Gait Recognition Based on Kinect Sensor

Mohammed Ahmed Naseer Al-Jawad Azhin Sabir

University of Buckingham Applied Computing Department

ABSTRACT

This paper presents gait recognition based on human skeleton and trajectory of joint points captured by Microsoft Kinect sensor. In this paper Two sets of dynamic features are extracted during one gait cycle: the first is Horizontal Distance Features (HDF) that is based on the distances between (Ankles, knees, hands, shoulders), the second set is the Vertical Distance Features (VDF) that provide significant information of human gait extracted from the height to the ground of (hand, shoulder, and ankles) during one gait cycle. Extracting these two sets of feature are difficult and not accurate based on using traditional camera, therefore the Kinect sensor is used in this paper to determine the precise measurements. The two sets of feature are separately tested and then fused to create one feature vector. A database has been created in house to perform our experiments. This database consists of sixteen males and four females. For each individual, 10 videos have been recorded, each record includes in average two gait cycles. The Kinect sensor is used here to extract all the skeleton points, and these points are used to build up the feature vectors mentioned above. K-nearest neighbor is used as the classification method based on Cityblock distance function. Based on the experimental result the proposed method provides 56% as a recognition rate using HDF, while VDF provided 83.5% recognition accuracy. When fusing both of the HDF and VDF as one feature vector, the recognition rate increased to 92%, the experimental result shows that our method provides significant result compared to the existence methods.

Keywords: Human Gait recognition, Kinect sensor, Horizontal Distance Feature, Vertical Distance Feature.

1. INTRODUCTION

Biometrics is the automated use of physiological or behavioral characteristics to determine or verify identity of a person, the physiological biometrics examines physiological characteristics, like: iris, face, fingerprints, DNA, and hand geometry; the behavioral biometrics examines behavioral issues, such as: voice, signature, and gait [1]. The importance of automatic identification of people has increased during the past decades, especially in high security areas such as airports and banks [2]. Each person has distinguishable unique

characteristics that can be used for identification. The science of biometrics uses these unique characteristics to verify and identify the person's identity. Gait recognition is one of the important biometric tools that aims to recognize individuals (from distance) by their style of walking, psychological studies show that people have a reasonable ability to recognize other individuals by the way they walk [1]. A unique advantage of gait, as a biometric measure, is that it potentially offers recognition from a distance at low resolution whilst requiring no user cooperation, whereas other biometrics are likely to require a certain level of cooperation [3]. Gait recognition systems are based on analyzing video sequence recordings. Microsoft has produced a device called the Kinect sensor which supports this kind of analysis, as described below.

Microsoft Kinect is a product used with XBox 360 gaming console, with this device a participant can control games by body motion without carrying any other sensor. Kinect also enables the tracking of different types of information involving for example: the human skeleton, face geometry, and also creating different depths of the captured image. Using Kinect SDK for Windows with a set of APIs provides a very efficient tool to develop a gait recognition application [4].

Nowadays, most gait recognition applications use normal video cameras to record video sequences of people walking. Using these traditional videos, however, causes many difficulties when extracting features. In [5] and [6], features are manually determined as dot points on the subject in each frame. All these motion based methods present difficulties in extracting the features precisely, where Kinect can provide these feature more precisely, as well as providing details of the body skeleton.

2. RELATED WORK

Different ways of recognizing individuals by their gaits have been widely discussed in the literature for many years. The first work in this area was carried out by psychologists in 1971, when Johansson attached light points to the joints of people's bodies in a dark area. Participants then were asked to walk, run, or ride a bicycle [7] and [8]. The results suggested that people can recognize each other by their individual walking styles. Work carried out by Perry in [9] and Kumar et al. in [10], led to the hypothesis that gait is a characteristic that can be used as a biometric tool to identify people.

