Vision-based Action Understanding for Assistive Healthcare: A Short Review

Md Atiqur Rahman Ahad[1, 2], Anindya Das Antar[1], and Omar Shahid[1]

[1]University of Dhaka, Bangladesh

[2]atiqahad@du.ac.bd

Abstract

The scarcity of trained therapist, economic imbalance, and an increasing amount of elderly people are the reasons for poor rehabilitation treatment and inadequate healthcare facilities in many countries. Vision-based rehabilitation treatment, monitoring daily living, and advanced healthcare can improve technology that allows people with an injury to practice intense movement training without taking help from a therapist daily. This technology has remarkable basic notable benefits as vision-based systems are non-contact, precise, immune to electromagnetic interference, nondestructive, and they can be used for long range and multiple target monitoring. The objective of this survey paper is devoted to exhibiting a summary of the challenges and difficulties in this domain along with some solutions. Besides, in order to guide the researchers in this field, we have discussed available sensing devices in the field of computer vision that can be used for taking data in hospitals and rehabilitation centers. We have also analyzed some benchmark datasets regarding gestures, medical activities, sports and exercise actions, 3D actions, and so on with relevant information. Moreover, this article also provides a comparison among existing research works on some benchmark datasets related to this field of research.

1. Introduction

Rehabilitation process combines pharmacological and mostly psychotherapeutic treatments along with exercises and therapies in order to help the patients to get back or improve skills and functioning for daily living that has been lost due to sickness, injury, or birth complications. These difficulties may be a result of stroke, hip fractures, some forms of arthritis, neurological disorders, major multiple trauma, brain injury, and so on. In most of the cases, a large number of people demanding rehabilitation treatment are elderly people and suffering from paralysis due to injuries. For example, each year more than 60,000 people survive a stroke and other accidental injuries in the USA and similar figure exists in other countries. A recent report based on visual statistics and hospital discharge data shows that the annual incidence of traumatic brain injury in the USA can be approximated to 102 per 100,000 making injury the leading cause of death among American citizens who are under 45 years of age [24]. Approximately 80% survivors of these injuries lose arms, legs along with impaired mobility, balance, and coordination. Advanced rehabilitation systems strive to help them recover as soon as possible under proper treatment. It is a matter of fact that economic pressures on healthcare providers, shortage of young people for assistive care, and lack of advancement in medical sector play a major role behind the scarcity of leading-edge rehabilitation technology and assistive care. Many patients are noncompliant with the exercise plan prescribed by the professionals. This is another main barrier to the successful home exercise program (HEP) [34]. In the case of home rehabilitation system, improvement is dependent on self-exercise with little professional or quantitative feedback. Intensive and supervised training can play an essential role to improve the movement of patients after injury. Thus, the aim of rehabilitation engineering is to make technological advancement in order to allow the patients to practice intense movement exercise without the cost of a therapist who will always present to monitor the patients.

The advancement of computer vision-based systems can be used for the automation of the rehabilitation system for helping patients to overcome challenges and get back to their daily routines. A cheaper computer vision-based system will allow the injured patients to practice exercises at home or clinic with periodical communications with a therapist [62]. There has been a very limited range of research in this domain due to the complication of creating training data, managing agreement with hospitals and rehabilitation center authority for implementing cameras, arranging patients who are interested to participate in the experiment for data collection, and so on. For example, research work in

[66] presented the University of Idaho-Physical Rehabilitation Movement Dataset (UI-PRMD) for physical rehabilitation exercises. The basic goal of this work was to mathematically model the therapy movements and establish performance measure for the evaluation of patient persistence in performing the designated rehabilitation exercises.

Currently, there are a vast number of publicly available datasets associated with common human movements [17]. Those are widely utilized for tasks like gesture recognition, action recognition, fall detection, and pose estimation. Among these datasets, CMU Multi-Modal Activity (CMU-MMAC) [18] and Berkley MHAD (Multi-Modal Human Action Dataset) employ optical motion capturing systems for recording the actions and movements. Some famous datasets have been mentioned in [10] and [9], in the field of computer vision and human action recognition for assistive healthcare along with some information about existing research works. The existing therapy movement related datasets are restricted either in the extent of the actions or in the given data format. Research work in [23] shows a dataset named HPTE (Home-based Physical Therapy Exercises) related to the therapeutic movements. One major limitation of this dataset is that it provides only the depth streams and videos from the Kinect sensor. The data set does not provide the corresponding body joint positions or angles. In the field of physical rehabilitation, one dataset namely The EmoPain data set [11] was designed in order to study pain-related emotions. This dataset contains high-resolution face videos, audio files, full body joint motions, and electromyographic (EMG) signals from back muscles. Chin et al. [25] presented another dataset, which is restricted to three exercises of lower limbs performed by nine subjects using EMG electrodes. In addition, there are also some datasets dedicated to physical activity monitoring (e.g., by wearing heart rate monitors, inertial measurement units [27]), where wearable sensors have been mostly used. There are some problems regarding wearable sensors regarding comforts, medical complexities, clinical requirements, and patient needs in rehabilitation research. Wearable sensors are generally not enough for the robust, long-term monitoring of patients under real-life conditions [48]. Vision-based systems can be useful in this regard due to some inherent notable benefits such as noncontact, nondestructive, long distance, high precision, immunity to electromagnetic interference, and large-range and multiple-target monitoring [76]. The primary goal of this paper is to present a survey on advanced assistive healthcare and rehabilitation research to automate the medical system so that the exercise pattern and activity of patients can be monitored using computer vision to bring them back into a secured and healthy lifestyle. The remainder of this paper is organized as follows.

Section 1 provides a brief introduction to the domain of vision-based rehabilitation research and important applications of this field for patients with major and minor injuries. The challenges due to which there is still very less amount of research work in this domain have been discussed in section 2 along with some solutions to those problems. Section 3 describes some useful sensing devices that can be useful for data collection in hospitals and rehabilitation centers. Besides, we have described the typical environmental setting that is needed for the collection of data in an accurate manner. In order to introduce with the existing datasets that can be used by the new researchers have been described with useful pieces of information in section 4. Section 5 shows a comparison among previous research works on some benchmark datasets along with their methodology. Finally, we have drawn the conclusion in section 6 and discussed future challenges in this domain.

