XZ-Shape Histogram for Human-Object Interaction Activity Recognition based on Kinect-like Depth Image

M.A.AS'ARI[1], U.U.SHEIKH[2], E.SUPRIYANTO[1]
[1]Department Faculty of Biosciences and Medical Engineering
[2]Computer Vision, Video and Image Processing Research Group (CvviP), Faculty of Electrical Engineering
Universiti Teknologi Malaysia, Johor Bahru
Malaysia
amir-asari@biomedical.utm.my, usman@fke.utm.my, eko@biomedical.utm.my

Abstract: - This paper introduces XZ-shape histogram in recognizing human performing activities of daily living (ADLs) which focuses on human-object interaction activities based on Kinect-like depth image. The evaluation framework was formulated in order to compare XZ-descriptor with previous shape histogram as well as X-shape histogram and Z-shape histogram. Each descriptor was segmented into several cases according to number of shells and symbols used in vector quantization process which was executed using our own dataset called RGBD-HOI. This study showed that XZ-shape histogram managed to outperform the other 3D shape descriptors along with the excellent one that compares the performance inferred by the area under receiver operating characteristic curve (AUC-ROC).The results of this study not only demonstrate the implementation of 3D shape descriptor in the dynamic of human activity recognition but also challenge the previous shape histograms in terms of providing low dimension descriptor that capable in improving the discrimination power of human-object interaction activity recognition.

Key-Words: - human-object interaction, activities of daily living (ADLs), RGBD image, shape histogram Kinect-like depth image

1 Introduction

Monitoring activities of daily living (ADLs) plays a major part in assessing the health status of a person suffering with either cognitive [1] or physical impairment[2, 3] which is commonly done by human caregiver or healthcare practitioner. Recently, there are many investigations emerged on developing automated system for monitoring the activities of daily living (ADLs) which can be divided into vision-based and non-vision based system[4]. However, a rapid growing of the vision-based ADLs monitoring system development in this few years has promised the practicality of this sensing modality [4] over the non-vision based ADLs monitoring system: 1) manage to track and sense gross and fine human movements that represent ADLs; 2) provide rich of information such as spatial information, patient characteristics and anomaly actions obtained using a single vision-based sensing agent; 3) easily set up according to the conditions and environments; and 4) has high user or patient acceptance due to the non-invasive modality.

Human ADLs recognition has been investigated widely within the computer vision community [5-7]. However, most of the previous studies emphasized more on the activities without the manipulation of objects such as walking, running and jumping; which are out of healthcare community's interest. This is because the community of healthcare focuses on monitoring the home and indoor ADLs such as drinking, reading or answering a phone which are categorized into the activities that involve object manipulation. However, with the introduction of Microsoft Kinect,, there are several studies appeared proposing the human activity recognition [8-10] that includes the activities of human performing ADLs with specific to object manipulation [11, 12].

This present study involves with the ADLs recognition that focuses on object manipulation activities or human-object interaction based on Kinect-like depth image; since most of the previous studies used RGB based camera or video in recognizing human performing activities of daily living (ADLs) [5-7]. This study also introduces the XZ-shape histogram as motivated from shape histogram [13] as one of the common 3D shape descriptors which used for 3D object retrieval based on the object mesh surface. The implementation of XZ-shape histogram in recognizing human-object interaction activity was done because only little contributions were found in extracting 3D shape descriptor for the dynamic of human activity[14]; especially when the 3D surface is in a form of 3D point cloud which is obtained from Kinect-like depth image.

Initially, shape histogram is a histogram with a total number of 3D points resides in each defined shell. The defined shell is in sphere form within the 3D object surface with respect to the centroid. In this paper, we propose two forms of shell model; 1) shell in a form of plane with surface normal is in x-axis direction; and 2) shell in a form of plane with surface normal is in z-axis direction. Shape histogram from both shell models is concatenated to generate XZ-shape histogram. The developed histogram outperformed the previous shape histograms in recognizing human-object interaction activities based on Kinect-like depth image.

The paper is organized as follows. We review the existing approach in vision based human activity recognition and 3D shape descriptor in next section. After that, the shape histogram and our proposed XZ-shape histogram are presented in Section 3 before the evaluation framework is discussed in Section 4. The evaluation result is explained in Section 5 before we discuss and conclude in Section 6.

2 Related works on Human Activity Recognition and 3D Shape Descriptor

In this section, the existing vision-based human activity recognition is discussed as an overview of the current approach in extracting meaningful feature as well as general 3D shape descriptor that has been used for 3D object retrieval.

2.1 Human Activity Recognition

Human activity recognition has been widely investigated by the computer vision community. In general, the proposed approaches can be categorized into three level; 1) low-level; 2) middle-level; and 3) high-level [4] in parallel with the three levels of taxonomy activity. Since the previous studies were based on the RGB camera or video, there were several approaches developed according to the color or RGB image in extracting meaningful information to infer the human activity. The approaches has been reviewed in our previous study [4].

