BIOSCIENCES BIOTECHNOLOGY RESEARCH ASIA, August 2015. Vol. **12**(2), 1139-1144

Imrovement of Human Key Posture Recognition

Abay Kabylovich Nussipbekov[1], Yedilkhan Nesipkhanovich Amirgaliyev[2] and Minsoo Hahn

[1]Al-Farabi Kazakh National University, al-Farabi ave. 050038, Almaty, Kazakhstan
[2]Institute of Information and Computing technologies MES RK, Puskhina str. 125, 050010, Almaty, Kazakhstan.
[3]Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon, South Korea

DOI: http://dx.doi.org/10.13005/bbra/1765

(Received: 19 February 2015; accepted: 29 April 2015)

Human pose recognition is an important problem in such fields like scene recognition, robotics, multimedia systems an etc. It plays an essential role in full body gesture recognition. The obtained poses can be then used to detect gestures. In gesture recognition it is essential to capture all poses. Because missing some key pose may significantly decrease the gesture recognition accuracy. Most pose recognition approaches use silhouettes extracted from regular cameras that may not bring a favorable pose accuracy or different kind of special costumes that uncomfortable to wear. In this paper we propose to classify poses based on their skeleton appearance. For this we use depth camera called Microsoft Kinect. The main difference of our method is that we add some additional objects that take place in the scene into skeleton provided by Kinect camera. In another words we use an advanced type of skeleton unlike camera can provide. The skeleton can be modified upon the specific pose recognition task by modifying it which in turn increase the pose recognition accuracy. In our work we adopt it for golf postures recognition problem by tracking golf club head. The background of an image is subtracted by using depth image histogram, then we calculate object coordinates using HSV(Hue Saturation Value) color information and image moment. Finally we perform classification using Support Vector Machines. The obtained results demonstrate the efficiency of the proposed method.

Key words: Pose recognig, skeleton, Support Vector Machines, Microsoft Kinect

Human posture recognition is an interesting problem in computer vision domain. It have a wide range of different applications for such tasks like Human Computer Interaction (HCI), robotics, surveillance, multimedia systems and etc. Also computer games based on human gestures became popular in recent years.

Human posture recognition plays an important role in a full body gesture recognition. The obtained poses can be then used to detect gestures (Miranda et al., 2012). In gesture recognition it is essential to capture all poses. Because missing some key pose may significantly decrease the gesture recognition accuracy. Many of the existing works use only body information such as silhouettes and skeletons to recognize postures while there are also other important objects that are taking place in a scene and which can bring an important information. In this paper we are considering the special case of such situation which is golf playing scene. In golf playing scene golf club head can bring a useful information that can significantly improve the pose recognition accuracy. In figure 1 you can see that sometimes (last two postures on the right) human body pose may be not enough to recognize real golf player posture. And additional objects (golf club head) that are a part of scene can help in such situations.

* To whom all correspondence should be addressed.

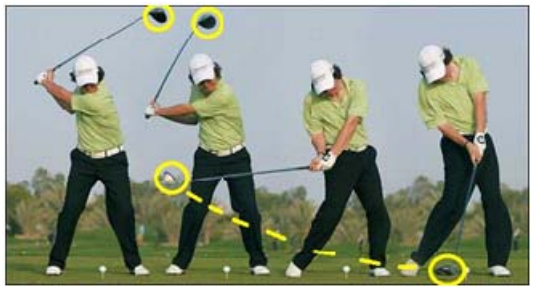

Fig. 1. Golf club head can considerably help in posture recognition

The task of pose recognition has been addressed in many works. Bo Peng and Gang Qian (Peng and Qian, 2008) used human silhouettes to recognize postures. In their work they performed tensor decomposition and projection using obtained silhouette images and then calculated body orientation coefficient vectors that they used for classification. Similarly, Mohiuddin Ahmad *et al.* (Ahmad and Lee, 2008) used silhouettes information. But unlike in previous study they applied a spatio-temporal silhouette representation to characterize human gestures. As another alternative, many methods have been developed using human body skeleton. A study by (Yoon and Kuijper, 2010) propose using segmented skeletal features to recognize human actions. Later, researchers from International Research Institute MICA (Le et. al., 2013) adopted human skeleton provided by Kinect camera to directly recognize postures. Other researchers also used human skeleton in their works (Xiao et al., 2012; Monir et. al., 2012). One of the significant work in this domain was done by (Raptis et. al., 2011). They have got high recognition results by using an angular representation of the human skeleton.

In our work we used Microsoft Kinect depth camera which is very popular nowadays. The difference of that camera from others is that unlike other RGB (red green blue) cameras it provides depth information using which we can capture 3D scenes. In our system we firstly separate foreground from the background by applying iterative tresholding method on depth histogram. Then we detect golf club head based on its HSV value and finally calculate its real world coordinates according to camera. The obtained coordinates are added into golf player skeleton so that we get an improved version of it. Next task is to extract a feature set for the classification step. We use the body parts inclination angle values as a feature vector. In the last step we perform a classification of poses by using Support Vector Machines (SVM).

