Kinect-Based Physiotherapy and Assessment: A Comprehensive Review

Fadilla Atyka Nor Rashid, Nor Surayahani Suriani, Ain Nazari
Department of Computer Engineering, Faculty of Electrical and Electronics Engineering
Universiti Tun Hussein Onn Malaysia, Batu Pahat 86400, Johor, Malaysia

Article Info	ABSTRACT
	Kinect-based physical rehabilitation grows significantly as a mechanism for clinical assessment and rehabilitation due to its flexibility, low-cost and markerless system for human action capture. It is also an approach to provide convenience for for patients' exercises continuation at home. In this paper, we discuss a review of the present Kinect-based physiotherapy and assessment for rehabilitation patients to provide an outline of the state of art, limitation and issues of concern as well as suggestion for future work in this approach. The paper is constructed into three main parts. The introduction was discussed on physiotherapy exercises and the limitation of current Kinect-based applications. Next, we also discuss on Kinect Skeleton Joint and Kinect Depth Map features that being used widely nowadays. A concise summary with significant findings of each paper had been tabulate for each feature; Skeleton Joints and Depth Map. Afterwards, we assemble a quite number of classification method that being implemented for activity recognition in past few years.

Corresponding Author:

Nor Surayahani Suriani,
Department of Computer Engineering, Faculty of Electrical and Electronics Engineering,
Universiti Tun Hussein Onn Malaysia,
Batu Pahat 86400, Johor, Malaysia.
Email: nsuraya@uthm.edu.my

1. INTRODUCTION

Physiotherapy, a component of modern healthcare, is concerned with the development, maintenance and restoration of body movement and functionalities after illness or injury [1]. Various types of diseases and illness need to be cure by doing several exercises in order to manage pain and prevent diseases. Medical experts or therapists has been instructed to assess and cure people who have movement impairment and incompetence to perform daily tasks due to an injury or illness.

Strokes, brain injuries, motor disabilities, sport injuries, post-accident injuries and Parkinson disease are the examples of diseases that undergo physiotherapy. Stroke, also known as cerebrovascular accident or brain attack, is one of the top five leading causes of death and one of the top 10 causes for hospitalization in Malaysia [2]. According to the World Health Organization (WHO), stroke ranks as the second leading cause of death. Physiotherapy can help stroke's patients gaining their muscle control and strength back depending on the severity of the stroke. 50% of stroke survivors endure from disabilities of motor function that requires continuous rehabilitation [3].

1.1 Types of Exercises

Patients usually will be given exercises or training within their own pace and tolerance levels. The training that adequate for a patient may not be equally adequate for others. There are several types of

physical therapy exercises which depends on patients' particular condition and physical capabilities. Some of them are as:
1. Range of Motion
2. Muscles Strengthening Exercises
3. Balance Exercises
4. Flexibility Exercises
5. Post-Surgery Exercises

Table 1 summarize the types of exercises together with the examples and conditions.

Table 1. Types of Exercises Correspond to the Conditions

Physical Therapy Exercises	Explanation	Examples of Exercises	Conditions
Range of Motion	Range of Motion (ROM) exercises helps you move your joints to prevent stiffness.	Active ROM Passive ROM Active-assisted ROM	Arthritis sport injuries Post-Surgical Healing Caution
Muscles Strengthening Exercises	Increasing muscles strength to gain better balance, mobility and ability to enjoy a normal lifestyle.	Squat One-Arm Row Modified Push Up Shoulder Press Knee Extension Bridging Sit to Stand	weight control pulmonary diseases Stroke Heart diseases
Balance Exercises	Balance exercises can help people with balance or people who have muscles weakness preventing them from sudden falls	standing one foot walking in a straight line Yoga, Tai Chi	Stroke Cardiac event Elderly Low Blood Pressure
Endurance Exercises	Increase breathing and heart rate improving the health of lungs and heart also improve person's overall fitness.	Waking Stairs Climbing	Parkinson's Cardiovascular diseases
Flexibility Exercises	Stretching can help improve human body to become more flexible and limber.	Hamstring Stretch Chest Stretch Calf Stretch Back Stretch Shoulder Stretch	Back pain Disc Diseases Parkinson's
Post-Surgery Exercises	Surgery patients experienced pain, muscle contractions and stiffness. Physical therapy can relieve these issues by gradually adjusting the physical conditioning	Head Lift Buttock Lift Walking	Depends on body parts surgeries

Currently, the patients need to attend therapists at the clinic and it's such a burden for the caregiver and the patients itself. It's so inconvenient for the patients especially for elderly and bed-ridden patients, to go back and forth once in a week for the physiotherapy. Aside from that, patients may need to wear assistive devices such as sensors throughout the session. This leads to unpleasant training for the users. In the other hand, there are limited physical therapy equipment and therapists allocated in clinic or hospital. Therefore, patients may have to wait until their turn to perform the exercises as the therapists unable to assess them or the equipment is being used by other patients.

1.2 Limitations of Exergames and Serious Games

Recent studies have demonstrated that Kinect Sensor can be utilized to evaluate clinically-relevant parameters of gait [4], [5] and posture [6]. Kinect-based virtual stepping treatment has been appeared to be powerful for post-stroke rehabilitation of gait [7].Boundless of applications and research on Kinect-based were conducted even now. The most leading applications on Kinect-based for rehabilitations exercises are Exercise Games and Serious Games.

Exercise games, known as exergames intent to integrate natural human motion and the entertainment to promote elderly exercise while serious games aim to concurrently rehabilitate motor-impaired users and monitored patients' progress. However, exercise games and serious games must come to the limits as the requirements for clinical data capture for specific limb movements cannot be achieved. Below are the summarization of the limitations for exergames and serious games [8]:
1. Designated games particularly for symptomatic use are restricted to non-occluding movements. This implies that standard stroke impairment level tests requiring pervasive occluding movement sets may be illogical for a Kinect-based system to capture.

