



Article

An Upper Extremity Rehabilitation System Using Efficient Vision-Based Action Identification Techniques

Yen-Lin Chen ¹ , Chin-Hsuan Liu ¹ , Chao-Wei Yu ¹, Posen Lee ^{2,*} and Yao-Wen Kuo ¹

¹ Department Computer Science and Information Engineering, National Taipei University of Technology, Taipei 10608, Taiwan; ylchen@csie.ntut.edu.tw (Y.-L.C.); chinhhsuanliu@gmail.com (C.-H.L.); david741002@gmail.com (C.-W.Y.); kent21221@gmail.com (Y.-W.K.)

² Department Occupational Therapy, I-Shou University, Kaohsiung 82445, Taiwan

* Correspondence: posenlee@isu.edu.tw; Tel.: +886-7-657-7711 (ext. 7516)

Received: 30 May 2018; Accepted: 10 July 2018; Published: 17 July 2018



Featured Application: This study proposes an upper extremity rehabilitation system using efficient action identification system for home based on color and depth sensor information, and can perform well under complex ambient environments.

Abstract: This study proposes an action identification system for home upper extremity rehabilitation. In the proposed system, we apply an RGB-depth (color-depth) sensor to capture the image sequences of the patient's upper extremity actions to identify its movements. We apply a skin color detection technique to assist with extremity identification and to build up the upper extremity skeleton points. We use the dynamic time warping algorithm to determine the rehabilitation actions. The system presented herein builds up upper extremity skeleton points rapidly. Through the upper extremity of the human skeleton and human skin color information, the upper extremity skeleton points are effectively established by the proposed system, and the rehabilitation actions of patients are identified by a dynamic time warping algorithm. Thus, the proposed system can achieve a high recognition rate of 98% for the defined rehabilitation actions for the various muscles. Moreover, the computational speed of the proposed system can reach 125 frames per second—the processing time per frame is less than 8 ms on a personal computer platform. This computational efficiency allows efficient extensibility for future developments to deal with complex ambient environments and for implementation in embedded and pervasive systems. The major contributions of the study are: (1) the proposed system is not only a physical exercise game, but also a movement training program for specific muscle groups; (2) The hardware of upper extremity rehabilitation system included a personal computer with personal computer and a depth camera. These are economic equipment, so that patients who need this system can set up one set at home; (3) patients can perform rehabilitation actions in sitting position to prevent him/her from falling down during training; (4) the accuracy rate of identifying rehabilitation action is as high as 98%, which is sufficient for distinguishing between correct and wrong action when performing specific action trainings; (5) The proposed upper extremity rehabilitation system is real-time, efficient to vision-based action identification, and low-cost hardware and software, which is affordable for most families.

Keywords: upper extremity identification; color and depth sensors; skeleton points; rehabilitation actions; home rehabilitation; computer vision

1. Introduction

Telemedicine and home-care systems have become the trend because of the integration of technology into the practice of medicine [1–5]. Some medical care can be applied at home using some simple system. If technology products are easy to operate, then patients or caregivers can conveniently perform daily self-care activities at home by themselves. This can improve the time spent on patient care, the quality of care, and the therapeutic benefits of care. Telemedicine and home care systems can also reduce the costs and time associated with transportation between the hospital and home for follow-up care or treatment [6–9].

In rehabilitation, the person receiving treatment exhibits impaired mobility. Treatment typically takes a long time to achieve a positive effect, and if the treatment is stopped or interrupted, then a functional decline or reversal of progress may occur [10,11]. Therefore, rehabilitation is a very lengthy process that takes a heavy psychological, physical, and economic toll on the patients and families of patients [12]. If an efficient rehabilitation system can be designed for use at home that helps patients to perform the movements that they must perform repeatedly every day to maintain their mobility and physical function, then with such a system, transportation challenges and costs are eliminated. Moreover, infection risks for patients with weakened immune systems can be avoided because frequent hospital stays are eliminated too. A home-based rehabilitation program can also make rehabilitation more flexible and enables more frequent exercise. In sum, transportation burdens are reduced, families have more time, and exercises can be more frequently performed.