MATERIALS AND METHODS

Preprocessing

In order to add golf club head into skeleton we have to find its coordinates. For this purpose we have to apply some preprocessing steps.

Firstly we perform background subtraction. In our case we want to eliminate all the unnecessary part from our image in order to make object detection in next step easier. Depth camera makes its convenient to subtract background efficiently. In depth image each pixel demonstrates not color information but rather how far it stands from camera. We convert depth measurements from the depth image provided by camera into byte data type in the range 0-255 so

Fig. 2. Grayscale depth image. Each pixel demonstrates its distance from camera

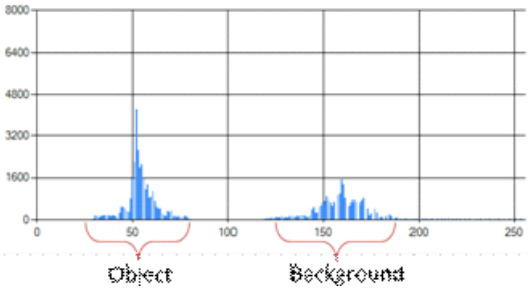

Fig. 3. Histogram of depth image. Two peaks are belonging to tracking object and background

that we can work with grayscale image. In figure 2 you can see sample of depth image of human body sitting in front of the camera. The dark pixels are closer to the camera than bright ones. Next, we create histogram like the one shown in figure 3.

A histogram is a function m_i that counts the number of occurrences of specific pixel that fall into some range called bins. The number of bins in grayscale image is 250. The total number *n* of pixels is thus:

$$n = \sum_{i=1}^{k} m_i \ , \qquad ...(1)$$

where *k* is the number of bins.

Obtained histogram gives us clear picture of body and its background. The problem now is to separate two given peaks from each other. We do it by firstly calculating approximate initial threshold level which will be used then in the next step. We find initial threshold level by firstly applying mean filter to the obtained histogram graph. Then we find two peaks in the graph and then calculate their average which will be approximate initial threshold value. Peaks can be found by using such ready functions in matlab like findpeaks(). The next step is to refine the threshold value. It could be done by the method proposed by T.W. Ridler and S. Calvard (Ridler and Calvard, 1978). The overview of this method is demonstrated in figure 4. The main principle here is to use our previously found initial threshold value as a first value of switching function *f(s)* which separates digitalized image signals (pixels) into background and object according to threshold value. In other words those signals that are below threshold value are considered to be foreground and others background. When the whole image has been separated a new value of threshold is calculated again as an average of background and foreground values. Then *f(s)* is calculated again using new value of threshold and all this process repeats several times until threshold remains unchangeable.

After subtracting background and leaving only golf player with golf club we have to detect club head on the given image. We decided to do it by using its color information. Object detection by color has been addressed in many works (Vo et al., 2010; Iwai et al., 1996). In our work we will use OpenCV library for simplicity. The detection will be done by HSV (Hue, Saturation, Value) tresholding. We chose HSV because its more natural for object detection by color. Firstly we take an image, convert it to HSV and then divide it into channels. In other words we create four images; one for storing image in HSV format and other three for dividing an image into separate channels H, S and V. Knowing the color we are looking for, we can create HSV threshold values manually, however there are plenty of ready utilities for that purpose for example cvRangeS(). Finally we apply logical operation AND (cvAND(h_range, s_range, v_range)) upon our three threshold images. In other words, we will have white pixels only for those parts which are in all three images simultaneously. Otherwise we will have logical 0 which means black pixel.

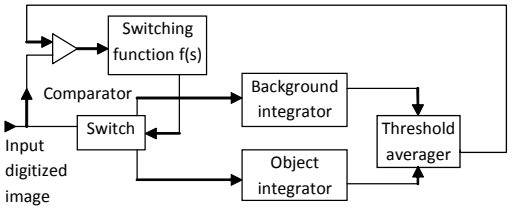

Fig. 4. Overview of image processor for iterative threshold selection proposed by Riddler and Calvard (Riddler and Calvard , 1978)

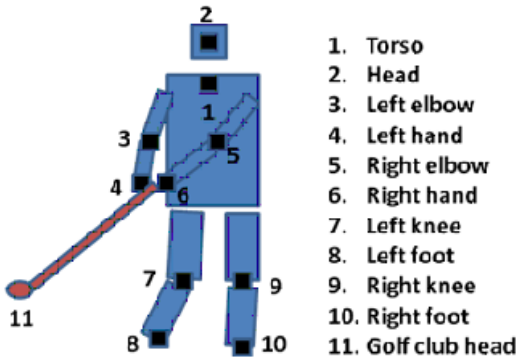

Fig. 5. Human body parts with golf club head

Next, we find coordinates of golf club. Knowing which pixels belong to object, we can calculate the x and y coordinates of the center of an object on that image. This is done by calculating a weighted average of all pixels which belong to an object (Bunner et. al., 2007):

$$\begin{pmatrix} c_x \\ c_y \end{pmatrix} = \frac{\sum_{x,y} m(x,y) \begin{pmatrix} x \\ y \end{pmatrix}}{\sum_{x,y} m(x,y)}, \quad ...(2)$$

where *m (x,y)* is the mask image at given coordinates.

