Real-time Rehabilitation and Fitness System using Depth Sensor

Syyam Noor
Department of Computer Science
COMSATS University Islamabad
Lahore 54000, Pakistan
syyamnoor@gmail.com

Muhammad Hasham
Department of Computer Science
COMSATS University Islamabad
Lahore 54000, Pakistan
muhammad.hasham.2311@gmail.com

Zulfiqar Habib
Department of Computer Science
COMSATS University Islamabad
Lahore 54000, Pakistan
drzhabib@cuilahore.edu.pk

Allah Bux Sargano
Department of Computer Science
COMSATS University Islamabad
Lahore 54000, Pakistan
allahbux@ciitlahore.edu.pk

Abstract—**This paper presents an action verification system for rehabilitation of patients with limited mobility. In this regard, a kinect sensor is used to capture the image-sequences of patients for identification of different actions in real time during rehabilitation exercisers at home. Then, a set of most discriminating features is extracted from a human skeleton. These features are fed to the classifier for action recognition. In order to evaluate the performance of the proposed system, we build a data set for patients with frozen shoulders and Overhead shoulder issues. The results confirm an excellent performance of the proposed system. In this paper, our contribution is three fold. 1) we have constructed a data set of overhead shoulder exercise performed by 8 different actors. 2) A novel set of features has been proposed for verification of these two exercises in real time. 3) we have employed 5 different classifiers for classification and presented a comparative analysis.**

Index Terms—**Computer Vision, Rehabilitation, KNN, Machine Learning, Classification**

I. Introduction

Physical Therapy or Physiotherapy is a health care profession, in which Physical Therapist provides treatment to individuals for developing, maintaining and restoring the maximum functional capacity throughout life. Human movement is considered vital to the health and well-being of individuals in Physio therapy, the history of physio therapy and the processes involved has been discussed in brief [1]. Physiotherapists work in different areas such as Outpatients Dept, Intensive care unit, Woman's health, Orthopedics, Occupational health, terminally ill, Neurological rehabilitation [2,3]. Pakistan is a nation under growth, most of the health care departments are under the course of advancement. Physiotherapy is amid one of the most ignored health care fields in this country due to the lack of reforms and knowledge [4]. In accordance to 2012 statistics, nationwide disability rate was 2.65% which accounts for around 5 million individuals with disabilities (PWDs) in Pakistan. Rural areas have more PWDs (65.7%) than urban areas (34.3%). [5]. Rehabilitation exercises are recommended by the physiotherapists and surgeons to reduce pain or regain muscle movements for the patients who have undergone surgery, have Frozen Joint syndrome, and Adhesive Capsulitis that causes decreased range of motion (ROM) in joints. With proper treatment like steroid joint injections, heat and cold therapy and physical exercises this syndrome can be treated. Physiotherapists normally recommend movement and exercise to help improve in a person's mobility and function. This include: exercises designed to improve movement and strength in a specific part of the body these usually need to be repeated regularly for a set length of time. Mobility exercises play a very vital role in a quick recovery from frozen joints syndrome and they help restore strength and muscle movements back. Seeking professional help in this regard is very important, because a medical professional understands the critical details of various movements and exercises. Technological advancement in rehabilitation domain has emerged over the past few years. There have been various advancements made in the field of rehabilitation related mobility exercises that allows patients to regain their muscle movements. The main motivation behind making this motion capture application is to assist patients as well as medical experts in performing and monitoring the patients exercises, respectively. Motion capturing is the process of recording the human body movement and to animate the digital character model in 2D or 3D. There are 2 types of motions capture systems: optical and non-optical. The non-optical systems are further divided into 3 categories: inertial systems, mechanical motion systems, and magnetic systems. Among these, inertial motion capture is widely used, it uses IMUs (inertial measurement units) with built in sensors to sense the movement and position. These typically comprise of accelerometers, gyroscopes and occasionally magnetometers. Similarly, the optical systems are further divided into 2 sub categories: with markers and without markers capture system.