2. Diagnostic potential for extremities is limited to gross movements, as fine movements of the hand and foot are currently outside the Kinect's capture sensitivity.
3. Games targeted at rehabilitation may be prone to "cheating" which means unnatural.
4. Appropriate response to failure and poor performance, if not accounted for during game design, can inherently limit positive outcomes due to demotivation
5. The advantages of the games mainly considered for here and now with small sized studies.

In addition, the games itself having cons as they are not suitable for all ages. For an example, exergames and serious games are built only for young and middle ages people. It is inadequate for elderly as elderly might have secondary disabilities such eye-sight, hearing, speech problems thus they can't focus on the screen as well as the games. The games also irrelevant for bed-ridden patients.

2. HUMAN ACTION RECOGNITION

Vision based human action recognition is an orderly way to recognize and perceive the movement of people in camera captured content. It composes of fields such as Biomechanics, Machine Vision, Image Processing, Artificial Intelligence, and Pattern Recognition [9]. Human activities can be classified into four categories which is actions, gestures, interactions and group activities [10]. Motion recognition composes of many actions such as walking, sitting, standing, running, waving, etc. Figure 1 defines the steps in human motion recognition system.
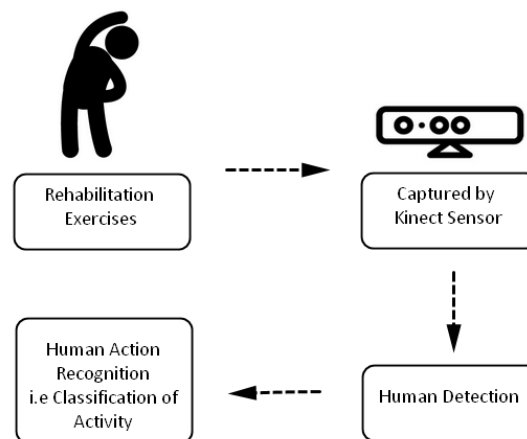

Figure 1 General framework of human action recognition

2.1 Detection

Human detection is an early stage for a human action recognition. Detecting human can be divided by many categories depends on human body parts such as upper limb, lower limb, hands, arms, head, legs etc. Each of body parts need different detection methods as to match the accuracy of the evaluation.

Computer vision and machine learning algorithm has been adopted to confront the problem of human detection in videos. There are many solutions and techniques introduced based on various scenarios including variations in illumination and poses, as well as background clutter. Dalal & Triggs report impressive results on human detection [11] by implementing Histogram of Gradient (HoG) as low-level features and outperformed other features such as wavelets[12], PCA-SHIFT [13] and shape contexts[14]. Zhu et al. proposing a rejection cascade using HoG features to improve the detection speed [15] whilst Zhang et al. come up with a multi-resolution framework in order to cut down computational cost[16].

In contrast, Lowe proposed the Scale-Invariant Feature Transform (SIFT) which has high accuracy and low computation time [17] which also being employed by Khaledian et al. for hand gesture recognition [18]. Ke and Sukthankar attempts to further improve the method by introducing PCA-SIFT [13]. Next, Speeded Up Robust Features by Bay et al. being introduced as it shown to yield comparable or better results to SIFT[19].

Perhaps, the most recent, promising approach is detection by RGB-D camera such as Kinect sensor. This detection can be split into two common methods which is Skeletal Joints and Depth Mapping.

I.   Skeleton Joints

Skeleton joints involve in combination number of joints which define body parts such as head, shoulders, neck and arms. This process describes by a very huge number of dimensions and its describes unique individuals such as their shapes, sizes, postures, motions, etc. Each version of Kinect has different number of joint types that made up a skeleton. For version 1 (Figure 2a) is made up 20-joint types, while version 2 (Figure 2b) made up of 25-joint types with additional 5 joints from Kinect v1. The additional joints in Kinect v2 are Spineshoulder, HandTipLeft, ThumbLeft, HandTipRight, and ThumbRight. The details of each joint number have been explained in Table 2.
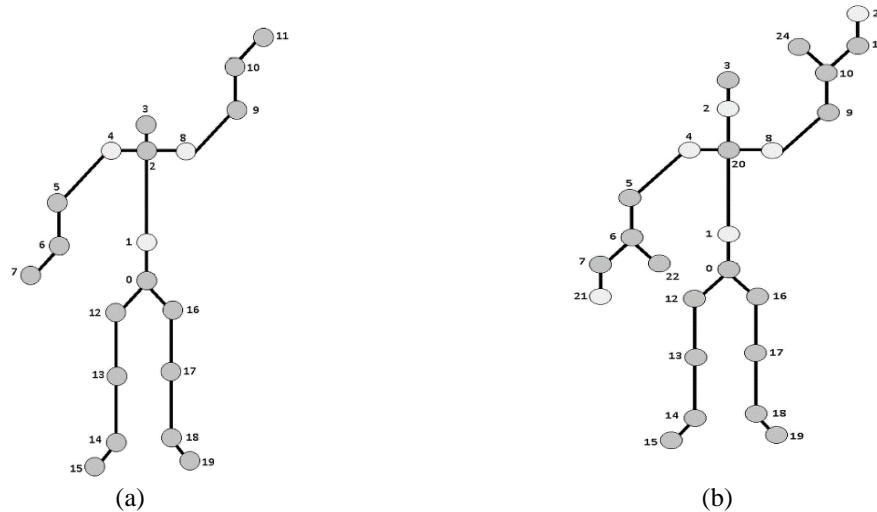

Figure 2. (a) 20-Joints of human body in action recognition through Kinect Version 1;
(b) 25-Joints of human body in action recognition through Kinect Version 2