2. Challenges of this Domain

The primary goal of using computer vision in advanced assistive healthcare and rehabilitation research is to track the improvement of patients by monitoring their daily actions or prescribed activities by the doctors. In most of the previous researches, wearable camera [63] or camera networks [21] had been used for the real-time continuous monitoring of physical activities. There are some factors that limit the accuracy and increase the difficulties in this field of research. In this section, we have demonstrated those challenges with examples and suggested some possible solutions.

2.1. Data Collection and Labeling Training Data

Most of the subjects in this field of research are patients with injuries or some diseases who require rehabilitation care to get back to a healthy life. It is tough to create realistic datasets due to this reason in this field of research. This is a common practice to engage students for creating a normal vision dataset with daily actions like eating, brushing teeth, performing a gesture, etc. In contrary, it will not be realistic if we engage students to perform actions and exercises like injured patients to monitor their daily improvement. This is also costly and tough to manage real patients to create the dataset. There are also limitations in terms of law and this will require an agreement with the hospital authority. Installations of cameras in some rooms of rehabilitation care with permission of the authority can solve this problem to some extent. If we can take daily data of real patients in hospitals or rehabilitation care using camera, then this can be a useful factor for creating a realistic dataset for this domain.

2.2. Dataset Diversity

This is a big challenge to create a dataset with diversity in healthcare and rehabilitation research using computer vision. A good diverse dataset should be created with more

subjects, gender variation (male and female), age variation (children, young, and old), and so on. Besides, the actions should be performed multiple times and for multiple days to get a specific amount of data. This is tough to fulfill this requirement when the subjects are patients. A collaborative project engaging patients from different hospitals can be a possible solution to this problem. We can also collect and share data from other countries by posing some rules and regulations. Osaka University inertial sensor-based gait database [50] is one of the largest databases in terms of diversity, which includes at most 744 subjects. Among them, 389 are males and rest are females with ages ranging from 2 to 78 years for gait-based personal authentication. The Osaka University Multi-View Large Population Dataset (OU-MVLP) [52] is another diverse dataset with 10,307 subjects (5,114 males and 5,193 females with various ages, ranging from 2 to 87 years).

2.3. Intra-class and Inter-class Variations

There are wide variations in terms of performance for many actions while performing exercise actions suggested by trained doctors and therapists. For example, hip flexion with a hold, knee extension, wrist bend movement, wrist side movement can vary in speed and stride length. There are also anthropometric differences between patients while performing these actions due to injury. An ideal human action recognition method for medical application should be able to speculate over changes within one class. The different action classes should be distinguished properly too. This will be more challenging for growing number of action classes as there will be more amount of overlap among classes. In some cases, a class label distribution might be a suitable solution.

2.4. Environment and Lab Settings

The environmental conditions can vary up to a large extent while performing actions based on day or night, outdoor or indoor, etc. If the dataset is created under controlled lab set up, there is a high chance of lower performance in real time application in the hospitals or inside the houses. It is hard to localize patient actions in cluttered or changing environments. Besides, we can get occluded body parts of a person in the recording. Lighting condition also plays an important role in this regard.

2.5. Number and Position of Cameras

This is a tough challenge in terms of accuracy, cost, and set up to fix the amount and position of cameras for creating the dataset. It is difficult to track the actions of patients using a single camera. Observation of actions from different viewpoints can lead to very different image observations. Occluded viewpoint problems and issues can be alleviated using multiple cameras especially by combining the observations from multiple views into a consistent representation. If the background is dynamic, it can enhance the complexity of localizing the person in the image by robustly observing the motion. In the case of a moving camera, these challenges become even harder. The camera position also needs to be fixed according to the application. The camera can be placed into the head of the patients, shirt buttons, in a table or inside the room.

2.6. Spatio-temporal Variations

Normally, it is expected that actions are segmented in time, which removes the load of the segmentation from the recognition task. But there can be a large deviation in the rate of performance of an action. While using motion features, the action recording rate has an extensive effect on the temporal extent of the action. We can call a human action recognition algorithm robust if it is invariant to various rates of execution.

2.7. Dataset Availability

Another challenge of this field lies in the fact that most of the advanced and good datasets related to medical activity data are not publicly available for research purpose. Most of the research projects in this domain are funded by companies or hospitals who do not permit the public availability of those datasets, created under their monitoring. Privacy issue can be another problem in vision-based research. Most of the datasets are not publicly available because of the privacy cencern of the subjects in RGB images and videos. Depth images and skeleton data are not strictly private but they can also violate the privacy as they contain some sort of personal information. This is a big challenge for new researchers to work in this field without finding a proper dataset due to the rules and regulations of the authority of not publishing it. It is tough for a new researcher to create a public dataset by self-funding as it requires a large amount of cost for setting up the experiment and environment for taking data.

2.8. Reconstruction of Features from Images

The sequence of frames may be distracted or in other words, there could be nominal motion in sequence due to the wind as it may cause the camera sways back. In this case, we can't avoid false detection without a robust maintenance scheme. In some cases, new background object may be inserted and it might be detected as foreground. It happens due to the unavailability of a robust maintenance scheme. Sometimes, shadows of moving object may appear as a foreground object. These problems hamper the perfect reconstruction of features from images.

2.9. Storage Space and Computational Time

The power consumption of devices and computational time are two important factors to consider in vision-based

rehabilitation research as most of the devices are limited in storage space. Being two equivalent factors it is necessary to find the balance between computation and storage to maximize efficiency. In the healthcare system, there will be a huge amount of data from different cameras and sensor from a large number of patients. The models built by researchers should be efficient enough with faster processing time in real-time. The tradeoff between processing time and efficiency should be dealt with care.

3. Available Sensing Devices for Action Understanding

In order to design a vision-based measuring system for creating a dataset or for real-time application, the most important part is the image acquisition device consisting of different types of digital cameras, lens, and image grabber. In addition to this, a high processing computer and an image processing software platform act as the critical parts to obtain the desired parameters in structural monitoring. In previous days, only RGB images were used in most of the computer vision datasets. After incorporating the concepts of depth images and skeleton data, most of the present datasets contain depth images and videos. In this section, we will discuss some sensory types of equipment that are commonly used for data collection in order to capture motion and gesture. These sensors and cameras can be used by researchers who are interested in creating a dataset for capturing body movements during therapy sessions under rehabilitation care. Some of these devices are explored below along with their specifications.