978-1-7281-6625-4/19/$31.00 ©2019 IEEE
DOI 10.1109/FIT47737.2019.00048

Optical motion capture demands different cameras placed around the object, and reflective sensors are placed on the object at the main points of concern. The sensor throws an infrared light that is reflected off of the reflective sensors. The reflected beams are then perceived by the motion capture camera. More than one camera makes a 3D model of that respective object. Thus, more cameras mean better outcome. The marker less capture method is currently being widely used for motion capture in the area of computer vision. The main advantage of using Marker less systems is that you do not require objects to wear special equipment for tracking. A Kinect sensor is great example of marker less capture system. Microsoft Kinect Sensor, has been widely used in various applications of computer vision. Comparing to other marker less capture system, Kinect uses a comparatively cheaper IR RGB-D sensor to get the real-time depth information of the environment. Essentially, it operates just like any other 3D structured-light depth sensor, and 3D coordinates of an object in the environment. Kinect sensor was initially made for video games for X-Box gaming console by Microsoft, the technology has become so popular among computer scientists that people have started using it widely for different fields of studies, such as object reconstruction [6] or augmented reality applications [7]. The kinect sensor helps in getting the coordinates of joints in a human skeleton. That is the reason why we have opted kinect device for our experimentation. There are other rehabilitation systems as well, but they require markers and sensor gloves for data acquisition that adds up to cognitive load of patient and extra burden [8]. Kinects camera detects 25 points *figure 2* on body without the need to wear any physical devices e.g sensors, electronic gloves, physical markers or any other hardware. As far as the precision is concerned, its accuracy matches shoulder to shoulder with the above mentioned techniques, not only it removes hurdles in treatment by adding real-time interaction with the system but also reduces the cost associates to traditional rehabilitation. There has been some work done in dealing with the fitness/health related issues with the help of technology, for example: Tai-chi and karate [9], dancing actions (k-pop) [10], sports actions (cricket, baseball, tennis etc.) [11], and dance learning [12], but nobody has yet performed the validation of mobility exercises required by patients with frozen joint syndrome. Moreover, our interface will consist of a single motion capture camera, where most of the previous work done related to this field has been carried out using multiple depth sensors.

Another way to solve this problem is to call a home therapist, but again apparently it is also not financially feasible [13]. A better solution which people can think of is the process of tele-rehabilitation, in which patient can receive rehabilitation services over internet and telecommunication network. With this approach, patients can follow and act upon the instructions from remote rehabilitation therapist or memorandums from rehabilitation centers to perform rehabilitation exercises at home on advised days. [14].

In order to accomplish this, we curated a data set to carry out our experiments. We are using eleven most discriminated features that will help us classify the correctness of the exercise. These features are then shaped into feature vectors and are fed to a various ML models for classification. We have designed a BP-Artificial neural network, and also used SVM, KNN and random forest classifiers for classifying the results. In literature, we infer that KNN model completely outperforms its counterparts. The detailed results are given in the related sections.

The remainder of this paper is organized as follows. In Section 1, we describe the proposed methodology and the process of image acquisition and feature extraction steps. In section 2, we present the description of our data set and feature selection. In Section 3, we present the description of different classification techniques that we carried out. In Section 4, we present the experimental results for baseline comparisons between different classifiers on our data set. Finally, Section 5 includes our conclusion.

II. Proposed Methodology

Our proposed methodology is based on a sequence of actions. Images of human body are acquired from Kinect. To select distinct features we use feature extraction which comes handy in calculating angles and distances between joints. The feature vector is provided to a proposed classifier. The classifier is trained before using augmented data set. This helps us to obtain classified results for exercise. The underlying block diagram (Fig 1.) shows the sequence of steps.

A. Image acquisition

The Kinect device contains two essential components that work collectively to acquire image, this includes: VGA video camera (*RGB* color), a depth sensor. The video camera detects *RGB* components as well as physical features. It has a resolution of 640x480. The depth sensor on the other hand contains a monochrome CMOS sensor and infrared projector that makes a 3D projector throughout the controlled environment. In this section, just like any vision system we will do image acquisition. Here, we describe the process of finding angles and distance between body parts that includes shoulders, wrists, spine and arm joints of the human skeleton. The angle measurement process consists of three steps. First X, Y and Z coordinates are obtained from human skeleton joints. Next, relative position of each coordinate of joints are obtained and dot product is calculated. Finally, vectors are formed and their magnitude is used to calculate angle. On the other hand, for distance we calculate X, Y and Z coordinates of joints on which we apply Euclidean distance formula. These above mentioned X ,Y and Z coordinates are calculated using Kinect depth sensor.