Table 2. Kinect Skeleton Joints type for V1 and V2

20 -Joints Kinect V1				25-joints Kinect V2			
0	HipCenter	10	WristRight	0	SpineBase	13	KneeLeft
1	Spine	11	HandRight	1	SpineMid	14	AnkleLeft
2	Shoulder Center	12	HipLeft	2	Neck	15	FootLeft
3	Head	13	KneeLeft	3	Head	16	HipRight
4	ShoulderLeft	14	AnkleLeft	4	ShoulderLeft	17	KneeRight
5	ElbowLeft	15	FootLeft	5	ElbowLeft	18	AnkleRight
6	WristLeft	16	HipRight	6	WristLeft	19	FootRight
7	HandLeft	17	KneeRight	7	HandLeft	20	SpineShoulder
8	ShoulderRight	18	AnkleRight	8	ShoulderRoght	21	HandTipLeft
9	ElbowRight	19	FootRight	9	ELbowRight	22	ThumbLeft
				10	WristRight	23	HandTipRight
				11	HandRight	24	ThumbRight
				12	HipLeft		

Skeleton Joints features are incompatible to work alone as it inadequate to identify various human actions. Hence, there are many developed novels of visual representations and machine learning methods in order to fully achieved skeleton features in human action recognition.

Raptis et al. successfully used skeleton positions in a real time dance classification by employing Principle Component Analysis (PCA) on torso joints positions. This is done to determine a human torso surface as well as defining a human pose with the spherical angles within the limb joints postions and torso surface [20]. Fourier transform also being utilized over time to describe the temporal structure of actions.

Yang et al. proposed a new type of features based on position differences between joints, Eigenjoint, to represent actions which combine action information [21]. The positions difference is extracted from all the pairs in one frame, the joint of continuing frames and the joints of the initial frame with another frame to grab the structural of human postures. They also applied PCA to the features to extract the crucial data for action recognition, "eigenjoints", then undergo action classification by applying nearest-neighbor classifier.

Xia et al. introduced a novel approach for human action recognition with Histogram of 3D joint locations (HOJ3D). It extracts the histogram of spherical coordinates of the joint positions in a coordinates

system that uses the hip joint as origin [22]. They also employed Shotton et al.'s method to extract 3D joint location from a depth image by employing a local mode-finding approach based on mean shift with a weighted Gaussian Kernel to compute the confidence-scored 3D positon estimation of body joints [23].

Chaundry et al. illustrated bio-inspired dynamic 3D discriminative skeleton feature by using linear dynamic systems to model the dynamic medial axis structures of human parts. The paper considered a discriminative metric to compare sets of linear dynamics system for action recognition. Table 3 summarizes relevant papers on Kinect based for physiotherapy and assessment using skeleton joints. Each paper proposed different methods for different rehabilitation exercises.

Table 3. Kinect Skeleton Joint features for Physiotherapy and Assessment

First author / Year	Diseases & Exercises / motion	Dataset	Method used
J. Venugopalan, 2013. [24]	Traumatic Brain Injury -hand waving sideways -horizontal stretch -Vertical stretch -left and right hand extended -front hand stretch	-20 joints (x, y, z) -Quarternion (w, x, y, z) -30 frames -4 subjects (22-30 yo)	-Skeleton Normalization -Direct comparison -Cross Correlation -Dynamic Time Warping
H. Jiang, 2013. [25]	Stroke with Hemiplegia -Legs and Arms motion	-20 joints (x, y, z)	-Skeleton Normalization -Dynamic Time Warping
T.Y. Lin, 2013. [26]	Parkinson's Diseases -Seated Tai Chi Exercises with 18 forms.	-10 joints (x, y, z) -2 subjects (60 & 91 yo)	-Skeleton normalization Kolomogorov-Smirnov test
R. Staab, 2014. [33]	-jumping jacks -arm circle -arm curls	-20 joints (x, y, z)	-SVM -SVM with sigmoid Kernel -K-NN
H. D. Rosario, 2014. [45]	-motion of limbs -trunk, shoulder and hips.	-12 joints (x, y, z)	-NITE algorithm -Particle Filter
S. Li, 2014. [46]	Rehabilitation Exercises	-20 joints (x, y, z)	-Proposed algorithm -Kalman Filter
J.D.Lee, 2014. [27]	Movement Disorder -Tai-Chi exercises	-10 joints -1 subjects (60 yo)	-Skeleton Normalization -Fuzzy Logic
D Anton, 2015. [30]	Shoulder disorders -hands to mouth -shoulder extension -shoulder flexion -hands to head	-20 joints (x, y, z) -15 subjects (44- 83 yo)	-trajectory recognition -Dynamic Time Warping
Q. Wang, 2015. [47]	12 different exercises include 6 sitting poses	20 joints and 14 joints	-Unscented Kalman Filter -Kinematic Filtering
Cappecci. M, 2016. [28]	Physical impairment and disabilities -Arm lifting -squatting -pelvis rotation -trunk rotation -trunk tilting	-25 joints (x, y, z) -33 subjects (22 – 72 yo)	-Zero Velocity Crossing -Hidden Semi-Markov Model -Dynamic Time Warping
S. Sinha, 2017. [48]	Stroke neurological disorders -Active ROM exercises	20 joints (x, y, z)	-Point Cloud Segmentation -Cylinder Model Fitting -Kalman Filter
J. Richter, 2017. [31]	-Hip abduction exercises	20 joints (x, y, z)	-Local and Normalised Hierarchal Coordinates -Increment DTW -SVM
S. H. Han, 2017. [29]	Recovery states -postural correction	-20 joints (x, y, z)	-Z-Score Normalization -Deep neural network -Deep learning algorithm

Janani et al applies skeleton normalization for pre-processing data to overcome discrepancies when real time quantitative assessment of exercises performed by TBI patients at home matched with the template exercises performed in the clinic [24] as well as in Jiang et al [25], Lin et al [26], Lee et al [27], they employ normalization to compare and evaluate different skeleton models. Whereas, Cappecci e al utilizes Zero-Velocity Crossing (ZVC) to locate starting and ending points for human motion segmentation in order

to identify moving points [28]. Han et al scaling the values of joint points by implementing Z-Score Normalization as to improve the performance of deep learning algorithm [29].