Gait recognition approaches can be classified into two models: motion -based (model free), and model-based. Model-free is based on the binary representation of a silhouette to generate spatio-temporal patterns and state-space; these are methods used for representing the human body as described in [11] and [12]. In [11] mass vector was used as a feature that is based on the number of 1's representing the object. The authors used Dynamic Time Warping (DTW) for the matching scheme to overcome the problem of direct frame-by-frame matching, because people do not always walk using the same speed and style. Other researchers, however, in [12] proposed a different method of using a motion based. Here, they relied on step-forward and step-following templates, based on the Frame Difference History Image (FDHI).

Model-based approaches use the structure of the human body and its motion as a model, utilizing information from body parts such as the foot, knee, ankle, hip, wrist shoulder, torso, thigh, head, and hand, for extracting features [1], [3], [5], [6] and [13]. Measuring lengths and distances between different parts of the body has been used for constructing gait features. In [13] up to 20 subjects were proposed using simple activity-specific parameters (height, torso length, leg length, and step length) extracted from the double support phase (when both feet are on the ground) of the gait cycle. Singh and Jain [5] proposed dynamic features for representing the human body by constructing two triangles between three points: the first triangle created was based on the hands and the right and left heel, the second triangle used the same hands with the right and left toes, the third triangle was constructed by defining the intersection points of the two triangles and hand. They then calculated the mean value of the angles created based on the new triangle for each subject in one gait cycle. The authors of [6] used dynamic features to represent the human body by using one of the hands and both feet as a feature. The concept involves inserting dots on selected parameters of the subject in all frames, drawing a triangle among points, then calculating each angle of the triangle and finding the mean value of one cycle. Papers [5] and [6] used CASIA database A, that recorded data using a standard video camera, inserting dots manually on subjects to determine the points for extracting features.

Conversely, many research work were based on using Kinect sensor supported by its SDK. The authors of [1], [3], and [14] used Kinect sensors to record data to track 20 points in human skeletons from the head, through the hip, to the foot. A Kinect sensor is also used in different applications such as games, health care and security surveillances, Parajuli et al. in [15] used a Kinect sensor to present a health monitoring system of seniors to collect data on posture recognition (sitting versus standing) and gait recognition (normal walking versus abnormal walking). Preis et al. in [1] describes the extraction of 14 biometric features for representing the human body, using 12 static features (i.e. height, torso, both lower legs, right and left upper arms, the length of both upper legs, both thighs, both forearms), and 2 dynamic features (i.e. the step length and speed) for 9 subjects who were asked to walk from the right to left side view of the Kinect sensor, 8 times per person. They then used 3 different sets of features: the first set included the height, length of legs, length of the torso, and length of the left upper arm, the second static features (which are not used in the first set), and the third set used only dynamic features. Three different classifiers were used R1, C 4.5 and Naive Bayes these features achieved 62.7%, 76.1% and 90 % rates of recognition respectively, which comprises the best set of features compared with the recognition rates of the other two sets. Furthermore, Kumar and Venkatesh [3] claimed a recognition rate of over 90% with 20 subjects, again using skeleton points, recorded by a Kinect sensor. In this study, each subject walked towards the camera 10 times in 2 different scenarios (i.e. with a fixed and a then a moving camera). Five different parts of the skeleton were selected as a feature (i.e. spine, left arm, right arm, left leg and right leg).

The authors of [16] proposed 18 dynamic features, based on the changing angle in the lower limb joints. They then calculated the mean, maximum and standard deviation of the three angles for both legs in the half gait cycle and achieved a 43.6% recognition rate. Other researchers in [14] proposed 2 sets of features: the first set included static features comprising joints from the upper body (i.e. shoulder center, shoulder left, hip left, hip center, hip right and shoulder right) and from the lower body (i.e. hip center, hip right, knee right, ankle left, knee left and hip left). The second set consisted of 4 distances measured between the hands and legs, taken as the dynamic features. Using the Adaptive Neural Network for selection and classification of features resulted in the highest recognition rate of this approach of 86%. In our proposed method we will use a Kinect sensor for constructing 2 sets of

dynamic features; the Horizontal Distance Feature (HDF) and the Vertical Distance Feature (VDF). These features are extracted based on data related to the feet, knee, hand, shoulder, hip and height during one gait cycle.

