Kinect (In Chinese). J. Tianjin Polytech. Univ., 31: 34-41.

Wang, K., P. An, Y. Zhang, H. Cheng and Z.Y. Zhang, 2012. Real-time depth extraction and multi-view rendering algorithm based on Kinect (in Chinese). J. Optoelect. Laser, 23: 1949-1956.