Further on exercises assessment, Cappecci et al [28] compared a Histogram Semi-Markov Model (HSMM) based algorithm to monitor and evaluate rehabilitation exercises with Dynamic Time Warping algorithm resulting HSMM outperformed DTW as HSMM demonstrated scores correlated better with the clinical scores. However, Janani et al [24] stated that DTW surpass Direct Comparison and Cross-Correlation method as DTW able to give a large score and achieve higher separation between most of the similar and dissimilar videos. DTW also being used by Jiang et al [25] and Anton et al [30].

In the other hand, Ritcher et al intend to give continuous evaluation to the patients, hence, they employed the extension of DTW which is Incremental DTW [31] introduced by Khan et al. [32] that delegates the comparison between references exercises and exercises that patient currently performs. Then, they classify the motion using hierarchical SVM. Staab also implements SVM and also SVM with sigmoid Kernel to train some motion exercises since each exercise has a unique distribution in feature space [33]. Thus, concludes that a model that performs well for one exercise might not be suitable for another.

In brief, skeleton joints can provide reliable joint coordinates to the users with its real-time skeleton estimation algorithm. Its also has drawn a great attention [33-43], as it brings a great robustness to illumination, clustered background, and camera motion.

II.   Depth Maps

Depth imaging technology has advanced adequately over last few years, finally reaching a consumer price point with the launch of Kinect. Depth images provide depth information of an object or also known as z-information of an object in a real world. Depth maps can be collected through Stereo Camera, Laser Triangulation etc. It is also widely used in many 3D vision algorithms recently. The intensity values in an image represent the distance of the object from a viewpoint. As illustrated in Figure 3a) there is a bottle and an umbrella in an area. The depth image in Figure 3b) shows luminance in proportion to the distance of an object from the camera. The nearer object to the camera which is the bottle is darker while the further object, the umbrella, is lighter.

(a)   (b)

Figure 3. Example of comparison between normal and Depth image using Kinect v2.
a) Normal Image; b) Depth Image

In the last few years, solutions for activity recognition have been presented, they intended to extract features from depth data such as [49] where they presenting the adaptive spatial-temporal pyramid to improving in retaining the spatial and temporal orders. Truong et al. developed a simple novel method from hand gesture recognition that achieve accurately in real time using depth information from Kinect Sensor [50]. Author applying thresholds to the hand point that tracked by Bayesian Object Localization method in the depth image to determine the hand region.Next, Samad et al applied background segmentation with improved adaptive Gaussian mixture algorithm to the depth map to detect moving obejcts [51].

Whereas, raw depth map has been smoothen by applying two filtering methods which are pixel filtering and context filtering [52]. Then, the depth map encoded through proposed Local Ternary Direction Pattern (LTDP) feature descriptor and utilized by SVM classifier. The result turns out that LTDP outperformed others five existing descriptors (LBP, LTP, HOG, PHOG, CENTRIST) and the nonlinear effect of SVM classification task were reduced by using LTDP on depth map.

Yang et al. proposed HOG in Depth Motion Maps (DMM-HOG) which applies the HOG descriptor on depth motion maps. It is computed by taking the difference of the depth maps in two consecutive frames, thresholding the difference, and aggregating the difference over time. Then, they extract DMMs from the

front, top and side views [53]. Next, Xia et al. present a novel human detection method by depth information by Kinect and the results can adequately detect the persons in all poses and appearances also provide exactde estimation of the whole body contour of a person [54].

On the other hand, Ni et al. propose a method for action recognition by combining depth maps with RGB videos. The method is done by identifying the interest points in RGB videos, extracting HOG/HOF features and LDP features from RGB videos and depth sequences, respectively and concatenating the RGB and depth map features [55]. It shows that the information in RGB videos and depth map sequences are complementary to each other. There are massive studies on depth images approaching rehabilitation and physiotherapy assessment has been illustrated in Table 4.

Bakar et al [56] and Sosa et al [57] employs Region of Interest (ROI) in the segmentation phase, to minimize the area and remove unwanted objects that appears around while Sinha et al [58] proposed an algorithm in Depth-based segmentation and PCA to improve accuracy of Kinect for upper body rehabilitation applications.

Next, Yao et al [59] proposed a Kinect-based rehabilitation system for both therapists and patients. They evaluate their time sequence-based data by implementing Cross Correlation as the method is well known in detecting common periodicities. They also employed DTW to compare whether the patients done the exercises as the same rate as the skeleton frame sequences. It is to compare and find optimal alignment between two given time-sequences. Furthermore, Ye et al [60] utilizes DTW to compute a distance matrix for gait pattern extraction while Su et al [61] applying DTW measure the similarity of joint data between "at home exercises "and" in hospital exercises".

Su et al then evaluate the performance by using Adaptive Neuro-Fuzzy Inference System (ANFIS) which integrates a neural network and a fuzzy logic [61]. Nomm et al practiced Neural Network based model, NN-based ANARX (Additive Nonlinear Auto Regressive exogenous), in their monitoring system as it can adjust the system according to the specific needs of each patient [62] whereas Ye et al used NN-based on nonlinear autoregressive with exogenous (NARX) for gait phase classification and Enhanced Random Decision Forest (ERF) for missing features cases [60]. Next, Nahavandi et al trained a Random Decision Forest (RDF) for generalising a learning model in order to discriminate between seven RULA-scored sets of postures [63].