Rehabilitation involves the use of repetitive movements for maintaining or improving physical or motor functions [10]. After receiving a professional evaluation and a recommended exercise regimen to perform at home, a patient may only need to be observed and recorded by a rehabilitation supporting instrument during the basic training sections performed at home. A professional demonstrates an action or use of an instrument, and the patient can then operate a quality monitoring system or review visual feedback provided by a home rehabilitation system for quality self-monitoring of the motions performed during the daily rehabilitation training at home [13].

Some somatosensory games claim that they can achieve the effects of sports and entertainment, and some of these have been used for physical training and physical rehabilitation [14–16]. However, these games are usually designed to involve whole body activities, and most require a standing position [17–22] often in front of a camera at a distance of at least 100 cm. Beside most of those somatosensory games and human pose estimation are focus on whole body pose discrimination [22] neither pay attention on changes of range of motion of joints nor think about how muscles works in those poses.

Nowadays, there are many human pose estimation systems that extract skeletons or skeleton points from depth sensors [18–23]. However, determining the joint movements (include directions and angles) are necessary for rehabilitation applications. In [18–20], they derive a unified gaussian kernel correlation (GKC) representation and develop an articulated GKC and articulated pose estimation for both the full body and the hands. That study can achieve effective human pose estimation and tracking, but may have limitations in single joint movement determination, such as shoulder rotation, wrist pronation. In [21], the input depth map is extracted to match with a set of pre-captured motion exemplars to generate a body configuration estimation, as well as semantic labeling of the input point cloud, but as we know that body figure can be shown or defined as a point cloud but the real body skeletons is segmental, each movement comes from the angle change of the key joint rather than each point cloud of the body. In [23], the system grabs a color image as input and extract the 2D locations of anatomical keypoints, and then applies an architecture for jointly learning parts detection and parts association. The pros of [23] is 2D pose estimation of multiple people in images, that can be applied in social context identification assisting system for Autism or security monitoring system for home or public spaces, but may not be sufficiently accurate for medical assessment. In our study, we extract both color and depth information from RGB-D sensors, and rebuild a new upper extremity skeleton and skeleton points. To provide an efficient rehabilitation system that can assist the patients to train

specific joints and muscles, we need to obtain not only the pose changes, but also to find out these changes coming from which joints in real human body skeletons.

Although there are many pose estimation systems and somatosensory games have been developed and presented, most of those existent systems are performed based on extracting skeletons or skeleton points from a depth sensor and mainly focus on providing human pose estimation, rather than determining the joint movements (include directions and angles). In this study, we desire to obtain not only the pose changed, but also try to find out these changes come from which joint in real human body skeleton. Most of those systems and game sets may not focus on training some specific actions for the purpose of medical rehabilitation, have insufficient evidence base on medical applications, and might be unsuitable for someone who are unable to stand or stand for long, lack of standing balance, or only need to rehabilitate their upper extremity, such as patients with spinal cord injuries, hemiplegia, or advanced age.

To overcome the aforementioned challenges, this study proposes an action identification system for home upper extremity rehabilitation, which is a movement training program for specific joints movement and muscle groups, and have advantages such as cost-economic, suitable for sitting position rehabilitations, and easy for operating at home. The major contributions of the study are: (1) the proposed system is not only a physical exercise game, but also a movement training program for specific muscle groups; (2) The hardware of upper Extremity Rehabilitation System included a personal computer with personal computer and a depth camera. These are economic equipment, so that patients who need this system can set up one set at home; (3) patients can perform rehabilitation actions in sitting position to prevent him/her from falling down during training; (4) the accuracy rate of identifying rehabilitation action is as high as 98%, which is sufficient for distinguishing between correct and wrong action when performing specific action trainings; (5) The proposed upper extremity rehabilitation system is real-time, efficient to vision-based action identification, and low-cost hardware and software, which is affordable for most families.

2. The Proposed Action Identification System for Home Upper Extremity Rehabilitation

In the proposed system, we apply an RGB-depth sensor to capture the image sequences of the patient’s upper extremity actions to identify its movements, and a skin color detection technique is used to assist with extremity identification and to build up the upper extremity skeleton points. A dynamic time warping algorithm is used to determine the rehabilitation actions (Figure 1).