B. Calculation of angles

In the first stage of angle calculation, X, Y and Z coordinates are calculated. The coordinates after calculation are used to calculate relative position of each joint with respect to other joints.Vectors are calculated with these coordinates.

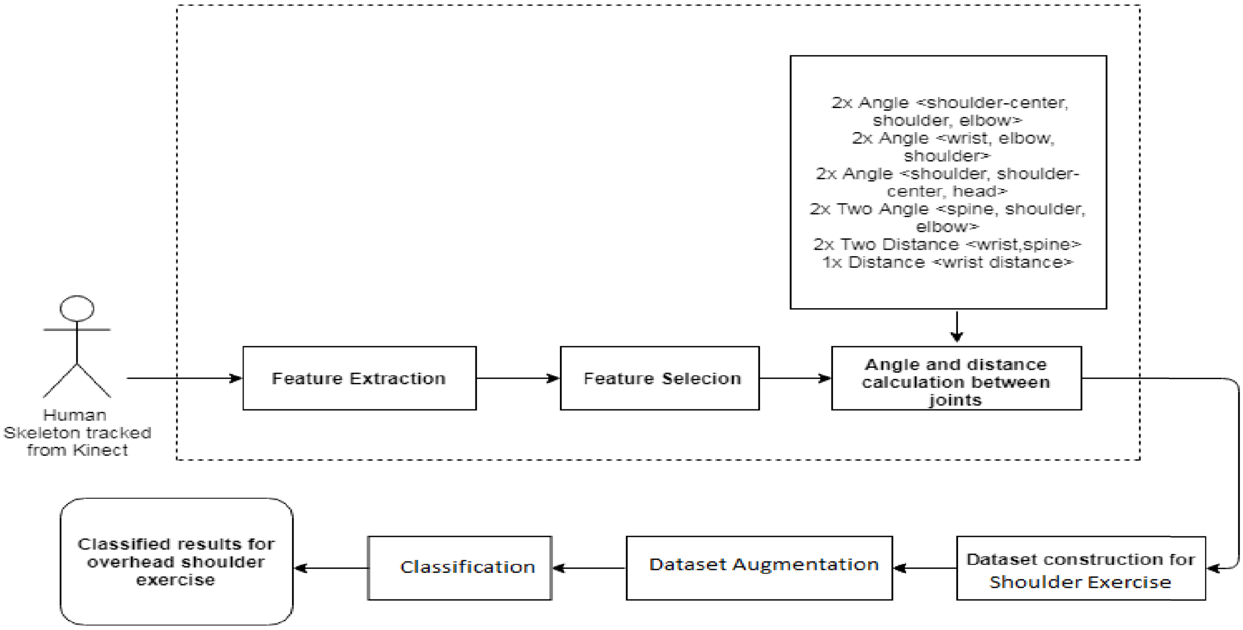

Fig. 1. Block Diagram of Real-time Rehabilitation and Fitness System

Following are the equations used for calculating angle based attributes:

$$|X| = [(x_1 - x_2) * (x_3 - x_2)] \tag{1}$$

$$|Y| = [(y_1 - y_2) * (y_3 - y_2)] \tag{2}$$

$$|Z| = [(z_1 - z_2) * (z_3 - z_2)] \tag{3}$$

$$X_p = |X| + |Y| + |Z| \tag{4}$$

$$X_{m1} = |\sqrt{x_1^2 + y_1^2 + z_1^2}| \tag{5}$$

$$X_{m2} = |\sqrt{x_2^2 + y_2^2 + z_2^2}| \tag{6}$$

$$X_t = X_p/(X_{m1} * X_{m2}) \tag{7}$$

The resultant angle we get is:

$$X_{angle} = (180/\pi) * \arccos(X_t) \tag{8}$$

In *Equation 1 - 4* we are calculating angle between 3 points in a human body with the help of Pythagoras theorem.