3. GAIT RECOGNITION PROPOSED METHOD

In this paper, we propose a model for gait recognition based on the skeleton that is provided by the Microsoft Kinect sensor. As explained before, a Kinect sensor supported with its SDK provides a high quality of a human skeleton for up to two persons. Kinect also provides RGB image and depth image, but we used a skeleton model only. In general, our system consists of four stages: firstly the creation of an application for skeleton recording and creating a database by recording skeleton information using Kinect, secondly the detection of a gait cycle for each subject, and thirdly the extraction of features in two different sets: VDF and HDF, this will be described in more details later. Finally, K-Nearest Neighbour (KNN) is used as a classification method, as shown in figure 1.

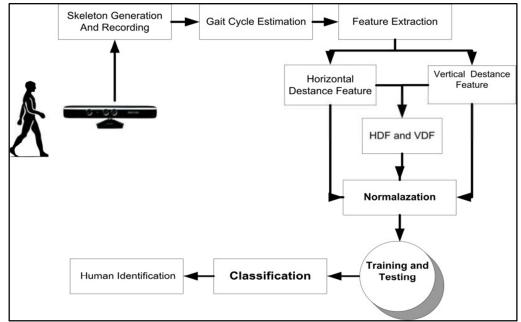


Figure 1.Over view of gait recognition system.

3.1.System Review

Kinect technology is a human machine interface that has been attracting more interest from researchers. In this paper we propose gait recognition system based on Kinect. Kinect is providing 20 useful points (see Table 1) presenting human skeleton from head over hip to foot more precisely that can have a major role in gait recognition as compared to other types of cameras.

| Joint Number | Joint Name | Joint Number | Joint Name | Joint Number | Joint Name | | | |
|-----------------|--------------------|-----------------|-------------------|-----------------|-------------|--|--|--|
| 1 | Hip Center | 8 | Hand Left | 15 | Ankle Left | | | |
| 2 | Spine | 9 | Shoulder Right | 16 | Foot Left | | | |
| 3 | Shoulder Center | 10 | Elbow Right | 17 | Hip Right | | | |
| 4 | Head | 11 | Wrist Right | 18 | Knee Right | | | |
| 5 | Shoulder Left | 12 | Hand Right | 19 | Ankle Right | | | |
| 6 | Elbow Left | 13 | Hip Left | 20 | Foot Right | | | |
| 7 | Wrist Left | 14 | Knee Left | | | | | |

Table 1. Joint points name and numbers

3.2.Data Recording

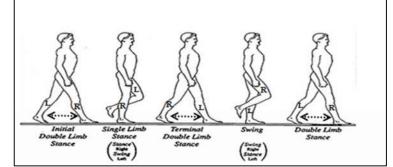
As the Kinect is relatively new there is no standard database available in the public domain, so it was necessary to build our own database using the Kinect camera before we started to conduct our experiment. We tested the accuracy of Kinect measurements of distances between Kinect and subject by recording 10 short records for 5 persons at different distances such as 3.5m, 3m and 2.5m. We then compared the accuracy of these distances. We decided the optimum distance between Kinect and the subject to be 2.5m because at this distance the skeleton information presents more clearly than at the other distances and provides more than one gait cycle. The height of the Kinect sensor was 0.6m, the angle of the Kinect was 0 degrees and it recorded at approximately 30 frames per second. For the experiment, 20 participants (16 males and 4 females) walked across in the front of the Kinect sensor from right to left at an angle of 90 degree. They were asked to walk normally 10 times, providing a total of 200 records, as shown in figure 2.