3.1. RGB Video Camera

Earlier vision-based datasets primarily consisted of RGB data using different types of video cameras used for electronic motion picture acquisition. While using RGB cameras there are several factors that need to be taken care of, for example, environmental noise, occlusions, cluttering, and so on [60]. Besides, another important area of research is to choose the number of cameras and fix the distance between subject and cameras. Omnidirectional cameras or also known as 360-degree cameras are also used in computer vision technology. Their field of view covers almost the entire sphere or at least a full circle in the horizontal plane. The FIOR 360 degree, Starcam are some examples of omnidirectional cameras. Webcams are most common while creating datasets on the primary basis for collecting RGB data. Though RGB data collection is a common approach in vision-based research works, this approach has some limitations. The varying length of video sequences affect the recognition result in the case of RGB data. Moreover, it is also difficult to estimate pose using RGB data.

3.2. Depth camera

Depth cameras can be helpful to record depth images and videos along with skeleton data, which can solve the problem of RGB cameras to a large extent. Kinect sensor [5] has been developed by Microsoft, which contains an RGB video camera, a depth sensor, and a multi-array microphone. It detects motion and captures the physical image of a person in front of it. The RGB camera can detect red, green, and blue color components with body-type and facial features. RGB camera has a pixel resolution of 640×480 pixels and 30 fps frame rate. The depth sensor is a combination of a monochrome CMOS sensor and an infrared projector, which helps to create the 3D imagery throughout the room. Depth camera has a distance range of $0.7m \sim 6.0m$. It is possible to measure the distance of each point of the subject's body using this sensor by near-infrared light transmission and measuring its "time of flight" after it reflects off the objects. The microphone array has a feature to isolate the voices of the player from other background noises. We can detect and track 48 different joint points on each subject's body. Kinect sensor have been used in several vision-based research works [61], [30], [35], [4], [66], etc. because of providing depth images and skeleton data. Intel RealSense [7] is another depth camera with stereo D4 vision processor. It has up to 1280×720 depth stream output resolution and the output frame rate is up to 90 fps. Besides, it has a color image signal processor to adjust images and scale color data along with an active infrared projector to illuminate objects to enhance the depth data. Though depth cameras have lots of advantages over RGB cameras, depth cameras do not work well in an outdoor environment and for distant object tracking. After a limited range, we get too much noisy data using a Kinect sensor and RealSense. So, if the data is not available, then it is difficult to decipher.

3.3. Motion Capture System

There are multiview cameras and expensive motion capture systems (MoCap) in order to record the movement of objects or people. They are used in military, entertainment, sports, medical applications, and for validation of computer vision and robotics. We have summarized some motion-sensing systems in this section.

Vicon optical tracking system: This device comes with the feature to capture motion by recording the object or people movements. This device can be used to monitor the gait pattern of patients suffering from Parkinson's disease. The motion capture technology can also be widely used by sports therapists, VFX studios, neuroscientists, and in the field of robotics and computer vision. The powerful new core engine in Vicon tracker can track multiple objects at higher camera counts. This system can process data in 1.5ms, at more than 500 fps. This is up to five times less than other systems. Research work [66] created University

of Idaho-Physical Rehabilitation Movement Dataset (UI-PRMD) using Vicon optical tracker and a Microsoft Kinect sensor for the motion capturing.

Xtion: This device can be used for motion-sensing applications [8]. Xtion PRO LIVE has the feature to sense RGB images to capture the users' image. This can be beneficial to human detection, digital signage, security system, and so on. It can also track people's hand motions without any delay and it has 8 predefined poses. The Xtion PRO LIVE development solution can also be used in developer mode to track a users' whole body movement. This device also has an audio stream application to control voice and any other voice recognition purposes.

3.4. Gaming Interfaces

Gaming interfaces are also used nowadays for collecting data in vision-based research field for tracking gestures, sudden action changes, monitoring hand and leg movements, and so on. We have summarized some gaming interfaces here that can be used for vision-based data collection.

Exergaming Interfaces: Exergames or active video games provide interfaces that need active involvement and the endeavor of physical force by participants [64]. These games are created in such a way that it can track body motions providing fun and exercise opportunity. These gaming interfaces can be used for rehabilitation programs and research. We can engage young patients especially children to play the game and we can collect data regarding their body movement pattern.

PlayStation Move: The PlayStation Move [6] interface contains the Move Eye and the Motion Controller, which is designed by Sony Computer Entertainment. Move Eye is basically an RGB camera with 640x480 pixels at 60 fps or 320x240 pixels at 120 fps. It also contains a directive microphone. The wand includes a 3D accelerometer, a 3D gyro sensor, and a geomagnetic sensor. The Move clearly has the fastest response. This interface is used for 3D gesture and scene recognition. In spite of good performances, this interface is limited to low temporal resolution, difficulty in occluded motion recognition, difficulty in recognition of motion that does not change depth information (e.g., arm axial rotation), and so on.

Eye Toy: This is a color digital camera device compatible with PlayStation2 [3]. This device uses computer vision and gesture recognition technology to process images taken by the camera. This device can be used for motion and color detection. It also contains a microphone for the sound interaction.

3.5. Other Devices

Wii Remote Plus and Sensor Bar: Wii is a small computer based on 700 MHz Power PC processor and 64 MB of external memory [12]. Remote Plus is a Wii controller that contains an IR camera (128x96 pixels), 3D gyro sensor, and 3D accelerometer sensor. Wii Sensor Bar is designed using highlighting IR LED units to estimate controller's position. The unit has IR LED clusters in both sides respectively. This device can be used to detect hand motion with relatively high temporal resolution but it is difficult to detect 3D hand positions.

Wii Fit: Wii Fit contains a balance board as the basic interface with multiple pressure sensors in it [28]. By using the sensor data it is possible to get 2D information of player's gravity center and we can also calculate load transition. Data are transmitted at a high temporal resolution (100 Hz). We can recognize 3D hand motion with high temporal, high spatial resolution, and robustness.