In contrast, Collins et al achieved to recognize several human actions done by stroke patients with high certainty by employing HON4D as a global descriptor [64]. Nghia et al proposed an algorithm to compute discriminative features, depth of wrist, by building a mapping table between the differences of bone joint depth and head depth as Kinect provided [65]. Consequently, depth maps is a good approach for detecting human action as it insensitive to changes in lighting conditions, hence it even works in low ambient light condition [66]. Depth maps properties also convenience to works with specific feature descriptors which spikes their evaluation performance [48], [65].

2.2 Recognition and Classification

Recognizing human activities in an video sequences or images can be quite challenging due to complication of an area [68] such as background clutter, partial occlusion, change in scale, lighting etc. Therefore, classification is needed to classify and recognize the action of a human action to solve the recognition and localization problem.

Hidden Markov Models (HMMs) are known to have high classification rates and are favoured for classifying dynamic gestures. [67]- [71] all use HMMs straightforward manner for gesture classification. Whilst, [74] employed Riemannian shape space to represent distance between curves and using k-Nearest Neighbour (k-NN) as the classifier. k-NN classifiers prominent for static pose also because of their high classification rates other than simple to be implemented.

In [23], Randomized Decision Forests which is a forest is an ensemble of trees, is be used as the classifier and have been proven faster and effective classifiers for many tasks [73-75] and can be implemented efficiently on the GPU [78]. Random forests also being implemented in [77], [78] for action recognition.

On the contrary, Neural Networks (NN) and Support Vector Machine (SVM) are also commonly used for poses and motion recognition. Suriani [81] employs SVM to identify a person state whether they in a normal or anomaly movement for fall detection while doing home-based rehabilitation exercises. Ciresan et al. used Convolutional Neural Networks (CNN) [82], while Toshev et al. implement Deep Neural Networks (DNN) to recognize human poses and activities [83]. Du et al. divided human skeleton into five segments and used each of the parts to train a hierarchical recurrent neural networks [36].

Table 4. Kinect Depth Map Features for Physiotherapy and Assessment

First author / Year	Diseases & Exercises / motion	Dataset	Method used
Y. J. Chang, 2013. [84]	Cerebral Palsy Upper limb rehabilitation	-Depth Information	-Kolmogorov-Smirnov Test
B.Penelle, 2013. [85]	Lower limb Whole Body	-Depth Images	- GPU based -Particle Filter
S. Nomm, 2013. [62]	Motor Functions Therapeutic exercises for human limbs -Dynamic ROM	-Depth information	-Neural Networks -NN-based ANARX
T. Watanabe, 2014. [86]	Elderly Lower Limb Chair Exercise -toe lift -heel lift -Knee extensions -thigh Lifts -open-leg exercises	-Depth Images -7 subjects (75 – 85 yo)	-Average Recognition Rate
L. Yao, 2014. [59]	-Rehabilitation Exercises	-Depth Images	-DTW
C. J. Su, 2014. [61]	Post-injuries -shoulder rehab exercises	-depth information	-DTW -Neural Network -Fuzzy Logic
M.Z. A. Bakar, 2015. [56]	Wrist Hand Injury Hand Deviation	-3D data -Depth Data -RGB Data	-Region of Interest (ROI) -Hand Contour - K-Curvature
L. Omelina, 2016 [87]	Face recognition	-Depth Images	-Local Binary Patterns with Chi Square -K-Mean
G. D. Sosa, 2015. [57]	Multiple Sclerosis -shoulder elevation -shoulder abduction -hip abduction	-RGB-D 2D with 30fps -4 subjects (24 – 39 yo)	-Region of Interest (ROI) -background subtraction -human silhouette -skeletonization
S. Sinha, 2016. [58]	Upper body -ROM exercise	-RGB Depth -10 subjects (21- 55yo)	-Principal Component Analysis (PCA) -Proposed algorithm
M. Ye, 2017. [60]	-Stroke -walking exercises	-depth map	-DTW -NN based algorithm (NARX) -Kernel Filter -Enhanced Random Decision Forest
D. Nahavandi, 2017. [63]	Musculoskeletal Disorder -Shoulder -Elbow -Trunk -Neck	-depth images -Rapid Upper Limb Assessment Scoring (RULA)	-DCF Feature Extraction -Random Decision Forest
J. Collins, 2017. [64]	Stroke - arm extensions -chest sway -waking	-depth data	-Histogram of Oriented 4D Normal (HON4D)
V. T. T. T. Nghia, 2017. [65]	-Flexion -Elevation -Abduction	-skeleton joint -depth images	-proposed algorithm
R. Samad, 2018.			

3. LIMITATIONS AND FUTURE WORKS

This paper has studied several methods for Kinect-based physiotherapy and assessment that can be used to develop technology and innovative services in assisted living applications. However, there are limitations and consequences within the methods discussed.

In the context of skeleton joints, detection may become less accurate and robust when joints not labelled properly as it leads to confusion of points especially between the ankle and knee points. Thus, it will be resulting to false detection.Furthermore, determining inner joints such as elbow and knee joints that are hidden and being kept out also caused unsastified results [88].

Next, we disclosed another constraint which is high interclass similarity where similar joints are being detected for different activities. For an example, in CAD-60 dataset, both drinking water and brushing teeth activities were including the similar joints and appeared to be errornous on classifying each of the activity. In addition, we also found another challenging reason, high intraclass variability limitations. There are several same actions that being performed different ways by the same subject for example, the subject performed the same action by left hand or right hand, or both hand indifferently.

However, for depth map, occlusion is the most problem encountered by researchers. Human image sometimes being occluded by another objects or persons hence degrading the detection performance [89]. Also, using only one depth camera can increases the persistent of background or foreground occlusions. Another drawback that being focused on depth map is cluttered background which eventually causing poor detection as the subject is not clear and difficult to detect. Other than that, as the distances to the camera increases, the accuracy of depth maps degrades and noisy.

Hence, in the future, we intend to develop an algorithm or adapting some existing algorithm in order to minimize the limitations. Next, we will be considering more adequate models and combine them to further enhance the performance of detection under our proposed framework.