The joints that are extracted from human skeleton are shown in figure 1. The dot product of magnitude of vectors is calculated to make sure angles are precise. As in *Equation 5* we use cosine function. As cosine calculates angle between adjacent side and hypotenuse as we are dealing with three dimensional (3D) domain. The cosine formula associates all pairs of vectors in the space with the inner product of the vectors which is scalar in nature. The inner product would allow the to calculate intuitive geometrical dimensions such as the length of a vector and the angle between two vectors. It would also assistance to calculate of orthogonality between vectors (zero inner product). Angles are taken for minimum and maximum range for a specific posture.

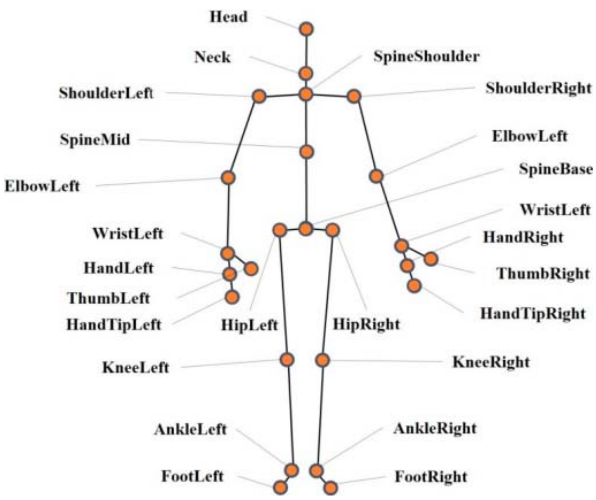

Fig. 2. Joints detected from human skeleton by Kinect sensor

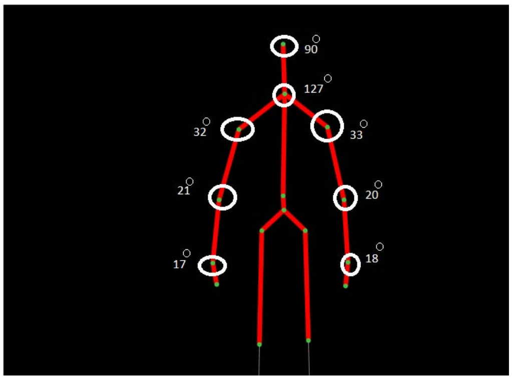

Fig. 3. 8 core angles calculated to distinguish each movement.

C. Calculation of distance

The next step includes the calculation of the distance between concerned body parts. Various distance measures are calculated from one body part to another. The incorporated Kinect sensor uses an infrared camera to produce a depth image that can be used in a controllable environment [16]. The depth image detected for calculating distance is processed from which joint data is extracted. The joints are used to calculate body proportions and in our case distance. Kinect has a field of view which is visible which is used by depth sensor which produces subsequent depth frames [17]. Each depth frame is a grid of 512424 points. With the help of calculated depth frames, a total of 25 human body joints are identified which are used to calculate the coordinates in the 3D space. Since, Cartesian coordinate system is used, every single joint has 3 values which are *X*, *Y*, and *Z*. The projection is done in a Cartesian coordinate system. The (0, 0, 0) point is the position of the sensor. Every other point is measured in terms of the position of the sensor. The first step is to obtain *X*, *Y* and *Z* coordinates from each joint that is detected by Kinect of human skeleton. The *Z* coordinate is not the linear distance between the point obtained from joint and the Kinect sensor. Apparently, it is the distance between point and the sensor. After it has been calculated the relative position of each coordinate is taken with respect to other coordinate system. Finally, Euclidean distance comes into play.

Euclidean distance between points x and y

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2} \quad (9)$$

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (10)$$

We use Euclidean distance to calculate distance of one body part relative to another. Euclidean distance is preferred in this scenario because it remains same in all directions and is rotation variant [18]. The illustrated figure 2 shows the distance being calculated.