Figure 2. Experiment place

3.3. Gait Cycle Estimation

The gait cycle is a walking pattern which starts from a heel-strike of a single foot to the next same heel-strike. The gait cycle is classified into 2 parts: stance phase and swing phase. The stance phase is the first part of the cycle that comprises 60% of the cycle and consists of 4 components (loading response, mid-stance, terminal stance, and pre-swing). The swing phase comprises the second, 40% part of the cycle of the cycle and is divided into 3 components (initial, middle and terminal swing) [17]; see figure 3.



http://www.oandp.org/jpo/library/popup.asp?xmlpage=1997_01_010&type=image&id=f2

Figure3. Representation of human walking in one gait cycle

Typically, there are two techniques for estimating gait cycle. The first technique is the mid-stance or local minima, where both feet are close together [18]. The second is the double support phase where the distance between the two feet is at the maximum. In our proposal we used the double support phase to determine gait cycle, measuring the distance between the ankles instead of feet for better accuracy.

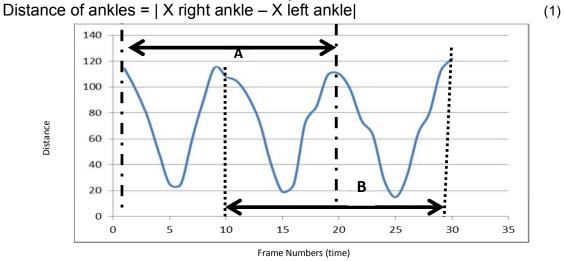


Figure 4: Gait cycle generation.

As explained above any gait cycle consists of three maximum distances between the feet. Thus, according to Figure 4, there is one gait cycle and one extra step, we can define "A" as a gait cycle and "B" also can be considered as another gait cycle, as shown in figure 4.

3.4. Feature Extraction

In our study the process of feature extracting is totally different from the techniques based on normal cameras because Kinect sensor is used for recording human skeleton information. As already mentioned, the Kinect sensor examines 20 joint points of the human skeleton, thus our dataset includes X-axis and Y-axis for 20 joints, totaling 40 attributes. Sixteen attributes have been selected for extracting features as follows: X and Y of both ankles, X of both feet, X of both knees, X of both wrists, X of both shoulders, Y of the head, Y of the right wrist, Y of the right shoulder, and Y of the center hip. Relying on these points, we have extracted two sets of dynamic features: the first set is HDF, and the second set is VDF. These two sets will be explained below.

3.5.Horizontal Distance Feature (HDF)

The Horizontal Distance Feature is based on measuring the changes of distances between skeleton joints according to the X-axis during one gait cycle. We proposed four features; the step length (HD1), the distance between right and left knees (HD2), the distance between right and left wrists (HD3), and the distance between right and left shoulders (HD4), as shown in fig. 5. Here, instead of measuring the distance between the left and right feet to detect step length we used the left and right ankles for better accuracy. Then we calculated the mean, skew and standard deviation for each of these measurements in one gait cycle. For creating the feature vector we used the mean and standard deviation to represent the temporal changes of distances in all frames during one gait cycle. The following set of equations represents the construction of the feature vectors (figure 5).

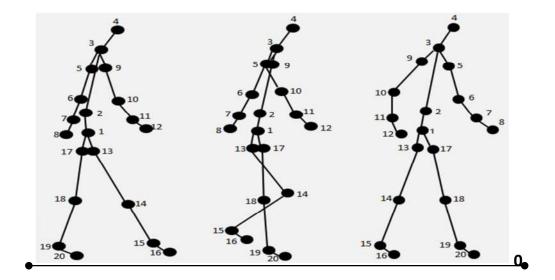


Figure 5: Horizontal and Vertical distances extracted from the skeleton.