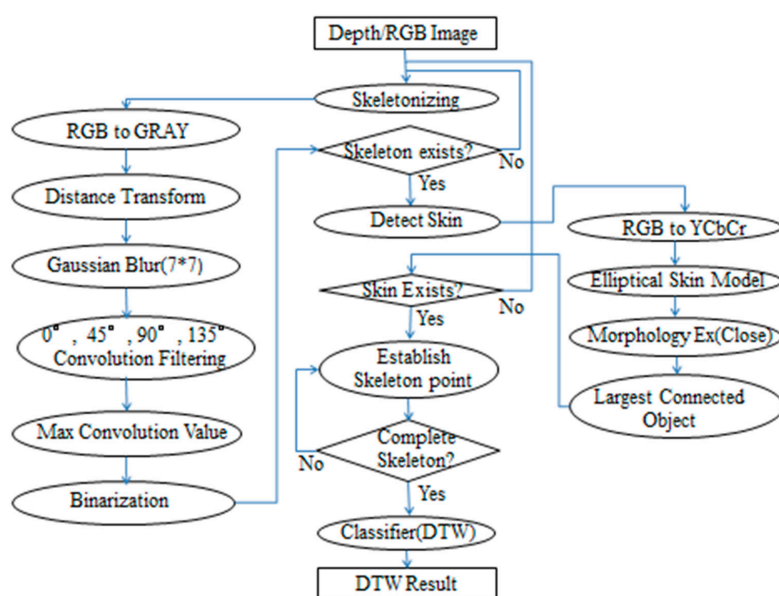


Figure 1. System block diagram of the proposed upper extremity rehabilitation system.

2.1. Experimental Environment

We suppose the user is sitting in front of the table to use this system. If the patient performs rehabilitation actions in a sitting position, the patient saves energy compared with standing, and the patient can focus his or her mind on the motion control training of the upper extremities. Moreover, if patients have limited balance abilities, they can sit at a table or desk, and the table or desk can provide support and prevent falls. The proposed system is set up similarly to a study desk or computer table at home, which provides a familiar and convenient environment. Caregivers are not required to prepare an additional space to set up the rehabilitation system, and the patient can engage in rehabilitation in a small space (Figure 2).

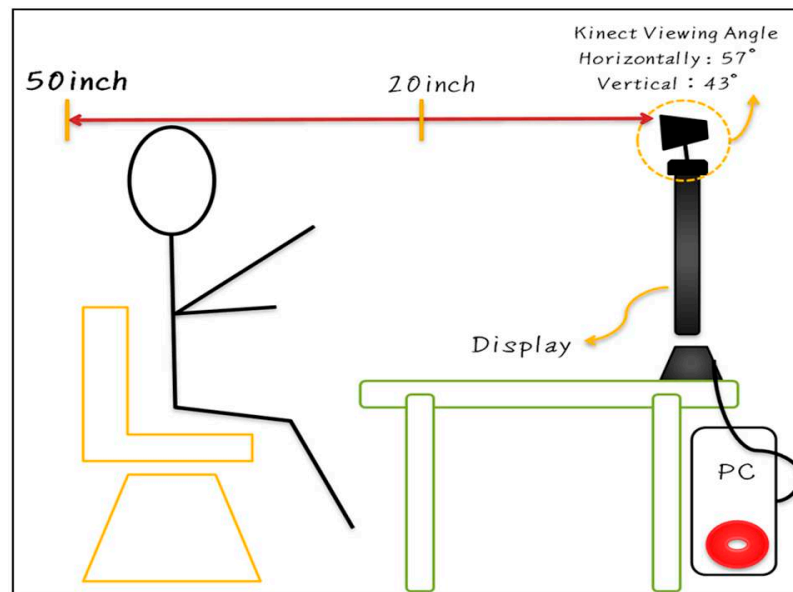


Figure 2. Conceptual lab setup.

The hardware of the upper extremity rehabilitation system includes a personal computer and a Kinect camera.

The first step of the vision-based action identification technique is to identify the bones of the patient. Many methods exist to accomplish this task. In this study, we suppose that the user's sitting posture is similar to that of sitting at a table. In this situation, the bones of their upper extremities can be identified accordingly. Second, through the human skin detection approach, skeleton joint points can be determined as well. After the human skeletal structure is established, motion determination is conducted. Through the motion determination process, we can determine what type of motion the user just performed, and then inform the patient or caregiver whether the rehabilitation action was correct or not. The main purpose of these processes is to create a home rehabilitation system that can provide efficient upper extremity training programs for patients and their caregivers.