However, with the introduction of Kinect to the research community [15], many studies found exploring different perspectives like the depth information or combination between color (RGB) and depth information for object classification [16-18], human detection [19], automated sign language interpretation [20, 21] and human activity recognition [16, 17, 22]. A study done by X, [8] was the pioneer study in accessing depth information from Kinect in order to interpret the human activity. It was done by establishing a-bag-of-3D point from the depth image before inferring the human activity from action graph. However, there were also a few studies done by combining the RGB and depth information in recognizing human activities [9-12, 23]. Since, spatio-temporal based descriptor generated based on the recent research interest in recognizing human activity by using the RGB camera or video, there were many studies found [9, 10, 23] that implement such descriptor in RGBD image for the similar interest. Study in [23] formulated hyper cuboid 4D from gradient which is taken from the interest point of RGB and depth image. Interest point was selected based on 2D Gaussian filter in spatial domain and 1D Gabor filter in temporal domain for both RGB and depth image. However, investigation in [9] recommended that, it is important to select the interest point solely on the RGB image before the bag-of-words was generated as descriptive histogram; while correspondence interest point of depth image was used to obtain depth information that separates the descriptive histogram into several depth channels. In line with this study, Zhao [10] performed the Histogram of Gradient (HOG) and Histogram of Flow (HOF) from the interest point of RGB image. However, local depth pattern which is adapted from local binary pattern (LBP) was generated from the correspondence depth interest point before classifying the human activity. Another approach was proposed by [11] which is modeling the probabilistic graphical model for human activities based on the joint of 3D point skeleton provided in Kinect. However, to our knowledge the only work which was focused on the human-object interaction activity was proposed by Koppula [12]. The study later was extended as in [11] introduced features of object manipulation as the contextual features to be used in improving the human activity recognition.

Thus, this study put a highlight on the human-object interaction activity recognition. However, this study proposes a new 3D shape descriptor as well as implementing the existing 3D shape descriptor which is initially used in the 3D object retrieval.

2.2 3D shape descriptor

3D shape descriptor can be categorized into four types: (1) Global based descriptor, (2) Local based descriptor, (3) View based descriptor, and (4) Graph-based descriptor. Originally, the 3D shape descriptor was designed for 3D object retrieval which is useful for the field of archeology, biology, anthropology and industrial part designing community. 3D object in a form of 3D mesh surface is commonly used for 3D object retrieval since such form has the capability to illustrate complex shape in a small memory capacity as compared to the 3D point cloud or 3D primitive form [24].

Global based descriptor describes the 3D object in terms of global shape or overall shape. The initial attempt of the descriptor was to generate the 3D object volume, moment and Fourier transform coefficients[25]. Other than that, a study in [26] suggested the convex-hull from the 3D object to be the 3D shape descriptor while several other studies, concentrated on extracting the shape [13] and shape distribution [27, 28]. However, there were also several investigations that demonstrate local based descriptor, which it describes the shape based on the geometric relation between local points in 3D object surface with neighbor points. The examples of local based descriptors are spin image [29] and curvature based descriptor [30, 31]. Graph-based descriptor interprets the 3D object shape in a form of simple informative skeleton such as Medial Scaffold [32] and Reeb Graph[33]. Meanwhile for view-based descriptor, the 3D object is illustrated into 2D view images first before determining the descriptor from the 2D view image. The example of approaches used in this category are Light-Field Descriptor [34], Characteristic view descriptor [35] and elevation descriptor [36].

3 Shape Histogram and XZ-Shape Histogram for Human Object Interaction Activity Recognition

3.1 Shape Histogram and XZ-Shape Histogram

The aim of this study is to introduce the XZ-shape histogram and demonstrate a shape histogram from Kinect-like depth image in recognizing the human-object interaction activities. Therefore, both descriptors were implemented by using our very own dataset called RGBD-HOI dataset (see Fig. 1); consists of a pair of RGB and depth sequence performing eight possible activities including: 1) answering a phone call (An); 2) brushing teeth (Br); 3) drinking from a mug (Dr); 4) lighting a flashlight (Li); 5) making a phone call (Ma); 6) pouring from a mug (Po); 7) spraying from a spray bottle (Sp); and 8) typing using a keyboard (Ty); each activity was performed by 12 subjects. However, this study only highlights on the extraction of depth image. The proposed 3D shape descriptor was recommended to be implemented from the depth image in a form of 3D point cloud. Thus, several approaches were executed during preprocessing in order to obtain 3D point cloud. During preprocessing, fixed bounding box per sample sequence was determined to highlight the region of interest as illustrated in Fig.2. After that, multilevel thresholding (fixed minimum and maximum threshold value to 680 and 830) was formulated from resultant image in order to remove background pixels within the bounding box before converting the retained pixels into 3D points cloud (see Fig.2) by using the approach that has been proposed in [17].