4.   CONCLUSION

Human activity understanding has attracted widespread interest for many researchers. This field becomes one of the most active research topics in computer vision. We have provided a comprehensive analysis of the existing, publicly available of Kinect-based Physiotherapy and Assessment. There are various applications of Kinect in the field of Physiotherapy and Assessment recently. We have explained types of exercises, limitations of exergames and serious games for rehabilitation. These direction in future works are vast to enhance methods for Kinect Assessment system to accurately for a patient completing rehabilitation exercises, also the data can be clinically used.

REFERENCES

[1]   Lin J. F. S. Automated Rehabilitation Exercise Motion Tracking. University of Waterloo. 2012.
[2]   Loo K. W, Gan S. H. Burden of Stroke in Malaysia. *Int. J. Stroke.* 2012; 7(2):165–167.
[3]   Kelly-Hayes M, Beiser A, Kase C. S, Scaramucci A, D'Agostino R. B, Wolf P. A. The influence of gender and age on disability following ischemic stroke: the Framingham study. *J. Stroke Cerebrovasc. Dis.* 2003; 12(3):119–126.
[4]   Stone E. E, Skubic M. *Passive in-home measurement of stride-to-stride gait variability comparing vision and Kinect sensing.* 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011; 6491–6494.
[5]   Clark R. A, Bower K. J, Mentiplay B. F, Paterson K, Pua Y.-H. Concurrent validity of the Microsoft Kinect for assessment of spatiotemporal gait variables. *J. Biomech.* 2013; 46(15):2722–2725.
[6]   Clark R. A. Validity of the Microsoft Kinect for assessment of postural control. *Gait Posture.* 2012; 36(3):372–377.
[7]   Lloréns M. D. N. R, Alcañiz M, Font C. C. Balance Recovery Through Virtual Stepping Exercises Using Kinect Skeleton Tracking: A Follow-Up Study With Chronic Stroke Patients. *Stud. Health Technol. Inform.* 2012; 181:108–112.
[8]   Webster D, Celik O. Systematic review of Kinect applications in elderly care and stroke rehabilitation. *J. Neuroeng. Rehabil.* 2014; 11(1):108.
[9]   And G. V. K, Patil V. H. A Study of Vision based Human Motion Recognition and Analysis. *CoRR.* 2016; abs/1608.0.
[10]   Aggarwal J. K, Michael S. R. Human activity analysis: A review. *ACM Comput. Surv.* 2011; 43(16):1-43
[11]   Dalal N, Triggs B. *Histograms of oriented gradients for human detection.* 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005; 1:886–893.
[12]   Mohan A, Papageorgiou C, Poggio T. Example-based object detection in images by components. *IEEE Trans. Pattern Anal. Mach. Intell.* 2001; 23(4):349–361.
[13]   Ke Y, Sukthankar R. *PCA-SIFT: a more distinctive representation for local image descriptors.* Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. 2004; 2:II-506-II-513.
[14]   Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* 2002; 24(4):509–522.
[15]   Zhu Q, Yeh M.-C, Cheng K.-T, Avidan S. *Fast Human Detection Using a Cascade of Histograms of Oriented Gradients.* 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). 2006; 2:1491–1498.
[16]   Zhang W, Zelinsky G, Samaras D. *Real-time Accurate Object Detection using Multiple Resolutions.* 2007 IEEE 11th International Conference on Computer Vision. 2007; 1–8.
[17]   Lowe D. G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* 2014; 60(2):91–110.
[18]   Khaledian M, Menhaj M. B. Real-time vision-based hand gesture recognition using SIFT features. *Indones. J.*

Electr. Eng. Comput. Sci. 2015; 15(1):162–170.

[19] Bay H, Tuytelaars T, Van Gool L. SURF: Speeded Up Robust Features BT - Computer Vision – ECCV 2006. 2006; 404–417.

[20] Raptis M, Kirovski D, Hoppe H. *Real-time Classification of Dance Gestures from Skeleton Animation*. Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. 2011; 147–156.

[21] Yang X, Tian Y. L. *EigenJoints-based action recognition using Naive-Bayes-Nearest-Neighbor*. 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2012; 14–19.

[22] Xia L, Chen C. C, Aggarwal J. K. V*iew invariant human action recognition using histograms of 3D joints*. 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2012; 20–27.

[23] Shotton J. *Real-time human pose recognition in parts from single depth images*. CVPR 2011. 2011; 1297–1304.

[24] Venugopalan J, Cheng C, Stokes T. H, Wang M. D. *Kinect-based rehabilitation system for patients with traumatic brain injury*. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2013; 4625–4628.

[25] Jiang X, Huaqiang, Jie. *Kinect-Based Rehabilitation Training Assistant System Research And Implementation.* 2013 International Conference on Software Engineering and Computer Science. Atlantis Press. 2013.

[26] Lin T. Y, Hsieh C. H, Lee J. D. A *Kinect-Based System for Physical Rehabilitation: Utilizing Tai Chi Exercises to Improve Movement Disorders in Patients with Balance Ability*. 2013 7th Asia Modelling Symposium. 2013; 149–153.

[27] Lee J. D, Hsieh C. H, Lin T. Y. *A Kinect-based Tai Chi exercises evaluation system for physical rehabilitation.* 2014 IEEE International Conference on Consumer Electronics (ICCE). 2014; 177–178.

[28] Capecci M. *Physical rehabilitation exercises assessment based on Hidden Semi-Markov Model by Kinect v2*. 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). 2016; 256–259.

[29] Han S.-H, Kim H.-G, Choi H.-J. *Rehabilitation posture correction using deep neural network.* 2017 IEEE International Conference on Big Data and Smart Computing (BigComp). 2017; 400–402.

[30] Antón A. I. D, Goñi A. Exercise Recognition for Kinect-based Telerehabilitation. *Methods Inf Med*. 2015; 54(2).