The last step in getting object coordinates is to transform them into world coordinates. This is performed by the special Kinect SDK (software development kit) method Map Depth Point To Skeleton Point (Depth Image Format. Resolution320x240Fps30, depthPoint), where depthPoint is taken from previously calculated (*x*,*y*) coordinates. Finally we get real X and Y coordinates of golf club head which will be used in the next pose recognition section.

Pose Classification

Now having obtained coordinates of golf club head we add it to the skeleton obtained from depth camera and get adapted skeleton for golf pose recognition. In our work we use 2D representation of human skeleton for simplicity (see figure 5). Coordinates information can be used as a feature set for classification task but because of that different people have different body sizes it would be better to use inclination angle values of body parts and golf club head which will make the system person independent. It also makes obtained skeleton invariant to the orientation of camera. The angle representations is similar to (Raptis et.al., 2011) except that we use only inclination angle values. The joints are divided into first-degree joints and second-degree. The first-degree joints are those that are directly related to body torso and other joints that are adjacent to first-degree joints are supposed to be second-degree.

We employ SVM (Support Vector Machines) as a classifier (Burges, 1997). SVM is well known because of its high accuracy and its capability to do a non-linear classification efficiently and work with high dimensional data.

RESULTS

In our experiment we used Microsoft Kinect camera. The distance between camera and golf players was 2 meters. The color of golf club head was black. The application was written on C# .NET (data collection) and matlab (data classification).

We used training data set $D=\{x_i,y_i,\}_{i=1}^{l}$, $X_i \in R^p$ where l is the size of training feature set, x_i is p dimensional vector and y_i is one of the predefined posture. All the dataset was manually labeled. We chose 4 key poses for the basic gesture in golf game such as "Back swing" which is initial pose, "Before impact" the moment before striking the ball, "Impact" moment and "End of swing" that is last pose in the gesture. We used about 700 different poses of several different people, 70% of them for training and the remaining 30% for testing. SVM constructs a hyperplane $\omega.\Phi(x)+b=0$ which classifies given postures.

Table 1. Pose recognition accuracy confusion matrix before adding golf club into skeleton

Pose	Back swing	Before impact	Impact	End of swing
Back swing	93	0	0	0
Before impact	4	63	32	1
Impact	3	37	68	1
After impact	0	0	0	98

Table 2. Pose recognition accuracy confusion matrix after adding golf club into skeleton

Pose	Back swing	Before impact	Impact	End of swing
Back swing	95	0	1	0
Before impact	3	89	9	2
Impact	2	11	87	4
After impact	0	0	3	94

DISCUSSION

In table 1 we demonstrate the result of pose classification experiments before adding golf club head into the skeleton. As you can notice the classification fails for such poses like "Before impact" and "Impact" which is not good because detecting the moment of impact is very important for the following gesture recognition task for example. However in table 2 we modified skeleton and got promising results 89% instead of 63% and 87% instead of 68%.

Generally the obtained results are considered to be high enough results in the field of gesture recognition and are comparable to the works in the related fields. For example (Miranda et al., 2012) and (Samanta et. al., 2012) used posture recognition in their studies. They haven't shown posture recognition results in their works but the gesture recognition accuracy results which was demonstrated in the subsequent step and that were high enough shows that the posture recognition accuracy was also almost the same and comparable to our results.

On the other hand, we can compare our work with those whose final aim was pose recognition rather than gesture recognition. The classification of poses through using stereo vision and 3D modeling by Pellegrini *et. al.* (Pellegrini et. al., 2008) has got almost the same results on average. Lutz Goldmann *et. al.* (Goldmann et. al., 2004) in their work have got high results (95.59%), but they mention that the recognition of the view in their work is still a problem (77.84%). The studies provided by (Peng and Qian, 2008), (Le et. al., 2013) and (Monir et.al., 2012) obtained very high results which is in some cases is higher than our results. But they used simple poses like standing, sitting, lying, bending, moving and etc. while in our work some of the poses were very close to each other and thus harder to distinguish. In other words our aim was to demonstrate how additional objects can be integrated into skeleton and thus increase the recognition accuracy.