| HD2= abs (D (Jx (18) - Jx (14))) | (3) |
|----------------------------------|-----|
| HD3= abs (D (Jx (11) - Jx (7))) | (4) |
| HD4= abs (D (Jx (9) - Jx (5))) | (5) |
| MeanH= mean {HD1, HD2, HD3, HD4} | (6) |
| StdH= Std {HD1, HD2, HD3, HD4} | (7) |
| SkwH = Skew {HD1, HD2, HD3, HD4} | (8) |
| HDF = {MeanH, StdH, SkwH} | (9) |

Where D is distance, Jx is horizontal joint coordinate

3.5.1. Vertical Distance Feature (VDF)

This set of features is based on changes in the distance between a skeleton joint and the ground according to the Y-axis. Here we proposed six features to represent the human body, as follows: the participant's height (VD1), the height of right wrist (VD2), the height of right shoulder (VD3), the height of right and left ankles (VD4 and VD5). We then calculated the means and the standard deviations of all of these measurements in one gait cycle. Finally, we created a triangle based on the hip-center and the distance between right and left Feet (VD6). The mean of VD6 was calculated during one gait cycle and added to the feature vector. The following set of equations represents the construction of the feature vectors (figure 5).

| VD1= D (Jy (4) - Ly (0)) | (10) | | | |
|--|------|--|--|--|
| VD2= D (Jy (11) - Ly (0)) | (11) | | | |
| VD3= D (Jy (9) - Ly (0)) | (12) | | | |
| VD4= D (Jy(19) - Ly (0)) | (13) | | | |
| VD5= D (Jy (15) - Ly (0)) | (14) | | | |
| VD6= 1/2 *(Jx(20) – Jx(16)) * Jy(1) | | | | |
| MeanV= mean {VD1, VD2, VD3, VD4, VD5, VD6} | | | | |
| StdV= Std {VD1, VD2, VD3, VD4, VD5} | | | | |
| VDF = {MeanV,StdV} | (18) | | | |
| Where D is distance. Iv is vertical joint and I v is land line | | | | |

Where D is distance, Jy is vertical joint and Ly is land line.

4. EXPERIMENTAL RESULTS

We tested our method using the database that we created, this human skeleton information dataset was recorded by a Kinect sensor and contains data from 20 subjects captured from side views, and 10 records per subject were recorded. This database is a set of (x, y) coordinates that represent each joint point captured during each gait cycle. After collecting the data, data pre-processing tasks were carried out, such as cleaning data, gait cycle estimation, feature extraction, and etc. resulting in two sets of meaningful features that are ready to be processed by the recognition system.

In our approach, KNN was used as a classification method that depends on the shortest distance. After testing the K value from K=1 to K=10, we decided to set K to 1 in order to provide a better recognition rate.

We constructed 2 sets of feature vectors, HDF and VDF, as discussed above. We tested our system by using each set of feature vector separately and then combined both in one set. When HDF was used alone the recognition rate achieved 56%, while VDF provided 83.5% recognition accuracy; by combining HDF and VDF into one feature vector the recognition rate increased to 92% (figure 6).

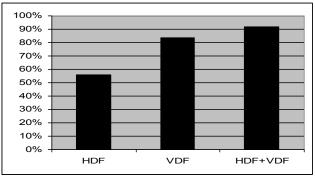


Figure 6: Recognition rate with the feature type.

5. COMPARISON WITH EXISTING WORK

Table 2 shows recognition rates reported in other studies of gait recognition based on Kinect sensors. These have been discussed in the related work section. Some of these studies tested different sets of features and used more than one classifier; we report only the highest accuracy rate they obtained.

| Related work | No. of Subject(person) | Recognition rate | |
|--------------|---------------------------|----------------------|--|
| [16] | 4 | 43.6% | |
| [14] | 5 | 86% | |
| | | 85.1% (all features) | |
| [1] | 9 | 91% (Static | |
| | | features | |
| Our Method | 20 | 92% | |