Shape histogram represents the number of 3D point resides in each designed bin. Defined bin can be modeled according to (1) shells model, (2) sectors model; or (3) combination between shells and sectors model. However, shape histogram was implemented with bin designed based on the shells model as such shape histogram is invariant to rotation. Fig.3 portrays the implementation of shape histogram to 3D point cloud; shells were built with respect to the centroid of the 3D point cloud, O_{sh}. O_{sh} was localized with coordinate x_o, y_o and z_o for each 3D point cloud and formulated as,

$$\left(x_o, y_o, z_o\right) = \left(\min x + r_x, \min y + r_y, \min z + r_z\right) \quad (1)$$

$$r_x = \frac{\left(\max x - \min x\right)}{2} \quad (2)$$

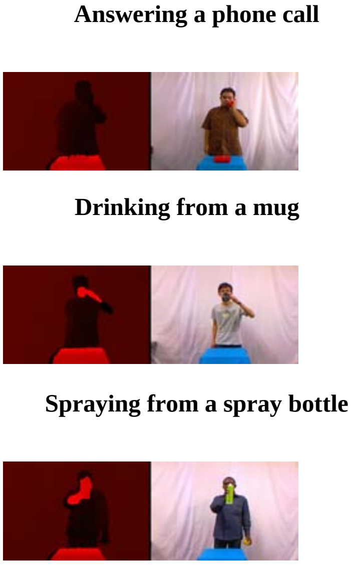
$$r_y = \frac{\left(\max y - \min y\right)}{2} \quad (3)$$

$$r_z = \frac{\left(\max z - \min z\right)}{2} \quad (4)$$

where x, y and z are the coordinates of the whole point exist in 3D point cloud. In this study, shape histogram with number of shells of P=500 and P=1000 were generated for comparison purpose.

XZ-shape histogram was generated by merging the X-shape histogram and Z-shape histogram. X-shape histogram was formulated based on shell model which is in a form of plane with surface normal in x-axis direction (see Fig.4). The drawback of X-shape histogram is that the descriptor incapable in differentiating the left-handed or right-handed subject (see Fig. 5(a)-(d)). Fig. 5 (b) and (d) display the X-shape histogram generated from left-handed subject (see Fig. 5(a)) and right-handed subject (see Fig. 5(b)) which are dissimilar in shapes. In order to overcome this problem, each generated X-shape histogram was flipped on condition that the frequency of the first bin is more than the last bin. The output for flipping the histogram can be identified in Fig. 5(e) that was formulated from X-shape histogram in Fig. 5(d).This mechanism managed to correct the X-shape histogram for depth image in Fig.5 (b) which was similar to X-shape histogram generated from depth image in Fig. 5(a). However, Z-shape histogram was prepared based on modeling the shell for several planes with surface normal in z-axis direction (see Fig. 6). Therefore, with the use of this approach, there was no issue of left-handed or right-handed subject.

This study evaluated the performance of X-shape histogram with P_x =5, 10, 20, 30 and 40; and Z-shape histogram with P_z =3, 5 and 10. The performance of each case was presented later in Section 4 before XZ-histogram was carried out based on the concatenation of the best X-shape histogram and Z-shape histogram.



Answering a phone call

Drinking from a mug

Spraying from a spray bottle

Fig.1: Example samples from RGBD-HOI dataset.

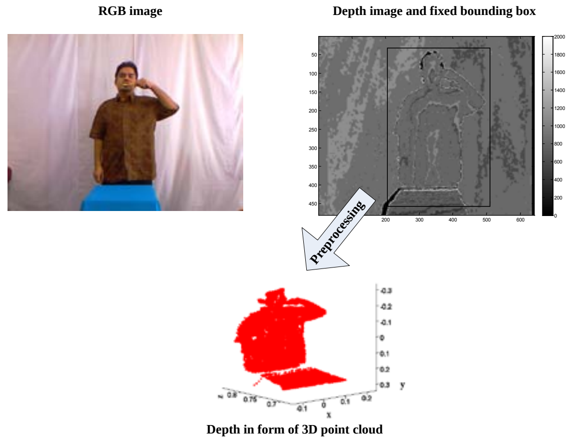

RGB image
Depth image and fixed bounding box
Preprocessing
Depth in form of 3D point cloud
x
y