2.2. The Proposed Methods

2.2.1. Depth/RGB Image Sensor

A home rehabilitation system should be easy to use and low cost; thus, we adopted the Microsoft Kinect RGB-depth (D) sensor to capture salient features. The Kinect depth camera and color camera have some differences, and some distortions occur between RGB color features and depth features extracted from the RGB-D sensors. Therefore, a calibration process to adjust the color and depth features of images is necessary to achieve a coherent image.

Figure 3 indicates the manner in which the angle of the depth camera and the angle of the color camera should be adjusted to calibrate depth information and color information to achieve a coherent image.

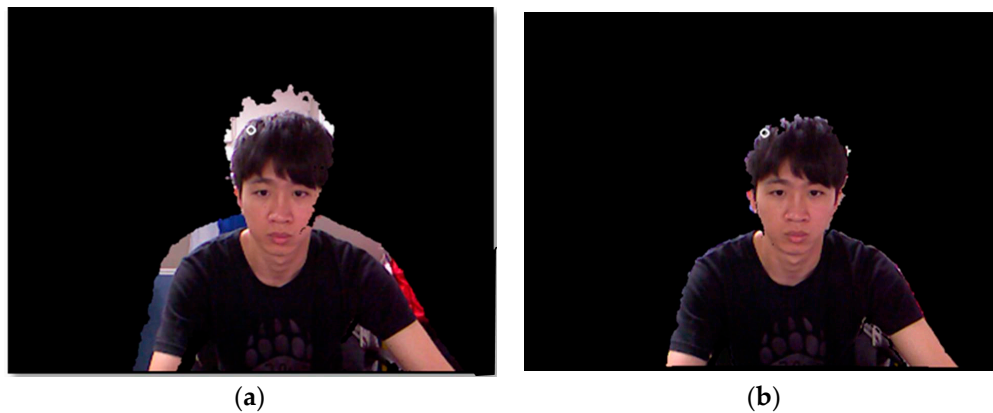


Figure 3. Calibration process to adjust the color and depth features to be coherent: (a) before adjustment and (b) after adjustment.

2.2.2. Skeletonizing

In this study, we detected and skeletonized the patient's upper extremities using OpenNI [24,25] techniques. OpenNI cannot identify the human skeleton immediately at a short distance, and human skeleton detection is difficult in a sitting position. Thus, we extracted the contours of the image, and then determined the human upper body bones from these contours, thereby identifying the joints of the upper body. This study did not directly adopt the skeletal determination of Open NI; rather, we applied the distance transform process for the body contour image in preprocessing [26,27].

- RGB to Gray color transform

If the background and color of the characters are too similar, mistakes can be made in the identification of the human body and errors can be made in the establishment of skeletal joint points. Therefore, the image is converted to gray-scale to remove the background.

- Distance Transform

A distance transform, also known as a distance map or distance field, is a derived representation of a digital image. The map labels each pixel of the image with the distance to the nearest obstacle pixel. The most common type of obstacle pixel is a boundary pixel in a binary image. One technique that may be used in a wide variety of applications is the distance transform or Euclidean distance map [28,29]. The distance transform method is an approximate Euclidean distance map. The distance transform of the image labels each object pixel of the binary image with the distance between that pixel and the nearest background pixel. For example, if the binary image has a pixel it is 1, and if there is no pixel it is 0 (Figure 4a). After distance transform, the farther away from the pixel value of 0, the greater the result that Euclidean distance will obtain. The pixel value of the center will change from 1 to 3 as depicted in Figure 4b. Thus, the distance transform features can highlight the outline of a skeleton frame (Figure 5).