III. Description of Data set

The data set was curated of different postures related to frozen shoulder exercise. For diversification, various people with different physical anatomy participated to prepare the data-set. The dataset consists of 1500 training instances, the extra data-points are imputed by the data augmentation algorithm (GAN), given some of the features, they predict the rest, the data-set augmentation has been described in the next subsection. The data set is further divided into train, test data-set with a 70-30 ratio, respectively. There is a balanced amount of positive and negative instances i.e. 51 percent correct and 49 percent incorrect. The data-set consists of total 11 distinct attributes of angles and distance measures. There are eight angle measurement attributes and three distance measurement attributes representing important motion features from a set of markers in each frame. These 8 core angle attributes and 2 distance attribute can be reduced to 4 and 1 respectively by using dimensionality reduction algorithm (PCA+LDA) but that's not the main focus of the experiment at the moment so we can leave it for later. Thus, it reduces the risk of over fitting and decreases computational complexity [19]. The values of angle-based attributes are in degrees and the distances are in meters that our Kinect sensor provides. Each training instance has a correct label attribute, which is a binary valued attribute that depicts whether the exercise being performed is correct or not. The angles and distances used are highlighted in *figure 3*.

The attributes of data set as highlighted in figure 3 are mentioned as follows:

- Two Angles $<$ shoulder-center, shoulder, elbow $>$
- Two Angles $<$ wrist, elbow, shoulder $>$
- Two Angles $<$ shoulder, shoulder-center, head $>$
- Two Angles $<$ spine, shoulder, elbow $>$
- Two Distances $<$ wrist,spine $>$
- One Distance $<$ wrist distance $>$

A. Data set Augmentation

Generative Adversarial Networks or GANs are one of the most fascinating areas in deep learning research and development due to their unbelievable ability to generate artificial results, specially while augmenting training data [20]. In this experiment, we have used GAN to build synthetic training data points as our data set is very small and learning from small data is a major issue in Deep Learning. One of the first questions we have with this approach is the association between the initial data set size and the number of synthetic data points that the GAN adds to the data set. We have 1650 real data points that we manually observed from people of different shapes and sizes, we used GAN to extend the data points to 5000. We can generate even more if we mix up the classic data augmentation methods with GAN data augmentation. We can use these hybrid methods to produce a larger data set. We deployed the GAN to produce 3500 more instances. Now we can deploy translational shifts on all

Model	Accuracy
KNN	0.86
SVM	0.85
Logistic regression	0.84
LinearSVC	0.73
BP-ANN	0.81

TABLE I
COMPARISON OF ACCURACY SCORES

instances, (including those generated by GANs), and increase our data set by three times.

B. Significance of using more angle-based attributes

The data set consists of 8 angle-based attributes out of 11 input attributes that constitutes in discriminating the correct from incorrect posture. Each angle is calculated from body joints that are extracted by Kinects depth sensor. The reason behind using more angles-based attributes than distance-based attributes is because angles because angles between body joints are independent of human body physique. Angle measurements would remain consistent irrespective of individual on the other hand the distance may vary between body joints as the height of patient's body will not always the same. This gives an added advantage and also justifies the reason of using more angle-based attributes in our data set. The 3 distance-based attributes that we are using are the most important ones as they count the most in overhead shoulder exercise and are almost consistent for every person.

IV. COMPARISON AND EXPERIMENTATION RESULTS

We carried out the classification on the dataset using different techniques, such as KNN, SVM and BP-ANN. The experimental results revealed that the KNN gave the best results with high accuracy. Among conventional methods for classification of our overhead shoulder exercise, such as SVM, Logistic regression, LinearSVC and KNN, the KNN worked the best with 86 percent accuracy. BP-ANN had a relatively lesser accuracy score, the reason being is that we dont have much data and according to research if we have small learning data we should always go for traditional learning algorithms such as: SVM, KNN etc. The table compares the results. As listed in Table 1, the proposed method generally led to better classification results than its counterparts alone.