Fig.2: Preprocessing from depth image

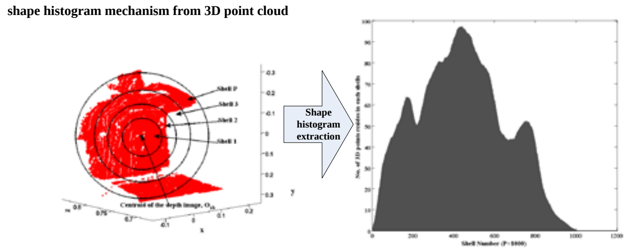

shape histogram mechanism from 3D point cloud
Shell P
Shell 3
Shell 2
Shell 1
Centroid of the depth image, O_c
Shape histogram extraction
No. of 3D points resides in each shells
Shell Number (P=1000)

Fig.3: Shape histogram generated from 3D point cloud

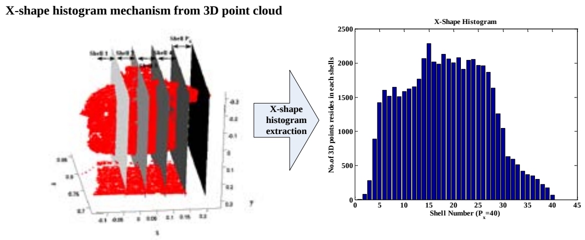
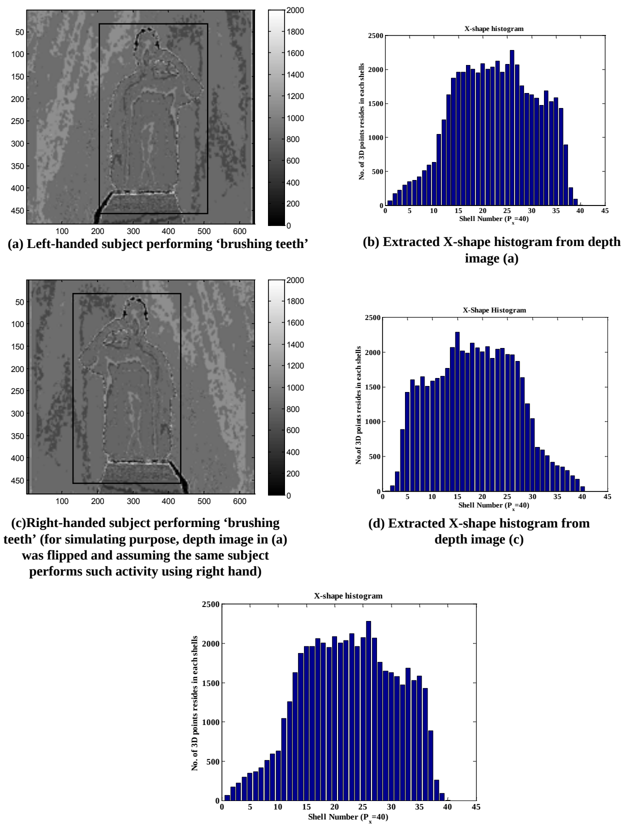

X-shape histogram mechanism from 3D point cloud
Shell 1
Shell 2
Shell 4
Shell P_x
X-shape histogram extraction
X-Shape Histogram
No.of 3D points resides in each shells
Shell Number (P_x=40)

Fig.4: X-shape histogram extracted 3D point cloud



(a) Left-handed subject performing 'brushing teeth'

(b) Extracted X-shape histogram from depth image (a)

(c)Right-handed subject performing 'brushing teeth' (for simulating purpose, depth image in (a) was flipped and assuming the same subject performs such activity using right hand)

(d) Extracted X-shape histogram from depth image (c)

(e) X-shape histogram of (d) after flipping histogram process which is similar to histogram in (b)

Fig. 5: The flipping histogram process in order to avoid right-handed and left-handed subject occur in generating X-shape histogram

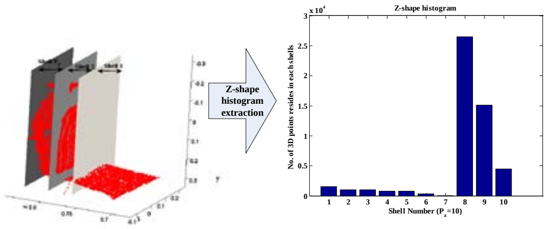


Fig. 6: Z-shape histogram extracted 3D point cloud2. Evaluation Framework

3.2 Evaluation Framework

For each sequence of depth frame per sample in dataset, a set of 3D shape descriptors $X_i = \{x_f\} \mid f = \{1,2,3\ldots F\}$ $x_f \in \mathbb{R}^d$ was formulated from each correspondence depth frame; where f is the frame index in each sample, i is the index for sample in the dataset and d is the 3D descriptor dimension. The main target is to obtain the receiver operating characteristic (ROC) for each 3D shape descriptor as the 3D shape descriptor performance is inferred by the ROC curve and area under ROC curve (AUC-ROC). Then, X_i was quantized into K number of possible symbols to produce $Y_i = \{y_f\} \mid y_f \in \{1,2,3\ldots K\}$. In this process, each 3D shape descriptor per frame in dataset was utilized in order to define K number of symbols which is based on K-mean approach before assigning each 3D descriptor into a symbol, y_f. Afterward, Y_i was used to establish the self-matching matrix M defined as,