3.6. Typical Environmental Setting

In the field of rehabilitation system-based research, every dataset is dedicated to a specific application, where an environmental setting is an important phenomenon. Environmental setup depends upon what types of action we want to recognize and in which application we want to use the data. Some datasets are taken from real life environment like CCTV footage from shop, gym or training center, whereas, in most of the cases datasets are created manually in the lab settings, considering many ideal cases for research purpose. Selection of environmental background, placement of subject, selecting camera position are the key factors of environmental setup for making a dataset. In Figure 1, we have shown a basic set up for a rehabilitation system environment for data collection using multiple cameras. The movement of the impaired arms or legs of the paralyzed patients are followed by multiple cameras. The simulated environment is shown in the monitor for observation. Datasets are formed by single view or multiple views using multiple cameras. The IXMAX dataset [74] can be a good example in this regard that captured sequences of actions from five different views by setting cameras in different angles of position on the experimental room. In that setup, other parameters were fixed like illumination settings and background setup. The camera positions were also fixed to make the frames scale invariant. The JPL-Interaction dataset [58] is a dedicated dataset to analyze activity from the first-person view. This dataset is aimed to level activities from the first-person view and visualize the activities from that perspective. Similarly, UT-Tower dataset [31] recorded nine types of actions using a stationary camera with jitter.

4. Benchmark Dataset Information

Due to challenges and difficulties regarding technical limitations, privacy issues, organizational authority, etc. in the process of data collection it is difficult to bring a robust research output in the field of advanced healthcare. Because of these difficulties, there are currently few datasets avail-

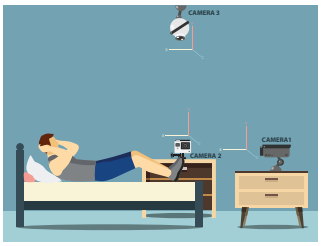

Figure 1. A basic experimental setup to collect data using multiple cameras for a rehabilitation system.

able in this field. In this section, we have gathered benchmark datasets from well-known repositories with relevant information like the number of subjects, gesture, instances, activity labels, and so on. These datasets contain data related to gesture, 3D activities, sports activities, exercise activities, medical activities, etc. that can be used for the research purpose in vision-based rehabilitation care. In this section, we have only summarized the vision-based datasets as there are some problems of wearable sensor-based systems including patient discomfort, missing data, etc. in the field of rehabilitation system-based research. Though some of the datasets mentioned in the following section are not directly related to the healthcare and rehabilitation system, these data can be helpful for the automation of therapy related systems after vital injuries, monitoring the daily living of elderly people to assist them, and coaching the users for behavior modification.

4.1. Gesture Datasets

Gesture related datasets can be used for gesture therapy, gesture recognition to measure the improvement of hand motion after an injury, gesture enabled remote control device for controlling wheel chair, and so on. In Table 1, we have summarized Kinect Gesture Dataset [4], ChaLearn Gesture Dataset (CGD) [2], Microsoft Kinect and Leap Motion [47], Creative Senz3D [49], and Microsoft Research Gesture 3D (MSR Gesture 3D) dataset [38]. These datasets are publicly available.

Table 1. List of publicly available gesture datasets with basic information.

Dataset	Subjects	Gestures	Instances	Year
Kinect Gesture Dataset [4]	30	12	6244	2016
CGD [2]	N.A	15	50,000	2011
Microsoft Kinect and Leap Motion [47]	14	140	1400	2014
Creative Senz3D [49]	4	44	1320	2015
MSR Gesture3D [38]	10	12	336	2012

N.A: Not Available.

4.2. 3D Action Datasets

In this section, we have listed some 3D action and activity datasets namely University of Wollongong Online Action3D Dataset [78], Microsoft Research (MSR) Action3D [41], Microsoft Research (MSR) Daily Activity 3D [68], and Multi-view Action Modeling and Detection dataset [69]. These datasets contain skeleton data, which can be used to detect human pose and actions more precisely. These actions can be monitored to coach the elderly people for behavior modification. For example, it is possible to monitor the action of "drinking water" and "taking medicine" so that we can coach the user to perform these actions properly for maintaining a healthy life. It is also possible to coach the users to walk and using stairs, if any person sits for a long time in a day by monitoring the action "working in a computer" and "using elevators". We have shown the basic information regarding these datasets in Table 2. These datasets are publicly available.

Table 2. List of publicly available 3D action datasets with basic information.

Dataset	Subjects	Activities	Files	Year
UOW Online Action3D [78]	20	20	N.A	2016
MSR Action3D [41]	10	20	567	2010
MSR Daily Activity3D [68]	10	16	960	2012
Multi-view Action Modelling and Detection [69]	10	10	N.A	2014

N.A: Not Available.

4.3. Sports Activity Related Datasets

Sports activity related datasets can also be used in this field of research to recognize human motion while performing sports activity. We can estimate based on the fitness of a person, either a specific sport activity is suitable for him to perform or not. In Table 3, we have summarized some sports activity related datasets namely University of Illinois at Urbana-Champaign (UIUC1 and UIUC2) datasets [65], University of Central Florida (UCF) dataset [55], UCF50 [54], and UCF-YouTube [43] dataset. These datasets are publicly available.

Table 3. List of publicly available sports datasets with basic information.

Dataset	Average duration	Video labels	Resolution	Frame per second (fps)	Year
UIUC1 UIUC2 [65]	3s	14	1024×768	15	2008
UCF [55]	3s	13	480×360	25	2008
UCF50 [54]	7s	50	320×240	30	2012
UCF YouTube [43]	6s	11	320×240	25	2009

4.4. Exercise Related Datasets

Exercise related datasets are very important to monitor the exercise patterns of injured patients in the rehabilitation research field. In Table 4, we have given some basic information about exercise related datasets namely KTH dataset

[39], i3DPost Multi-view Human Action [32], CASIA action database for recognition [72], and Tsinghua-Daimler Cyclist Detection Benchmark Dataset [42]. These datasets are publicly available.

Table 4. List of publicly available exercise datasets with basic information.

Dataset	Subjects	Actions	Sequences	Year
KTH [39]	25	6	2391	2004
i3DPost [32]	8	13	N.A	2009
CASIA [72]	24	15	1446	2007
Tsinghua-Daimler Cyclist Detection [42]	N.A	6	14674	2016

N.A: Not Available.