A detailed comparison is shown is *Table.1*. Which compares classifiers and its counterparts along-with BP-ANN.

A. Description of used classification techniques

We used *KNN*, *SVM*, *Logistic Regression*, *LinearSVC*, *Random Forest* and *BP-ANN* to compare evaluation results.

1) KNN: KNN is based on nearest neighbour in the vicinity. We calculated the value of K from 1 to 20, to check at which value of K do we obtain the best results. When K=6 the prediction results are most accurate. For KNN the range of K was altered from 1 to 20, the results can be seen in *figure 4*. We chose k=6 for our experiment as low value of k produce model of low bias, high variance and high value of k produce a model of high bias and low variance best model is found in the middle. We are getting 86 percent accuracy and Average precision-recall score of 0.76.

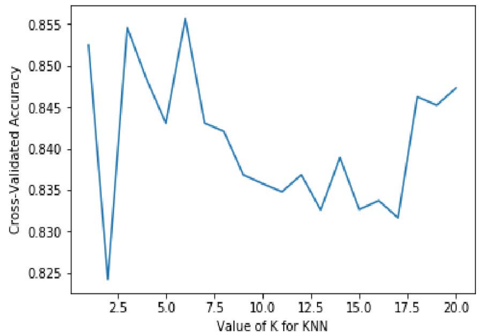

Fig. 4. Accuracy score for different values of K

2) SVM: SVM considers the outliers or the data points that are closest to the opposing class. When we used our data set on this classifier, the results were not significantly changes. The accuracy and average precision recall scores were almost similar to that of KNN's.

3) Logistic Regression: Logisitic Regression gave no more significant results or dropped accuracy. The results were almost same.

4) LinearSVC: LinearSVC however performed somewhat differently, the tolerance was kept 1e-4. The accuracy was dropped to 80 percent.

5) Random Forest: Random Forest which is based on decision trees gave results similar to that of KNN and SVM.

6) BP-ANN: Apart from the above mentioned classifiers, we evaluated BP-ANN on our data set as well. We made a multi-layer perceptron, with 11 feature input going into a 32-unit input layer and an output layer with "Sigmoid" as an activation function because it works well on binary classification problems. The model was trained with 15 epochs having batch size of 32. A pictorial representation of our back propagated neural network is also shown in *figure 5*.

V. CONCLUSIONS

We used Kinect to collect information of human skeleton joints. We compared different conventional methods and a BP-ANN for classification of our overhead shoulder exercise, such as SVM, RFCr, LR, LinearSVC and KNN, out of which the KNN and RFC gave the best results. The experimental results confirmed that our unique feature selection, selected features and our classifiers were able to classify the overhead shoulder exercise successfully. We got the highest accuracy on KNN (86 percent), the reason we got highest accuracy on KNN is because of the simplicity of our data set, if we get a large enough data set we can use BP-ANN. The positive results indicate that with a little bit alteration in the data set can help various applications of this kind, such as dance training

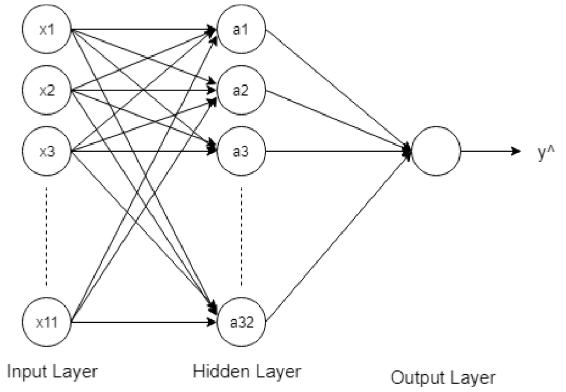

Fig. 5. Neural network representation

systems, bowling action validation systems and gym exercises systems etc. Our system is just focusing on on exercise for frozen shoulder, but we will design an all-encompassing physiotherapy rehabilitation system covering all the major mobility exercises that are required for the repossession of wronged joints.

ACKNOWLEDGEMENT

This research is supported by the PDE-GIR project which has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 778035.