[31] Richter J, Wiede C, Shinde B, Hirtz G. *Motion Error Classification for Assisted Physical Therapy - A Novel Approach using Incremental Dynamic Time Warping and Normalised Hierarchical Skeleton Joint Data*. Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM. 2017; 281–288.

[32] Khan N. M, Lin S, Guan L, Guo B. *A Visual Evaluation Framework for In-Home Physical Rehabilitation*. 2014 IEEE International Symposium on Multimedia. 2014; 237–240.

[33] Staab R. Recognizing specific errors in human physical exercise performance with Microsoft Kinect. Master's Theses Proj. Reports. 2014.

[34] Vemulapalli R, Arrate F, Chellappa R. *Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group.* 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014; 588–595.

[35] Vemulapalli R, Chellappa R. *Rolling Rotations for Recognizing Human Actions from 3D Skeletal Data*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016; 4471–4479.

[36] Du Y, Wang W, Wang L. *Hierarchical recurrent neural network for skeleton based action recognition.* 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015; 1110–1118.

[37] Veeriah V, Zhuang N, Qi G. J. *Differential Recurrent Neural Networks for Action Recognition*. 2015 IEEE International Conference on Computer Vision (ICCV). 2015; 4041–4049.

[38] Zhu W. Co-occurrence Feature Learning for Skeleton based Action Recognition using Regularized Deep {LSTM} Networks. *CoRR*. 2016; abs/1603.0.

[39] Shahroudy A, Liu J, Ng T.-T, Wang G. {NTU} {RGB+D:} {A} Large Scale Dataset for 3D Human Activity Analysis. *CoRR*. 2016; abs/1604.0.

[40] Liu J, Shahroudy A, Xu D, Wang G. Spatio-Temporal {LSTM} with Trust Gates for 3D Human Action Recognition. *CoRR*. 2016; abs/1607.0.

[41] Zanfir M, Leordeanu M, Sminchisescu C. *The Moving Pose: An Efficient 3D Kinematics Descriptor for Low-Latency Action Recognition and Detection.* 2013 IEEE International Conference on Computer Vision. 2013; 2752–2759.

[42] Gowayyed M. A. Histogram of Oriented Displacements (HOD): Describing Trajectories of Human Joints for Action Recognition. *CoRR*, 2013. 2013; abs/1611.0(7):2752–2759.

[43] Song S, Lan C, Xing J, Zeng W, Liu J. An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data. *CoRR*. 2016; abs/1611.0.

[44] Du Y, Fu Y, Wang L. Representation Learning of Temporal Dynamics for Skeleton-Based Action Recognition. *IEEE Trans. Image Process.* 2016; 25(7):3010–3022.

[45] De Rosario H, Belda-Lois J. M, Fos F, Medina E, Poveda-Puente R, Kroll M. Correction of joint angles from kinect for balance exercising and assessment. *J. Appl. Biomech.* 2014; 30(2):294–299.

[46] Li S, Pathirana P. N, Caelli T. *Multi-kinect skeleton fusion for physical rehabilitation monitoring*. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014. 5060–5063.

[47] Wang Q, Kurillo G, Ofli F, Bajcsy R. Remote Health Coaching System and Human Motion Data Analysis for Physical Therapy with Microsoft Kinect. *CoRR*, abs/1512.0(111965).

[48] Sinha S, Bhowmick B, Sinha A, Das A. Accurate estimation of joint motion trajectories for rehabilitation using Kinect. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2017; 3864–3867.

[49]   Yang X, Tian Y. *Super Normal Vector for Activity Recognition Using Depth Sequences*. 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014; 804–811.

[50]   T. Q. Vinh and N. T. Tri, "Hand gesture recognition based on depth image using kinect sensor," in *2015 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS)*, 2015, pp. 34–39.

[51]   Samad R, Yan L. W, Mustafa M, Abdullah N. R. H, Pebrianti D, Multiple Human Body Postures Detection using Kinect. *Indones. J. Electr. Eng. Comput. Sci.* 2018; 10(2).

[52]   Shen Y, Hao Z, Wang P, Ma S, Liu W. *A Novel Human Detection Approach Based on Depth Map via Kinect.* 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2013; 535–541.

[53]   Yang X, Zhang C, Tian Y. *Recognizing Actions Using Depth Motion Maps-based Histograms of Oriented Gradients*. Proceedings of the 20th ACM International Conference on Multimedia. 2012; 1057–1060.

[54]   Xia L, Chen C. C, Aggarwal J. K. *Human detection using depth information by Kinect.* CVPR 2011 WORKSHOPS. 2011; 15–22.

[55]   Ni B, Wang G, Moulin P. *RGBD-HuDaAct: A color-depth video database for human daily activity recognition.* 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). 2011; 1147–1153.

[56]   Bakar M. Z. A, Samad R, Pebrianti D, Mustafa M, Abdullah N. R. H. C*omputer vision-based hand deviation exercise for rehabilitation.* 2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE). 2015; 389–394.

[57]   Sosa G. D, Sánchez J, Franco H. *Improved front-view tracking of human skeleton from Kinect data for rehabilitation support in Multiple Sclerosis.* 2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA). 2015; 1–7.

[58]   Sinha S, Bhowmick B, Chakravarty K, Sinha A, Das A. *Accurate upper body rehabilitation system using kinect.* 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016; 4605–4609.

[59]   Yao A, Li, Xu, Hui, Li. Kinect-based rehabilitation exercises system: therapist involved approach. *Biomed. Mater. Eng.* 2014; 6(24):2611–2618.

[60]   Ye M, Yang C, Stankovic V, Stankovic L, Cheng S. *Gait phase classification for in-home gait assessment.* 2017 IEEE International Conference on Multimedia and Expo (ICME). 2017; 1524–1529.