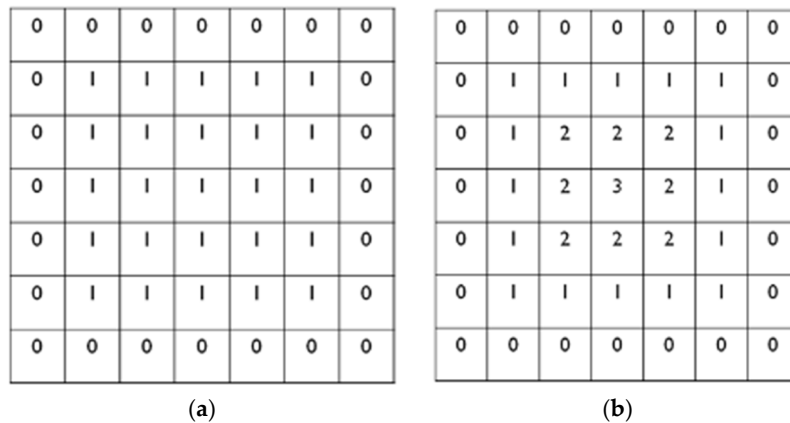


Figure 4. Distance transform process using Euclidean distance: (a) binary Image and (b) distance transform.



Figure 5. The distance transform process to highlight the outline of a skeleton frame: (a) before distance transform and (b) after distance transform.

- Gaussian Blur 7×7

The Gaussian smoothing method is applied to reduce the noisy features. After the Gaussian smoothing process is performed, the edges of the image become blurred, and that reduces pepper and salt noises simultaneously. The Gaussian blur convolution kernel sizes mostly range from 3×3 to 9×9 . This study adopted the convolution kernel size 7×7 , which could obtain the best results for the overall skeleton frame building process in our experiments (Figure 6).

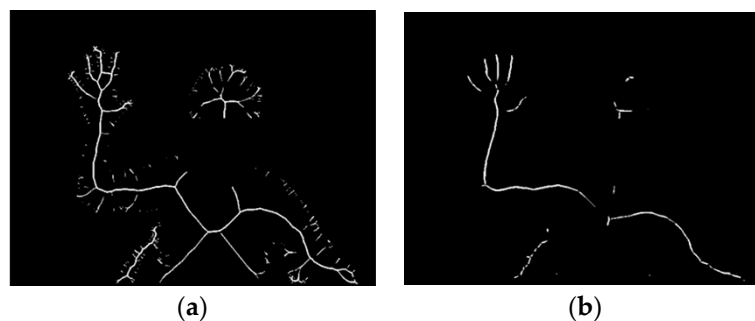


Figure 6. The process to highlight the outline of a skeleton frame: (a) before Gaussian 7×7 smoothing and (b) after Gaussian 7×7 smoothing.

- Convolution Filtering

Next, to strengthen the skeleton features, we used directional convolution filters to enhance the features. This study used the 5×5 -sized convolution kernel because it is faster than convolution kernel 7×7 and clearer than convolution kernel 3×3 to highlight the features of an image of the human body (Figure 7).

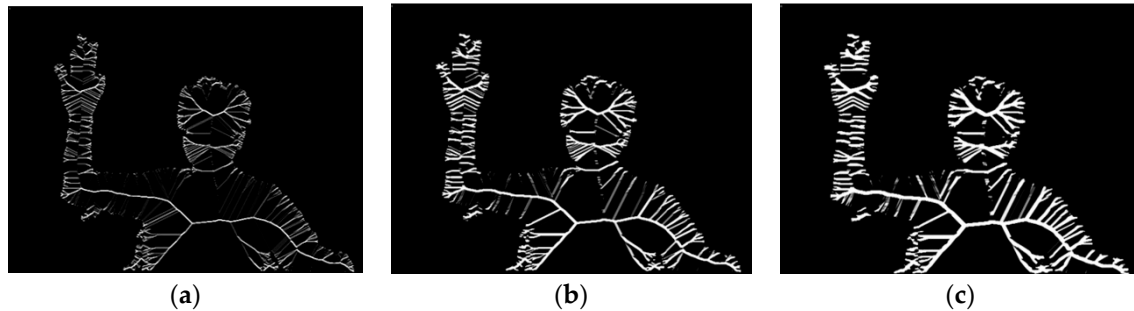


Figure 7. The effects of various convolution kernel processes to highlight the features of an image of the human body: (a) 3×3 convolution kernel, (b) 5×5 convolution kernel, and (c) 7×7 convolution kernel.