VI. REFERENCES

1) Bruce K. "History of the School of Physiotherapy". School Physiotherapy Centre for Physiotherapy Research. University of Otago. 2007 website:[http://physio.otago.ac.nz/about/history.asp].
2) American Physical Therapy Association:"APTA Vision Sentence for Physical Therapy 2020[www.apta.org]. 2010, Web site:[http://www.apta.org/vision2020/].
3) Grant J. Learning needs assessment: assessing the need. BMJ 2002; 324:156-9.
4) Naveed Babur, Muhammad Rashid Siddique, Farah Awan, Waqar. (2014). Future Of Physical Therapy In Pakistan Satisfaction Amongst Pakistani Physical Therapists About Their Profession. Isra Medical Journal. 06. 25-27.
5) Statistics PBo. Disabled population by nature of disability. Pakistan Bureau of Statistics. [online] 2012 [cited 2014 January 7]. Available from: URL: http://www.pbs.gov.pk/content/disabled-population-nature-disability.
6) Nguyen, C. V., Izadi, S., Lovell, D. (2012, October). Modeling kinect sensor noise for improved 3d reconstruction and tracking. In 2012 second international conference on 3D imaging, modeling, processing, visualization transmission (pp. 524-530). IEEE..
7) Casas, X., Herrera, G., Coma, I., Fernndez, M. (2012). A Kinect-based Augmented Reality System for Individuals with Autism Spectrum Disorders. In Grapp/ivapp (pp. 440-446).
8) Shapi'i, A., Bahari, N. N., Arshad, H., Zin, N. A. M., Mahayuddin, Z. R. (2015, March). Rehabilitation exercise game model for post-stroke using Microsoft Kinect camera. In 2015 2nd International Conference on Biomedical Engineering (ICoBE) (pp. 1-6). IEEE.
9) Lee, J. D., Hsieh, C. H., Lin, T. Y. (2014, January). A Kinect-based Tai Chi exercises evaluation system for physical rehabilitation. In 2014 IEEE International Conference on Consumer Electronics (ICCE) (pp. 177-178). IEEE.
10) Kim, D., Kim, D. H., Kwak, K. C. (2017). Classification of K-Pop dance movements based on skeleton information obtained by a kinect sensor. Sensors, 17(6), 1261.
11) Choppin, S., Wheat, J. (2013). The potential of the Microsoft Kinect in sports analysis and biomechanics. Sports Technology, 6(2), 78-85.
12) Saha, S., Ghosh, S., Konar, A., Nagar, A. K. (2013, June). Gesture recognition from indian classical dance using kinect sensor. In 2013 Fifth International Conference on Computational Intelligence, Communication Systems and Networks (pp. 3-8). IEEE..
13) Lange, B., Chang, C. Y., Suma, E., Newman, B., Rizzo, A. S., Bolas, M. (2011, August). Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor. In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 1831-1834). IEEE.
14) LaBelle, K. (2011). Evaluation of Kinect joint tracking for clinical and in-home stroke rehabilitation tools. Undergraduate Thesis, University of Notre Dame.
15) Iqbal, T. Ali, H. J Med Syst (2018).Generative Adversarial Network for Medical Images (MI-GAN)
16) Correa, D. S. O., Sciotti, D. F., Prado, M. G., Sales, D. O., Wolf, D. F., Osorio, F. S. (2012, May). Mobile robots navigation in indoor environments using kinect sensor. In 2012 Second Brazilian Conference on Critical Embedded Systems (pp. 36-41). IEEE.
17) Khoshelham, K. (2011, May). Accuracy analysis of kinect depth data. In ISPRS workshop laser scanning (Vol. 38, No. 1).
18) Wang, L., Zhang, Y., Feng, J. (2005). On the Euclidean distance of images. IEEE transactions on pattern analysis and machine intelligence, 27(8), 1334-1339.
19) Sahoolizadeh, A. H., Heidari, B. Z., Dehghani, C. H. (2008). A new face recognition method using PCA, LDA and neural network. International Journal of Computer Science and Engineering, 2(4), 218-223.
20) Radford, A., Metz, L., Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.