[61]   Su C.-J, Chiang C.-Y, Huang J.-Y. Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic. *Appl. Soft Comput.* 2014; 22:652–666.

[62]   Nomm S, Buhhalko K. *Monitoring of the Human Motor Functions Rehabilitation by Neural Networks Based System with Kinect Sensor.* IFAC HMS. 2013.

[63]   Nahavandi D, Hossny M. *Skeleton-free task-specific rapid upper limb ergonomie assessment using depth imaging sensors*. 2016 IEEE SENSORS. 2016; 1–3.

[64]   Collins J, Warren J, Ma M, Proffitt R, Skubic M. *Stroke patient daily activity observation system.* 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2017; 844–848.

[65]   Vo T. T. N, Tran A. T, Duong V. H, Wang J. C, Bao P. T. *Real time validating the accuracy of physiotherapy exercises.* 2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW). 2017; 329–330.

[66]   Rimkus K, Bukis A, Lipnickas A, Sinkevičius S. *3D human hand motion recognition system.* 2013 6th International Conference on Human System Interactions (HSI). 2013; 180–183.

[67]   Oreifej O, Liu Z. *HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences.* 2013 IEEE Conference on Computer Vision and Pattern Recognition. 2013; 716–723.

[68]   Vrigkas M, Nikou C, Kakadiaris I. A. A Review of Human Activity Recognition Methods. *Front. Robot. AI.* 2015; 2:28.

[69]   Veenendaal A, Daly E, Jones E, Gang Z, Vartak S, Patwardhan R. S, Sensor Tracked Points and HMM Based Classifier for Human Action Recognition. *Comput. Sci. Emerg. Res. J.* 2016; 5.

[70]   Afsar P, Cortez P, Santos H. *Automatic Human Action Recognition from Video Using Hidden Markov Model* 2015 IEEE 18th International Conference on Computational Science and Engineering. 2015; 105–109.

[71]   Uddin M. Z, Hassan M. M. A depth video-based facial expression recognition system using radon transform, generalized discriminant analysis, and hidden Markov model. *Multimed. Tools Appl.* 2015; 74(11):3675–3690.

[72]   Mahapatra A, Mishra T. K, Sa P. K, Majhi B. Human recognition system for outdoor videos using Hidden Markov model. *AEU - Int. J. Electron. Commun.* 2014; 68(3):227–236.

[73]   Jalal A, Kamal S, Kim D. Human Depth Sensors-Based Activity Recognition Using Spatiotemporal Features and Hidden Markov Model for Smart Environments. *J. Comput. Netw. Commun.* 2016:5-.

[74]   Devanne M, Wannous H, Berretti S, Pala P, Daoudi M, Bimbo A. D. *Space-Time Pose Representation for 3D Human Action Recognition BT - New Trends in Image Analysis and Processing.* ICIAP 2013: ICIAP 2013 International Workshops, Naples, Italy, September 9-13, 2013. Proceedings. 2013; 456–464.

[75]   Deng J, Xie X, Daubney B. A bag of words approach to subject specific 3D human pose interaction classification with random decision forests. *Graph. Models.* 2014; 76(3):162–171.

[76]   Handrich S, Al-Hamadi A, Lilienblum E, Liu Z. *Human bodypart classification using geodesic descriptors and random forests*. 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA). 2017; 318–321.

[77]   Rahmani H, Mahmood A, Huynh D. Q, Mian A. *Real time action recognition using histograms of depth gradients and random decision forests.* IEEE Winter Conference on Applications of Computer Vision. 2014; 626–633.

[78]   Pianu D, Nerino R, Ferraris C, Chimienti A. A novel approach to train random forests on GPU for computer vision

applications using local features. *Int. J. High Perform. Comput. Appl.* 2015; 30(3):290–304.

[79] Feng Z, Mo L, Li M. *A Random Forest-based ensemble method for activity recognition*. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2015; 5074–5077.

[80] L. Gan and F. Chen, *Human Action Recognition Using APJ3D and Random Forests*, vol. 8. 2013.

[81] Suriani N. S. Fall Detection Using Visual Cortex Bio-inspired Model for Home-Based Physiotherapy System BT - Advances in Machine Learning and Signal Processing. 2016; 47–57.

[82] Cirecsan D. C, Meier U, Masci J, Gambardella L. M, Schmidhuber J. *Flexible, High Performance Convolutional Neural Networks for Image Classification*. Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two. 2011; 1237–1242.

[83] Toshev A, Szegedy C. *DeepPose: Human Pose Estimation via Deep Neural Networks*. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014; 1653–1660.

[84] Chang Y.-J, Han W.-Y, Tsai Y.-C. A Kinect-based upper limb rehabilitation system to assist people with cerebral palsy. *Res. Dev. Disabil.* 2013; 34(11):3654–3659.

[85] B. Penelle and O. Debeir, "Human motion tracking for rehabilitation using depth images and particle filter optimization," in *2013 2nd International Conference on Advances in Biomedical Engineering*, 2013, pp. 211–214.

[86] Watanabe T, Ohtsuka N, Shibusawa S, Kamada M, Yonekura T. *Design of Lower Limb Chair Exercise Support System with Depth Sensor*. 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops. 2014; 104–111.

[87] Omelina S. L, Jansen B, Bonnechère B, Oravec M, Pavlovicova J, Van S. J. Interaction detection with depth sensing and body tracking cameras in physical rehabilitation. *Methods Inf. Med.* 2016; 55(1):70–78.

[88] Ding J, Wang Y, Yu L. *Extraction of Human Body Skeleton Based on Silhouette Images*. 2010 Second International Workshop on Education Technology and Computer Science. 2010; 1:71–74.

[89] Sung J, Ponce C, Selman B, Saxena A. *Unstructured human activity detection from RGBD images*. 2012 IEEE International Conference on Robotics and Automation. 2012; 842–849.