4.5. Medical Activity Related Dataset

In this section, we have summarized some datasets related to medical activities. In Table 5, we have given some basic information about University of Southern California (USC) Collaborative Research in Computational Neuroscience (CRCNS) dataset [19], Human Mortality Database (HMD) [29], and University of Idaho-Physical Rehabilitation Movement Data (UI-PRMD) [66]. These datasets are publicly available. These datasets are associated with general exercises performed by patients in physical rehabilitation programs and hospital environments. The objective of these datasets is to evaluate patient consistency while performing the prescribed rehabilitation exercises.

Table 5. List of publicly available medical datasets with basic information.

Dataset	Subjects	Activity	Application	Year
USC CRCNS [19]	520	Eye movement	Healthcare (detecting and monitoring eye problems)	2004-05
HMD [29]	59	Head movement	Health support (monitoring paralyzed patients)	2017
UI-PRMD [66]	10	Rehabilitation movement	Monitoring rehabilitation exercise	2018

4.6. Assistive Living Datasets

Most of the patients loss their ability to perform daily activities like cooking, brushing teeth, etc. after a severe injury especially related to hand. In this section, we have analyzed some daily activity related dataset to monitor the improvements of patients after providing therapy while performing these activities. We can also use these data to monitor the daily lifestyle of elderly people to maintain a healthy lifestyle and prevent accidents. Table 6 represents some benchmark datasets namely Breakfast Actions database [37], Max-Planck-Institut fur Informatik: MPII Cooking Activities Dataset [56], and Carnegie Melon University Multi-Modal Activity Database (CMU-MMAC) [53].

Table 6. List of publicly available exercise datasets with basic information.

Dataset	Video labels	Frame labels	Number of cameras	Year
Breakfast Actions [37]	10	48	4	2014
MPII Cooking Activity [56]	65	14	1	2012
CMU-MMAC [53]	43	5	5	2008

4.7. Emotion Recognition Related Dataset

Among the patients, who take rehabilitation treatment, a large number of people suffer from depression or mental stress that hampers their daily lives. Emotion and sentiment analysis can be a good technique to give them proper treatment that can improve emotional healthcare system. Table 7 represents some datasets namely The Amsterdam Dynamic Facial Expression Set (ADFES) dataset [67], Binghamton University 3D Facial Expression (BU-3DFE) dataset [77], Database of Kinetic Facial Expressions (DaFEx) [16], and STOIC dataset (A database of dynamic and static faces) [57] for emotion recognition.

Table 7. List of publicly available emotion recognition datasets with basic information.

Dataset	Subject	Number of emotions	Age	Year
Adfes [67]	22	9	18-25	2011
BU-4dfe [77]	100	6	18-70	2006
DaFEx [16]	8	7	25	2005
STOIC [57]	10	8	20-45	2007

5. Discussion

Vision-based datasets are made from different types of camera sensors and from different views or angles of cameras. Using these datasets, it is possible to make suitable models for desired activity recognition and the accuracy of models mostly depends on how much suitable data have chosen by the researchers. As now, in this paper we are focusing on vision-based rehabilitation, we have focused our vision to compare the previous works on some dedicated datasets for healthcare and rehabilitation system-based research in this section. We may also use other activity recognition datasets that are not dedicated to rehabilitation for the automation of assistive living and healthcare facilities. In that case, we need to process the data with some image processing algorithm for recognizing the activities that may be helpful in rehabilitation system-based research. HPTE (Home-based Physical Therapy Exercises) data set is dedicated to therapy actions [14]. It is recorded with a Kinect camera and it only provides video and depth streams. This dataset is made by five subjects and each of them performed each action six times and they give eight shoulder and knee exercise movements. The EmoPain data set [15] contains full body joint motions and high-resolution face videos. This dataset was mainly designed with a view to analyz-

ing the pain-related emotions in the rehabilitation system. A group of 28 healthy subjects and 22 patients performed 7 exercises. The patients were suffering from chronic lower back pain. The AHA-3D dataset [13] depicts both young and elderly subjects. Their performed exercise data have been taken in the form of 3D Skeleton data. There are 79 Skeleton videos and each contains 1 to 3 runs of the same exercise. In total, 21 subjects (5 male and 16 female) performed the exercises and among them 11 were young and 10 were elderly. The UI-PRMD data set [66] is the most popular and dedicated dataset for rehabilitation activity. This dataset is made by 10 subjects and they performed common physical rehabilitation exercises. The movements executed by the participants were collected concurrently with the Vicon and Kinect systems. The objective of the data set is to afford a basis for mathematical modeling of therapy actions, as well as for establishing performance metrics for the evaluation of patient consistency while performing the prescribed rehabilitation exercises.

There are also some gait databases that are publicly available. Gait analysis could enable us to finely characterize gait singularities to pinpoint potential diseases or abnormalities in advance. CASIA gait database [1] is a very popular database for gait recognition. Loughborough University's institutional repository suggested a database made by Kinect sensor for gait recognition. A group of twenty participants (12 men and 8 women) performed 10 activities set, finally resulting in the capture of 200 activities with a total of 60,225 frames [40]. 122 subjects were used to create the USF dataset by the University of South Florida. Each person walked around an ellipse in front of cameras [59] and collected 1870 sequences in total. Robotics Institute, Carnegie Mellon University prepared another dataset namely CMU Mobo. A 3D room had been used to take data from 25 subjects by walking in a treadmill. There are four different walking patterns for each individual: slow walk, fast walk, incline walk, and walking with a ball [36]. Southampton Dataset consists of 12 subjects walking around an inside track at different speeds. Data from each person were taken wearing various shoes, clothes, and without or with various bags [51].

In this section, we have also analyzed previous works on some benchmark data sets namely Microsoft Research (MSR) Action3D [65], MSR Daily Activity3D [55], University of Central Florida Kinect (UCF-Kinect) [30], Multiview 3D Event [73], University of Texas Kinect (UT-Kinect) [75], Berkeley Multimodal Human Action Database (MHAD) [53], University of Texas Dallas Multimodal Human Action Dataset (UTD-MHAD) [20], and AHA-3D dataset [15]. We have summarized the previous research works in these datasets in Table 8.