CONCLUSION

This paper addresses an important task in gesture recognition "Pose classification". We demonstrated that sometimes using human skeleton in order to classify poses may not be enough and we can solve the problem by adding additional objects that take place in the scene. In our work we showed how these objects can be added into human skeleton. Firstly we performed an iterative background subtraction method on the depth image histogram. After that we calculated the coordinates of golf club head on the image and then transformed them into real world coordinates. Finally after getting modified golf player skeleton we used its body parts angle information as a feature set for the SVM classifier. The obtained results demonstrate the efficiency of the proposed method. In our work we work only with 2D angle values however the model can be extended into 3D which will allow to capture more complex poses such as human body turnings, also golf club head can also be tracked upon its depth information rather than color (HSV) which will make the recognition independent of lighting conditions. So our next plan is to use 3D representation of poses and continue our work by using it in full body gesture recognition problem.

REFERENCES

1. Ahmad, Mohiuddin, and Seong-Whan Lee. "Recognizing Human Actions Based on Silhouette Energy Image and Global Motion Description." *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition* 2008; **1**–6. doi:10.1109/AFGR.2008.4813435. http://ieeexplore.ieee. org/lpdocs/epic03/wrapper.htm?arnumber=4813435.
2. Bunner R., F. Doepke, B. Laden, edited by Hubert Hguyen. *GPU Gems 3*. 2007. Ch. 26
3. Burges, Christopher J C. "A Tutorial on Support Vector Machines for Pattern Recognition" 1997; **43**: 1–43.
4. Goldmann, Lutz, Mustafa Karaman, and Thomas Sikora., "Human Body Posture Recognition Using MPEG-7 Descriptors.", Proceedings of the SPIE, 2004; **5308**: 177-188
5. Iwai, Yoshio, Ken Watanabe, Yasushi Yagi, and Masahiko Yachida. "Using Colored Gloves": 1996; 662–666.
6. Le, Thi-Lan, Minh-Quoc Nguyen, and Thi-Thanh-Mai Nguyen. "Human Posture Recognition Using Human Skeleton Provided by Kinect." *2013 International Conference on Computing, Management and Telecommunications (ComManTel)* 2013; 340–345. doi:10.1109/ComManTel.2013.6482417. http://ieeexplore.ieee.org /lpdocs/epic03/wrapper.htm?arnumber=6482417.
7. Miranda, Leandro, Thales Vieira, Dimas Martinez, Thomas Lewiner, Antonio W. Vieira, and Mario F. M. Campos. "Real-Time Gesture Recognition from Depth Data through Key Poses Learning and Decision Forests." *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images* 2012; 268–275. doi:10.1109/SIBGRAPI.2012.44. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6382766.
8. Monir, Samiul, Sabirat Rubya, and Hasan Shahid

Ferdous. "Rotation and Scale Invariant Posture Recognition Using Microsoft Kinect Skeletal Tracking Feature." *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)* (November): 2012; 404–409. doi:10.1109/ISDA.2012.6416572. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6416572.

9. Pellegrini, Stefano, and Luca Iocchi. "Human Posture Tracking and Classification through Stereo Vision and 3D Model Matching" (about 10): 1–18.
10. Peng, Bo, and Gang Qian., "Binocular Dance Pose Recognition and Body Orientation Estimation via Multilinear Analysis." *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* 2008; 1–8. doi:10.1109/CVPRW.2008.4562970. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4562970.
11. Raptis, Michalis, Darko Kirovski, and Hugues Hoppe. "Real-time Classification of Dance Gestures from Skeleton Animation." *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation - SCA* '2011; **11**: 147. doi:10.1145/2019406.2019426. http://dl.acm.org/citation.cfm? doid= 2019406.2019426.
12. Samanta, Soumitra, Pulak Purkait, and Bhabatosh Chanda. "Indian Classical Dance Classification by Learning Dance Pose Bases" (Icd).
13. Using, Picture Thresholding, and Iterative Selection Method. "Picture Thresholding Using" 1978; **8**: 630–632.
14. Vo, Nam, Quang Tran, Thang Ba Dinh, Tien Ba Dinh, and Quan M. Nguyen. "An Efficient Human-Computer Interaction Framework Using Skin Color Tracking and Gesture Recognition." *2010 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future (RIVF)* (November): 2010; 1-6. doi:10.1109/RIVF.2010.5633368. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5633368.
15. Xiao, Zheng, Fu Mengyin, Yang Yi, and Lv Ningyi. "3D Human Postures Recognition Using Kinect." *2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics* 2012; 344–347. doi:10.1109/IHMSC.2012.92. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6305696.
16. Yoon, Sang Min, and Arjan Kuijper. "Human Action Recognition Using Segmented Skeletal Features." *2010 20th International Conference on Pattern Recognition* 2010; 3740–3743. doi:10.1109/ICPR.2010.911. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm? arnumber =5597576.