$$M(i,j) = m\left(Y_i, Y_j\right) \text{ s.t } \quad i = j = \{1,2,\ldots S \times C\} \tag{5}$$

where S was defined as number of sample per activities class which in this case is 12 and C is the total number of activity classes that were evaluated. All eight activity classes in the dataset were used in the evaluation process. $m\left(Y_i, Y_j\right)$ is the matching measurement function that was used to obtain matching value between sample Y_i and Y_j that represents the row and column for each element in matrix M. Thus, the arrangement for both row and column of matrix M were designed as follows:

$$Y_c = \{Y_n\} \mid s = \{1,2,3\ldots.S\} \tag{6}$$

$$Y = \{Y_c\} \mid c = \{1,2,3\ldots.C\} \tag{7}$$

Fig.7(a) demonstrates the arrangement for both row, (6) and column, (7) of matrix M.In this study, edit distance [37] was done by matching the measurement function $m\left(Y_i, Y_j\right)$ as both Y_i and Y_j are in one dimensional sequence.

In order to plot ROC curve for each 3D shape descriptor, each self-matching matrix M was operated with correspond to ground-truth matrix M_t (see Fig.7(b)) in order to produce the true match t_m, true non-match t_n, false match f_m and false non-match f_n that were defined as,

$$t_m = \sum_i^{SC}\sum_j^{SC}\{M_t(i,j) \times M'(i,j)\} \tag{8}$$

$$t_n = \sum_i^{SC}\sum_j^{SC}\{(1 - M_t(i,j)) \times (1 - M'(i,j))\} \tag{9}$$

$$f_m = \sum_i^{SC}\sum_j^{SC}\{(1 - M_t(i,j)) \times M'(i,j)\} \tag{10}$$

$$f_n = \sum_i^{SC}\sum_j^{SC}\{M_t(i,j) \times (1 - M'(i,j)\} \tag{11}$$

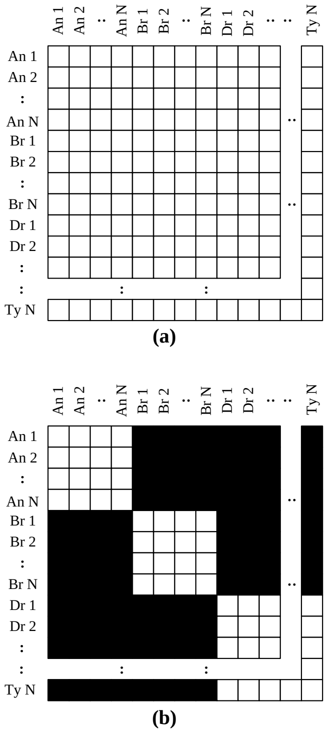


Fig. 7: (a) Self-matching matrix M (partially shown); both row and column were formulated using (2) and (3). (b) Ground-truth matching matrix, M_t which was built for self-matching matrix in (a).

M_t was formed from binary element which the size is similar to the size of M, $(S\times C)\times(S\times C)$ pixels but consists of white blocks with $(S\times C)$ pixels that is replicated within M in a diagonal direction. $M^{'}$ is the binary version of M , produced by thresholding M with threshold value of ε. Thus, the ROC can be plotted on the condition that ε was varied from 0 to 1 and produced true-positive rate (TPR) and false-positive rate (FPR) by using the following equations,

$$TPR=\frac{t_m}{t_m+f_n} \tag{12}$$

$$FPR=\frac{f_m}{f_m+t_n} \tag{13}$$

For each 3D shape descriptor case, three ROC curves were generated since the quantization in producing a sequence of symbol Y_i was based on K-means that suffers with initial seed error. Thus, the mechanism was repeated for three times before inferring the 3D shape descriptor performance based on the ROC average that represents such 3D shape descriptor.

4 Results and Discussion

The purpose of this study is to demonstrate the proposed XZ-histogram as motivated from shape histogram to be implemented in recognizing the activity of daily living (ADLs) which is specific to the human-object interaction activity based on the Kinect-like depth image. Therefore, before the performance of XZ-shape histogram can be concluded, several investigations were carried out which started with the implementation and evaluation of shape distribution in the RGBD-HOI dataset. Then, the same mechanism was executed to X-histogram and Z-histogram before finding the best performance of X-histogram and Z-performance which to be merged later in order to form XZ-histogram. Since, a lot of cases for each type of 3D shape descriptor were simulated (by varying K, P , P_x and P_z .), it is very difficult to observe ROC curve that represents each case in a graph. Thus, the performance of each 3D descriptor case was inferred based on the AUC-ROC curve that was displayed in a form of table (see Table 1-5).