Table 2: Comparison According to Recognition Rate

Comparing our approach with those of other researchers, based on accuracy rates as a measurement is, however, problematic for several reasons. The first of these is that there is a lack of a benchmark database; no standard database is available in the public domain. Each researcher who used a Kinect sensor, including us, therefore created his or her own database; there are several differences between these databases in terms of the sample size, the number of records per person, the distance between subject and sensor, the height of the sensor, and so on; each of these measurements has an effect on the recognition rate. Secondly, different pre-processing approaches were used; some researchers used the double support phase for estimating gait cycles, others used mid-stance, which may have produced different results. The use of various classification techniques is another factor making comparisons inconsistent. In the case of non-existing standard databases, the best way to improve comparisons between our and other researcher's approaches is to reimplement their approach into a framework and test it in our database. Due to

time constraints, we implemented the same set of static features that were proposed in [1] and, compared to the proposed features using our database, we found that their technique provided 71% recognition rates, whilst ours provided 92%.

Our approach seems to be practical and has the following positive points: firstly, we used our database that includes skeleton information for 20 people. Secondly, we proposed two sets of new and meaningful features for human recognition.

6. CONCLUSION AND FUTURE WORK

We proposed a gait recognition approach based on skeleton point trajectories obtained from a Kinect sensor. This sensor examines a new area of features and provides more robust skeleton information for gait recognition. In our approach we proposed two sets of features: HDF and VDF; these were classified separately by using KNN with CityBlock as a distance function, and were then combined to form one set of features. The recognition rate achieved for HDF, VDF, and fusing both (HDF+VDF) were 56%, 83.5%, and 92% respectively. The result outperformed the existing techniques, for two reasons; firstly by adding a new set of features that depends on the vertical rather than the horizontal feature (it should be noted, however, that these features cannot be extracted for a traditional gait silhouette), and secondly by identifying the feature point more precisely compared with the techniques that used traditional cameras.

Since gait recognition by using Kinect sensors is relatively new there are many issues that need further research. We summarize our future work as follows; Extending current features by extracting new features and using a different type of classifier to achieve an improved recognition rate, proposing approaches for gender classification based on human gait using Kinect sensors, Proposing a technique to overcome gait recognition challenges, i.e. establishing how carrying objects or wearing different clothes affects normal walking; we will try to find a method to improve the performance in these situations, and extending the current dataset and creating a new one based on the Kinect sensor locating different Variation, walking directions, and angles

REFERENCES

- [1]. Preis, J., et al. Gait recognition with kinect. in 1st International Workshop on Kinect in Pervasive Computing. 2012.
- [2]. Das, R., An introduction to biometrics. Military Technology, **29**(7): p. 20-27 2005..
- [3]. Kumar, M. and R.V. Babu. Human gait recognition using depth camera: a covariance based approach. in Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing. ACM 2012.
- [4]. Jana, A., Kinect for Windows SDK Programming Guide: Packt Publishing 2012.
- [5]. Singh, J.P. and S. Jain. Person identification based on gait using dynamic body parameters. in Trendz in Information Sciences & Computing (TISC), IEEE 2010.
- [6]. Arun Kumar Jhapate and J.P. Singh, Gait Based Human Recognition System using

Single Triangle. IJCST, p. 128-131:2011.