Table 8. Analysis of previous vision-based research works on some benchmark datasets for the healthcare and rehabilitation treatment.

Dataset	Method/Model	Accuracy (%)
MSRAction3D [65]	Convolutional Neural Networks (ConvNets) [70] [71]	100
	TriViews + Portable Format for Analytics (PFA) [22]	98.2
	Decision-Level Fusion (SUM) [79]	98.2
MSR Daily Activity3D [55]	τ-test [44]	95.63
	DL-GSGC + Total Productive Maintenance (TPM) [45]	95
	3D joint + CS-MLtp [46]	92.5
	Depth Volumetric Spatial Feature Representation (VSFR) [26]	89.7
UCF-Kinect [30]	Hierarchical model [35]	98.7
Multiview 3D Event [73]	4D human-object interaction (4DHOI) [73]	87
UT-Kinect [75]	Grassman manifold [61]	95.25
Berkeley MHAD [53]	Hierarchical hidden Markov model (HHMM) [53]	79.1
UTD-MHAD [20]	Convolutional Neural Networks (CNN) [33]	99.54
AHA-3D [15]	Kolmogorov-Smirnov test	88.29

6. Conclusion and Future Challenges

This paper provides a review of the research and progress in the area of advanced assistive healthcare and rehabilitation system by using the computer vision-based technology. We have discussed challenges of this field that are related to the data collection, class variations, environmental noises, positions of installed cameras, rules and regulations of hospital authority, and so on. The aim of this survey is to analyze possible solutions of this challenges so that researchers can work for the automation of healthcare system and rehabilitation facility. We have discussed about some available sensing devices based on vision-based technology that can be used by researchers while making training data. Besides, benchmark datasets related to gesture, medical, sports, exercise, 3D action, and emotion have been described with useful informations so that researchers can find it useful to compare and select the perfect dataset according to the field of research. The performance of previous research works has been analyzed based on their model selection and accuracy that can be useful in the case of comparison with new models for future researchers.

Due to the existence of major challenges in the path of advancement of computer vision-based rehabilitation system, there are still a few robust and good research work in this domain. There are still options for lots of progress in this field for the researchers. Real-time performance, activity monitoring of multiple patients at the same time, improvement score evaluation after therapeutic movements using the regression method, good performance in low light, etc. can be some bigger challenges that need to be addressed in future for obtaining good performances in the automation of the the healthcare system.

References

[1] CASIA gait database. http://www.

sinobiometrics.com. Accessed: 2019-02-25. 8

[2] ChaLearn gesture dataset. http://gesture.chalearn.org/data. Accessed: 2019-01-20. 6

[3] Eye toy. https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/. Accessed: 2019-01-13. 5

[4] Kinect gesture dataset. https://www.microsoft.com/en-us/download. Accessed: 2019-01-15. 4, 6

[5] Kinect sensor. https://www.microsoft.com/en-us/store/d/kinect-sensor-for-xbox-one/91hq5578vksc. Accessed: 2019-01-10. 4

[6] Playstation move. https://www.playstation.com/en-ae/explore/accessories/playstation-move-motion-controller/. Accessed: 2019-01-08. 5

[7] RealSense depth camera. https://click.intel.com/intelr-realsensetm-depth-camera-d435.html. Accessed: 2019-02-25. 4

[8] Xtion pro live. https://www.asus.com/3D-Sensor/Xtion_PRO/. Accessed: 2019-01-12. 5

[9] Md Atiqur Ahad. *Computer vision and action recognition: A guide for image processing and computer vision community for action understanding*, volume 5. Springer Science & Business Media, 2011. 2

[10] Md Atiqur Rahman Ahad. *Motion history images for action recognition and understanding*. Springer Science & Business Media, 2012. 2

[11] Mohiuddin Ahmad and Seong-Whan Lee. Human action recognition using shape and clg-motion flow from multi-view image sequences. *Pattern Recognition*, 41(7):2237–2252, 2008. 2

[12] Fraser Anderson, Michelle Annett, and Walter F Bischof. Lean on wii: physical rehabilitation with virtual reality wii peripherals. *Stud Health Technol Inform*, 154(154):229–34, 2010. 5

[13] João Antunes, Alexandre Bernardino, Asim Smailagic, and Daniel P. Siewiorek. Aha-3d: A labelled dataset for senior fitness exercise recognition and segmentation from 3d skeletal data. In *BMVC*, 2018. 8

[14] Ilktan Ar and Yusuf Sinan Akgul. A computerized recognition system for the home-based physiotherapy exercises using an rgbd camera. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(6):1160–1171, 2014. 7

[15] Min SH Aung, Sebastian Kaltwang, Bernardino Romera-Paredes, Brais Martinez, Aneesha Singh, Matteo Cella, Michel Valstar, Hongying Meng, Andrew Kemp, Moshen Shafizadeh, et al. The automatic detection of chronic pain-related expression: requirements, challenges and the multimodal emopain dataset. *IEEE transactions on affective computing*, 7(4):435–451, 2016. 7, 8

[16] Alberto Battocchi, Fabio Pianesi, and Dina Goren-Bar. Dafex: Database of facial expressions. In *International Conference on Intelligent Technologies for Interactive Entertainment*, pages 303–306. Springer, 2005. 7

[17] Matteo Bregonzio, Shaogang Gong, and Tao Xiang. Recognising action as clouds of space-time interest points. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1948–1955. IEEE, 2009. 2

[18] Fabrice Caillette, Aphrodite Galata, and Toby Howard. Real-time 3-d human body tracking using learnt models of behaviour. *Computer Vision and Image Understanding*, 109(2):112–125, 2008. 2

[19] Ran Carmi and Laurent Itti. The role of memory in guiding attention during natural vision. *Journal of vision*, 6(9):4–4, 2006. 7

[20] Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 168–172. IEEE, 2015. 8

[21] Ching-Hui Chen, Julien Favre, Gregorij Kurillo, Thomas P Andriacchi, Ruzena Bajcsy, and Rama Chellappa. Camera networks for healthcare, teleimmersion, and surveillance. *Computer*, 47(5):26–36, 2014. 2

[22] Wenbin Chen and Guodong Guo. Triviews: A general framework to use 3d depth data effectively for action recognition. *Journal of Visual Communication and Image Representation*, 26:182–191, 2015. 8