The first investigation started with the implementation of shape histogram by using our own RGBD-HOI dataset. Table 1 illustrates the AUC-ROC for each case in shape histogram. The performance of shape histogram depends on the P and K which are used to establish the evaluation mechanism. When P=500, a remarkable performance was seen when K was set to 100 before minimally declining when K was more than 100. However, there was significant increment performance for shape histogram with P=1000 when the number of K symbols were around 20 to 300. The results indicate that when P was set too small (P=500), the produced shape histogram was insufficient to describe the shape of 3D surface. This insufficient shape histogram lost more spatial information when the descriptor was quantized to a big number of K possible symbols since it quantized the insufficient meaningful information. However, when the P used to generate shape histogram is adequate to describe the shape of 3D surface (P=1000), the 20 to 300 number of K symbols was reported to increase the shape histogram performance because the quantization process had assessed a sufficient 3D shape information from shape histogram.

The second investigation involved with the experimentation of the mentioned X-shape histogram: (1) X-shape histogram without flipping histogram; 2) X-shape histogram with flipping histogram. The results for both 3D shape descriptors were illustrated in Table 2 and Table 3. Even there were insignificant cases in Table 3 which showed that the performance of X-shape histogram with flipping histogram process did not improve, the overall average of AUC-ROC,0.6395 in Table 3 was still outperformed the average AUC-ROC, 0.6384 in Table 2 . Besides that, the maximum AUC-ROC in Table 3 (0.6454) which was more than maximum AUC-ROC in Table 2 (0.6443). This shows that the implementation of flipping histogram improved the X-shape histogram performance. Moreover, in line with previous shape histogram, a variety of AUC-ROC value determined as in Table 2 and Table 3 was due to the amount of P_x which was to be considered as sufficient shells to generate X-histogram and the amount of K possible symbol used also depended on P_x .

The last investigation which was Z-shape histogram implementation in RGBD-HOI dataset was demonstrated in Table 4. As there were variety of K symbols and P_z, it seemed that the maximum performance of Z-shape histogram which was AUC-ROC-0.6476 occurred when the number of shell P_z used is 5 with number of K symbols used is 200 in vector quantization process

From the investigations, it was found that X-shape histogram with P_x=5 and K=300 (can be referred in Table 3) and Z-shape histogram with P_z=5 and K=200 (can be referred in Table 4) were selected as the best shape histograms to be combined as the new XZ-shape histogram. The new XZ-shape histogram was formulated again with the same evaluation framework in order to extract the performance in terms of AUC-ROC. Table 5 shows the XZ-shape performance as simulated by using K=20 to K=300 number of symbols in vector quantization process. The maximum performance of XZ-shape histogram based on the AUC-ROC is 0.6484 that occurred at K=300. The overall comparison was summarized in Table 6 that indicate the AUC-ROC for each 3D shape descriptor case (the best case) for shape histogram, X-shape histogram, Z-shape histogram and XZ-shape histogram. The proposed XZ-shape histogram achieved a remarkable performance as compared to rest of shape descriptor used in recognizing the human-object interaction activities via Kinect-like depth image.

Table 1. AUC-ROC for Several Shape Histogram based on Different K symbols and P number of Shell

		Number of cells used to establish shape histogram, P	
		500	**1000**
Number of K symbols used in vector quantization process	**20**	0.6402	0.6339
	100	0.6457	0.6393
	200	0.6431	0.6452
	300	0.6436	0.6455

Table 2. AUC-ROC for Several X-Shape Histogram (without flipping histogram process) Based on Different K Symbols and P_x Number of Shell

		Number of cells used to establish shape histogram, P_x				
		5	**10**	**20**	**30**	**40**
Number of K symbols used in vector quantization process	**20**	0.6343	0.6301	0.6323	0.6303	0.6299
	100	0.6417	0.6392	0.6386	0.6393	0.6386
	200	0.6422	0.6422	0.6401	0.6395	0.6408
	300	0.6443	0.6413	0.6387	0.6417	0.6419

Table 3. AUC-ROC for Several X-Shape Histogram (with flipping histogram process) Based on Different K Symbols and P_x Number Of Shell

		Number of cells used to establish shape histogram, P_x				
		5	**10**	**20**	**30**	**40**
Number of K symbols used in vector quantization process	**20**	0.6305	0.6298	0.6383	0.6288	0.6292
	100	0.6414	0.6417	0.6397	0.6409	0.6415
	200	0.6440	0.6433	0.6434	0.6413	0.6396
	300	0.6454	0.6442	0.6425	0.6411	0.6432