- [7]. Kozlowski, L.T. and J.E. Cutting, Recognizing the sex of a walker from a dynamic point-light display. Perception & Psychophysics, 21(6): p. 575-580 1977.
- [8]. Cutting, J.E. and L.T. Kozlowski, Recognizing friends by their walk: Gait perception without familiarity cues. Bulletin of the psychonomic society, 9(5): p. 353-356 1977.
- [9]. Perry, J. and J.M. Burnfield, Gait analysis: normal and pathological function: Slack 1993.
- [10]. Bharatkumar, A., et al. Lower limb kinematics of human walking with the medial axis transformation. in Motion of Non-Rigid and Articulated Objects, 1994., Proceedings of the 1994 IEEE Workshop on. IEEE 1994.
- [11]. Hong, S., et al. A new gait representation for human identification: mass vector. in Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on. IEEE 2007.
- [12]. Lee, C.-C., et al. Frame difference history image for gait recognition. in Machine Learning and Cybernetics (ICMLC), 2011 International Conference on. IEEE 2011.
- [13]. Bobick, A.F. and A.Y. Johnson. Gait recognition using static, activity-specific parameters. in Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on.. IEEE 2001.
- [14]. Sinha, A., K. Chakravarty, and B. Bhowmick. Person Identification using Skeleton Information from Kinect. in ACHI 2013, The Sixth International Conference on Advances in Computer-Human Interactions. 2013.
- [15]. Parajuli, M., et al. Senior health monitoring using Kinect. in Communications and Electronics (ICCE), 2012 Fourth International Conference on. IEEE 2012.
- [16]. Ball, A., et al. Unsupervised clustering of people from 'skeleton'data. in Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on. IEEE 2012.
- [17]. Sabir, A., N. Al-jawad, and S. Jassim. Gait recognition using spatio-temporal silhouette-based features. in SPIE Defense, Security, and Sensing. International Society for Optics and Photonics 2013.
- [18]. Lee, B., et al. Gait recognition system using decision-level fusion. in Industrial Electronics and Applications (ICIEA), 2010 the 5th IEEE Conference on. IEEE 2010.

پوخته

ئەم توێژینەومیە باس لە ناسینەومی خەللک دمکات بە گوێرمی شێوازی رێکردنیان, بە پشتبەستن بە ئێسکە پەیکەری مرۆڭ و جومگەکان, کە بەئامێری (Kinect Sensor) وێنه گیراوه. لەم کارمدا پشتمان بە دوو کۆمەللە لە شێوازی جولاو (Dynamic Feature) بەستوم کە لە يەك (gait cycle) دەرھێندراون. كۆمەللەی یەکەم بەناوی (HDF) واتا دووری ئاسۆیی, پشت بە دووری نێوان (پێيەكان,ئەژنۆكان,دەستەكان , شان) دەبەستێت.كۆمەللەى دووەم دوری ستونی (VDF) كە كۆمەلێك زانیاری گرنگ لەسەر شێومی جولامی مرۆڭ دەردەخات لە ڕێی پێوانە کردنی دووری نێوان سەر رووی زەوی و (دەست,شان, پێ).

(4) له رهگەزى مىٰ بوو كه بۆ ھەر بەشداربويەك (10) گرتەى فيديۆيى بەكارھاتووە.

بۆ گەيشتن بە ئەنجامەكان سێ سيناريۆى جياواز بەكارھاتووە, يەكەم سيناريۆ بەكارھێنانى كۆمەڵەى(HDF) كە رێژەى ناسينەوە تێيدا گەيشتە(٪56), سيناريۆيى دووەم (VDF) بەكارھێنرا كە رادەى ناسينەوى مرۆڭ تێيدا (٪83.5),سيناريۆيى سێيەم تێكەلكردنى ھەردوو سيناريۆيى پێشوو پيكەوە بەكار ھێنران كە رێژەى ناسينەوە ووتێيدا گەيشتە(%92).

ملخص

تتطرف البحث الى تعرف الى الانسان من حيث كيفية المشى اعتمادا على الهيكل العظمى و مفاصلة. صورت بكاميرا من نوع (Kinect Sensor) اعتمدنا فى هذة البحث على مجموعتين من اساليب المشى يخرجان هاتين المجموعتين من دائرة المشى واحدة.

المجموعة الاولى التي تسمى ب (HDF) تعتمد على مسافة بين (الكعب, الركبة, الايدى,الاكتاف).

المجموعة الثانية ولتى تستند على البعد العمودى(VDF) ولتى تستخرج مجموعة المعلومات هامة من اسلوب حركة الانسان اعتماداً على الابعاد بين سطح الارض و اليد,الاكتاف و القدم.