[23] Srikanth Cherla, Kaustubh Kulkarni, Amit Kale, and Viswanathan Ramasubramanian. Towards fast, view-invariant human action recognition. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2008. 2

[24] RM Chestnut, N Carney, H Maynard, P Patterson, NC Mann, and M Helfand. Rehabilitation for traumatic brain injury: Summary. 1998. 1

[25] Tat-Jun Chin, Liang Wang, Konrad Schindler, and David Suter. Extrapolating learned manifolds for human activity recognition. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 1, pages I–381. IEEE, 2007. 2

[26] Seong-Sik Cho, A-Reum Lee, Heung-Il Suk, Jeong-Seon Park, and Seong-Whan Lee. Volumetric spatial feature representation for view-invariant human action recognition using a depth camera. *Optical Engineering*, 54(3):033102, 2015. 8

[27] Olivier Chomat, Jéer?e Martin, and James L Crowley. A probabilistic sensor for the perception and the recognition of activities. In *European Conference on Computer Vision*, pages 487–503. Springer, 2000. 2

[28] Ross A Clark, Adam L Bryant, Yonghao Pua, Paul McCrory, Kim Bennell, and Michael Hunt. Validity and reliability of the nintendo wii balance board for assessment of standing balance. *Gait & posture*, 31(3):307–310, 2010. 5

[29] Xavier Corbillon, Francesca De Simone, and Gwendal Simon. 360-degree video head movement dataset. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, pages 199–204. ACM, 2017. 7

[30] Chris Ellis, Syed Zain Masood, Marshall F Tappen, Joseph J LaViola, and Rahul Sukthankar. Exploring the trade-off be-

tween accuracy and observational latency in action recognition. *International Journal of Computer Vision*, 101(3):420–436, 2013. 4, 8

[31] Antonio Fernández-Caballero, José Carlos Castillo, and José María Rodríguez-Sánchez. A proposal for local and global human activities identification. In *International Conference on Articulated Motion and Deformable Objects*, pages 78–87. Springer, 2010. 5

[32] Nikolaos Gkalelis, Hansung Kim, Adrian Hilton, Nikos Nikolaidis, and Ioannis Pitas. The i3dpost multi-view and 3d human action/interaction database. In *Visual Media Production, 2009. CVMP'09. Conference for*, pages 159–168. IEEE, 2009. 7

[33] Earnest Paul Ijjina and C Krishna Mohan. Human action recognition based on mocap information using convolution neural networks. In *Machine Learning and Applications (ICMLA), 2014 13th International Conference on*, pages 159–164. IEEE, 2014. 8

[34] Kirsten Jack, Sionnadh Mairi McLean, Jennifer Klaber Moffett, and Eric Gardiner. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. *Manual therapy*, 15(3):220–228, 2010. 1

[35] Xinbo Jiang, Fan Zhong, Qunsheng Peng, and Xueying Qin. Robust action recognition based on a hierarchical model. In *Cyberworlds (CW), 2013 International Conference on*, pages 191–198. IEEE, 2013. 4, 8

[36] Amir Kale, Naresh Cuntoor, B Yegnanarayana, AN Rajagopalan, and Rama Chellappa. Gait analysis for human identification. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 706–714. Springer, 2003. 8

[37] Hilde Kuehne, Ali Arslan, and Thomas Serre. The language of actions: Recovering the syntax and semantics of goal-directed human activities. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 780–787, 2014. 7

[38] Alexey Kurakin, Zhengyou Zhang, and Zicheng Liu. A real time system for dynamic hand gesture recognition with a depth sensor. In *EUSIPCO*, volume 2, page 6, 2012. 6

[39] Ivan Laptev. *Local spatio-temporal image features for motion interpretation*. PhD thesis, Numerisk analys och datalogi, 2004. 7

[40] Daniel Leightley, John Darby, Baihua Li, Jamie S McPhee, and Moi Hoon Yap. Human activity recognition for physical rehabilitation. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pages 261–266. IEEE, 2013. 8

[41] Wanqing Li, Zhengyou Zhang, and Zicheng Liu. Action recognition based on a bag of 3d points. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 9–14. IEEE, 2010. 6

[42] Xiaofei Li, Fabian Flohr, Yue Yang, Hui Xiong, Markus Braun, Shuyue Pan, Keqiang Li, and Dariu M Gavrila. A new benchmark for vision-based cyclist detection. In *Intelligent Vehicles Symposium (IV), 2016 IEEE*, pages 1028–1033. IEEE, 2016. 7

[43] Jingen Liu, Jiebo Luo, and Mubarak Shah. Recognizing realistic actions from videos "in the wild". In *Computer vision and pattern recognition, 2009. CVPR 2009. IEEE conference on*, pages 1996–2003. IEEE, 2009. 6

[44] Cewu Lu, Jiaya Jia, and Chi-Keung Tang. Range-sample depth feature for action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 772–779, 2014. 8

[45] Jiajia Luo, Wei Wang, and Hairong Qi. Group sparsity and geometry constrained dictionary learning for action recognition from depth maps. In *Proceedings of the IEEE international conference on computer vision*, pages 1809–1816, 2013. 8

[46] Kiajia Luo, Wei Wang, and Hairong Qi. Spatio-temporal feature extraction and representation for rgb-d human action recognition. *Pattern Recognition Letters*, 50:139–148, 2014. 8

[47] Giulio Marin, Fabio Dominio, and Pietro Zanuttigh. Hand gesture recognition with leap motion and kinect devices. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 1565–1569. IEEE, 2014. 6

[48] Eric McAdams, Asta Krupaviciute, Claudine Géhin, Etienne Grenier, Bertrand Massot, André Dittmar, Paul Rubel, and Jocelyne Fayn. Wearable sensor systems: The challenges. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3648–3651. IEEE, 2011. 2

[49] L Minto and P Zanuttigh. Exploiting silhouette descriptors and synthetic data for hand gesture recognition. 2015. 6

[50] Thanh Trung Ngo, Yasushi Makihara, Hajime Nagahara, Yasuhiro Mukaigawa, and Yasushi Yagi. The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. *Pattern Recognition*, 47(1):228–237, 2014. 3