Table 4. AUC-ROC for Several Z-Shape Histogram Based on Different K Symbols and P_x Number Of Shell

		Number of cells used to establish shape histogram, P_z		
		3	**5**	**10**
Number of K symbols used in vector quantization process	**20**	0.6374	0.6336	0.6414
	100	0.6445	0.6428	0.6475
	200	0.6470	0.6476	0.6468
	300	0.6469	0.6439	0.6472

Table 5. AUC-ROC for Several XZ-Shape Histogram Based on Different K Symbols

Number of K symbols used in vector quantization process	**AUC-ROC**
20	0.6360
100	0.6418
200	0.6426
300	0.6484

Table 6. AUC-ROC for different types of 3D shape descriptors

3D Shape Descriptors	**AUC-ROC**
Shape Histogram. P=500, K=100	0.6457
X-shape histogram. P=5, K=300	0.6454
Z-shape histogram. P=5, K=200	0.6476
XZ-shape histogram. K=300	0.6484

4 Conclusion

In summary, this study presents a new 3D shape descriptor, XZ-shape histogram in recognizing the human-object interaction activity based on the Kinect-like depth image. This study performs an intensive comparison of such descriptor with previous shape histograms; presents several cases for each descriptor by varying the number of shells used so as to generate the descriptor and number of symbols used during quantization process before finding the best descriptor for each case based on the AUC-ROC value. The result showed that the proposed XZ-shape histogram managed to show an outstanding performance as compared to the other best 3D shape descriptor cases. Moreover, XZ-shape histogram which consists of only 10 shells managed to outperform the previous shape histogram which consists of 500 shells. The results of this study therefore challenge the previous shape histogram in terms of providing a lower dimension descriptor but manage to improve the discriminating power of recognizing the human-object interaction activity. Therefore, it is suggested that the descriptor should be integrated with classifier module as to establish the whole human-object interaction activities recognition system. The descriptor as well need to be compiled with other modules in order to develop an automated activities of daily living (ADLs) monitoring system.

5 Acknowledgment

The author would like to express their gratitude to University Teknologi Malaysia (UTM) and the Minister of Higher Education (MOHE), Malaysia for supporting this research under Research Grant No. Q.J130000.7123.02J25.

References:

[1] J. Hoey, P. Poupart, A. v. Bertoldi, T. Craig, C. Boutilier, and A. Mihailidis, Automated handwashing assistance for persons with dementia using video and a partially observable Markov decision process, Computer Vision and Image Understanding, Vol. 114, 2010,pp. 503-519.

[2] J. S. Brach and J. M. VanSwearingen, Research Report Physical Impairment and Disability : Relationship to Performance of Activities of Daily Living in Community-Dwelling Older Men, Physical Therapy, Vol. 8, 2002,pp. 752-761.

[3] A. Paraschiv-Ionescu, C. Perruchoud, E. Buchser, and K. Aminian, Barcoding Human Physical Activity to Assess Chronic Pain Conditions, PLoS ONE, Vol. 7, 2012,p. e32239.

[4] M. A. As'ari and U. U. Sheikh, Vision based assistive technology for people with dementia performing activities of daily living (ADLs): an overview, in Fourth International Conference on Digital Image Processing (ICDIP), Kuala Lumpur, 2012, pp. 83342T-83342T.

[5] T. B. Moeslund, A. Hilton, and V. Krger, A survey of advances in vision-based human

motion capture and analysis, Comput. Vis. Image Underst., Vol. 104, 2006,pp. 90-126.

[6] Aaron F. Bobick, Movement, activity and action: the role of knowledge in the perception of motion, Philosophical Transactions of the Royal Society B:Biological Sciences, Vol. 352, 1997,pp. 1257-1265.

[7] G. Lavee, E. Rivlin, and M. Rudzsky, Understanding Video Events: A Survey of Methods for Automatic Interpretation of Semantic Occurrences in Video, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 39, 2009,pp. 489-504.

[8] L. Wanqing, Z. Zhengyou, and L. Zicheng, Action recognition based on a bag of 3D points, in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010, pp. 9-14.

[9] N. Bingbing, W. Gang, and P. Moulin, RGBD-HuDaAct: A color-depth video database for human daily activity recognition, in 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011, pp. 1147-1153.

[10] Y. Zhao, Z. Liu, L. Yang, and H. Cheng, Combing RGB and Depth Map Features for human activity recognition, in 2012 Asia-Pacific Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012, pp. 1-4.

[11] J. Sung, C. Ponce, B. Selman, and A. Saxena, Unstructured human activity detection from rgbd images, in 2012 IEEE International Conference on Robotics and Automation (ICRA), 2012, pp. 842-849.