[51] Mark S Nixon and John N Carter. Automatic recognition by gait. *Proceedings of the IEEE*, 94(11):2013–2024, 2006. 8

[52] Daigo Muramatsu Tomio Echigo Yasushi Yagi Noriko Takemura, Yasushi Makihara. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Trans. on Computer Vision and Applications*, 10(4):1–14, 2018. 3

[53] Ferda Ofli, Rizwan Chaudhry, Gregorij Kurillo, René Vidal, and Ruzena Bajcsy. Berkeley mhad: A comprehensive multimodal human action database. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pages 53–60. IEEE, 2013. 7, 8

[54] Kishore K Reddy and Mubarak Shah. Recognizing 50 human action categories of web videos. *Machine Vision and Applications*, 24(5):971–981, 2013. 6

[55] Mikel D Rodriguez, Javed Ahmed, and Mubarak Shah. Action mach a spatio-temporal maximum average correlation height filter for action recognition. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. 6, 8

[56] Marcus Rohrbach, Sikandar Amin, Mykhaylo Andriluka, and Bernt Schiele. A database for fine grained activity detection of cooking activities. In *Computer Vision and Pat-*

tern Recognition (CVPR), 2012 IEEE Conference on, pages 1194–1201. IEEE, 2012. 7

[57] Sylvain Roy, Cynthia Roy, Isabelle Fortin, Catherine Ethier-Majcher, Pascal Belin, and Frederic Gosselin. A dynamic facial expression database. *Journal of Vision*, 7(9):944–944, 2007. 7

[58] MS Ryoo, Thomas J Fuchs, Lu Xia, Jake K Aggarwal, and Larry Matthies. Robot-centric activity prediction from first-person videos: What will they do to me? In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 295–302. ACM, 2015. 5

[59] Sudeep Sarkar, P Jonathon Phillips, Zongyi Liu, Isidro Robledo Vega, Patrick Grother, and Kevin W Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE transactions on pattern analysis and machine intelligence*, 27(2):162–177, 2005. 8

[60] L. Shao, J. Han, D. Xu, and J. Shotton. Computer vision for rgb-d sensors: Kinect and its applications [special issue intro.]. *IEEE Transactions on Cybernetics*, 43(5):1314–1317, Oct 2013. 4

[61] Rim Slama, Hazem Wannous, and Mohamed Daoudi. Grassmannian representation of motion depth for 3d human gesture and action recognition. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 3499–3504. IEEE, 2014. 4, 8

[62] L Enrique Sucar, Gildardo Azcárate, Ron S Leder, David Reinkensmeyer, Jorge Hernández, Israel Sanchez, and Pedro Saucedo. Gesture therapy: A vision-based system for arm rehabilitation after stroke. In *International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 531–540. Springer, 2008. 1

[63] Fabrizio Taffoni, Diego Rivera, Angelica La Camera, Andrea Nicolò, Juan Ramón Velasco, and Carlo Massaroni. A wearable system for real-time continuous monitoring of physical activity. *Journal of healthcare engineering*, 2018, 2018. 2

[64] Kazumoto Tanaka, JR Parker, Graham Baradoy, Dwayne Sheehan, John R Holash, and Larry Katz. A comparison of exergaming interfaces for use in rehabilitation programs and research. *Loading...*, 6(9), 2012. 5

[65] Du Tran and Alexander Sorokin. Human activity recognition with metric learning. In *European conference on computer vision*, pages 548–561. Springer, 2008. 6, 8

[66] Aleksandar Vakanski, Hyung-pil Jun, David Paul, and Russell Baker. A data set of human body movements for physical rehabilitation exercises. *Data*, 3(1):2, 2018. 2, 4, 7, 8

[67] Job Van Der Schalk, Skyler T Hawk, Agneta H Fischer, and Bertjan Doosje. Moving faces, looking places: validation of the amsterdam dynamic facial expression set (adfes). *Emotion*, 11(4):907, 2011. 7

[68] Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan. Mining actionlet ensemble for action recognition with depth cameras. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1290–1297. IEEE, 2012. 6

[69] Jiang Wang, Xiaohan Nie, Yin Xia, Ying Wu, and Song-Chun Zhu. Cross-view action modeling, learning and recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2649–2656, 2014. 6

[70] Pichao Wang, Wanqing Li, Zhimin Gao, Chang Tang, Jing Zhang, and Philip Ogunbona. Convnets-based action recognition from depth maps through virtual cameras and pseudocoloring. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 1119–1122. ACM, 2015. 8

[71] Pichao Wang, Wanqing Li, Zhimin Gao, Jing Zhang, Chang Tang, and Philip O Ogunbona. Action recognition from depth maps using deep convolutional neural networks. *IEEE Transactions on Human-Machine Systems*, 46(4):498–509, 2016. 8

[72] Ying Wang, Kaiqi Huang, and Tieniu Tan. Human activity recognition based on r transform. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007. 7

[73] Ping Wei, Yibiao Zhao, Nanning Zheng, and Song-Chun Zhu. Modeling 4d human-object interactions for event and object recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3272–3279, 2013. 8

[74] Daniel Weinland, Remi Ronfard, and Edmond Boyer. Free viewpoint action recognition using motion history volumes. *Computer vision and image understanding*, 104(2-3):249–257, 2006. 5

[75] Lu Xia, Chia-Chih Chen, and Jake K Aggarwal. View invariant human action recognition using histograms of 3d joints. In *Computer vision and pattern recognition workshops (CVPRW), 2012 IEEE computer society conference on*, pages 20–27. IEEE, 2012. 8

[76] Xiao-Wei Ye, CZ Dong, and T Liu. A review of machine vision-based structural health monitoring: Methodologies and applications. *Journal of Sensors*, 2016, 2016. 2

[77] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J Rosato. A 3d facial expression database for facial behavior research. In *Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on*, pages 211–216. IEEE, 2006. 7

[78] Jing Zhang, Wanqing Li, Pichao Wang, Philip Ogunbona, Song Liu, and Chang Tang. A large scale rgb-d dataset for action recognition. In *International Workshop on Understanding Human Activities through 3D Sensors*, pages 101–114. Springer, 2016. 6

[79] Yu Zhu, Wenbin Chen, and Guodong Guo. Fusing multiple features for depth-based action recognition. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(2):18, 2015. 8