[12] H. S. Koppula, R. Gupta, and A. Saxena, Learning Human Activities and Object Affordances from RGB-D Videos, arXiv preprint arXiv:1210.1207, 2012.

[13] M. Ankerst, G. Kastenmüller, H.-P. Kriegel, and T. Seidl, 3D shape histograms for similarity search and classification in spatial databases, SSD' 99, 1999,pp. 207-226.

[14] P. Huang, A. Hilton, and J. Starck, Shape Similarity for 3D Video Sequences of People, Int. J. Comput. Vision, Vol. 89, 2010,pp. 362-381.

[15] Kinect. (2010, 04/02/2012). http://www.xbox.com/en-us/kinect.

[16] L. Bo, K. Lai, X. Ren, and D. Fox, Object Recognition with Hierarchical Kernel Descriptors, in Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[17] K. Lai, L. Bo, X. Ren, and D. Fox, A Large-Scale Hierarchical Multi-View RGB-D Object Dataset, in Proc. of International Conference on Robotics and Automation (ICRA), 2011.

[18] K. Lai, L. Bo, X. Ren, and D. Fox, A Scalable Tree-based Approach for Joint Object and Pose Recognition, in Twenty-Fifth Conference on Artificial Intelligence (AAAI), 2011

[19] L. Xia, C. Chen, and J. K. Aggarwal, Human Detection Using Depth Information by Kinect, in International Workshop on Human Activity Understanding from 3D Data in conjunction with CVPR (HAU3D), 2011.

[20] S. Lang, "Sign Language Recognition with Kinect," Bachelor, Institut für Informatik, Freie Universitä t Berlin, 2011.

[21] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti, American sign language recognition with the kinect, in Proceedings of the 13th international conference on multimodal interfaces, Alicante, Spain, 2011, pp. 279-286.

[22] K. Lai, L. Bo, X. Ren, and D. Fox, Detection-based Object Labeling in 3D Scenes, in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012, pp. 2169–2178.

[23] H. Zhang and L. E. Parker, 4-dimensional local spatio-temporal features for human activity recognition, in 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2011, pp. 2044-2049.

[24] F. Lafarge, R. Keriven, and M. Bredif, Insertion of 3-D-Primitives in Mesh-Based Representations: Towards Compact Models Preserving the Details, IEEE Transactions on Image Processing, Vol. 19, 2010,pp. 1683-1694.

[25] Z. Cha and C. Tsuhan, Efficient feature extraction for 2D/3D objects in mesh representation, in International Conference on Image Processing 2001, pp. 935-938 vol.3.

[26] J. Corney, H. Rea, D. Clark, J. Pritchard, M. Breaks, and R. Macleod, Coarse filters for shape matching, IEEE Computer Graphics and Applications, Vol. 22, 2002,pp. 65-74.

[27] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, Matching 3D models with shape distributions, in SMI 2001 International Conference on Shape Modeling and Applications, 2001, pp. 154-166.

[28] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, Shape distributions, ACM Trans. Graph., Vol. 21, 2002,pp. 807-832.

[29] A. E. Johnson and M. Hebert, Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 21, 1999,pp. 433-449.

[30] S. Heung-Yeung, M. Hebert, and K. Ikeuchi, On 3D shape similarity, in Proceedings CVPR '96, 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996, 1996, pp. 526-531.

[31] T. Zaharia and F. Preteux, Three-dimensional shape-based retrieval within the MPEG-7 framework, in Proceedings SPIE Conference on Nonlinear Image Processing and Pattern Analysis XII, 2001, pp. 133-145.

[32] M.-C. Chang and B. B. Kimia, Measuring 3D shape similarity by graph-based matching of the medial scaffolds, Computer Vision and Image Understanding, Vol. 115, 2011,pp. 707-720.

[33] M. Hilaga, Y. Shinagawa, T. Kohmura, and T. L. Kunii, Topology matching for fully automatic similarity estimation of 3D shapes, presented at the Proceedings of the 28th annual conference on Computer graphics and interactive techniques, 2001.

[34] D.-Y. Chen, X.-P. Tian, Y.-t. Shen, and M. Ouhyoung, On Visual Similarity Based 3D Model Retrieval, presented at the Computer Graphics Forum (EUROGRAPHICS'03), 2003.

[35] S. Mahmoudi and M. Daoudi, 3D models retrieval by using characteristic views, in 16th International Conference on Pattern Recognition, 2002, pp. 457-460 vol.2.

[36] J.-L. Shih, C.-H. Lee, and J. T. Wang, A new 3D model retrieval approach based on the elevation descriptor, Pattern Recognition, Vol. 40, 2007,pp. 283-295.

[37] P. H. Sellers, The theory and computation of evolutionary distances: Pattern recognition, Journal of Algorithms, Vol. 1, 1980,pp. 359-373.