Through the convolution filtering process, the areas that required handling were highlighted. We adopted four directions of convolution filtering. Through the four-directional convolution filtering, the contour pixels in each of the four directions are strengthened. The result highlights pixels in each single direction. Take 0 degree terms for an example, as depicted in Figure 8. We compute the maximal values from the four-directional convolution filtering results as the feature values of the corresponding feature points in the following process.

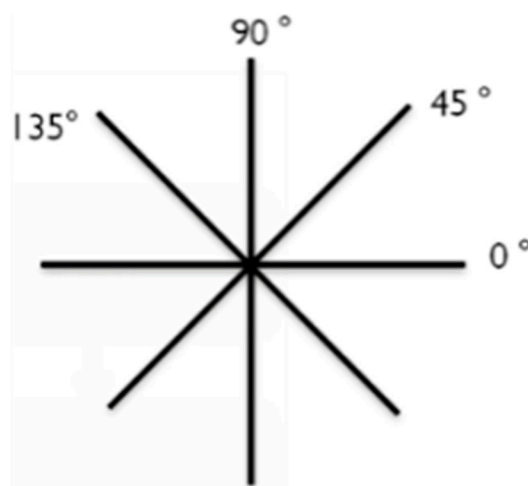


Figure 8. Four directions of convolution filtering.

- Binarization

We used a binarization process as presented in Equation (1) to exclude nonskeletal features. First, we examined the image from the bottom left corner of the origin (0, 0) to the top right corner of the screen (255, 255) to determine if pixels were present. To determine whether a pixel is skeleton or noise, the cut off threshold is set at 6.

$$\text{output Img}(x, y) = \begin{cases} 0, & \text{output Img}(x, y) < 6 \\ 255, & \text{output Img}(x, y) > 6 \end{cases} \quad (1)$$

Following the steps of distance transform, Gaussian blur, and convolution filtering, we could exclude non-skeleton noise and obtain the image of a human skeleton as depicted in Figure 9.

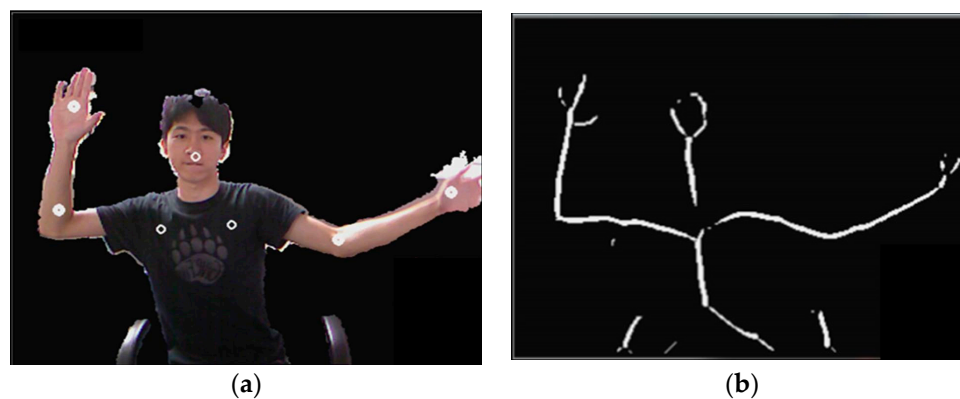


Figure 9. Results before and after skeletonizing: (a) before skeletonizing and (b) after skeletonizing.

2.2.3. Skin Detection

We integrated skin color detection into the skeletonizing process to enable the system to more accurately establish the skeleton.

- RGB to YCbCr (luminance and chroma) Transform

In this study, we applied the YCbCr elliptical skin color model [30] to detect the skin regions. The YCbCr color model decouples the color and intensity features to reduce the lighting effects in color features. Although the YCbCr color model consumes more computational time than the RGB model, it performs computations faster than the HSV (Hue-Saturation-Value) color model [28]. Therefore, using YCbCr color features from human bodies, we can extract skin regions with computational efficiency and more accurately establish the corresponding human skeletons.

- Elliptical Skin Model

We used an elliptical skin model to determine whether skin was present. We created an elliptical skin model where if the pixels outside the elliptical model do not match the skin color, then the pixels inside the ellipse represent the skin color.

- Morphological Close Operation

The RGB image must be converted into YCbCr, and the elliptical skin model is used to separate areas that may be skin color from those that are not. When an area in which the RGB image is likely to be the skin color is separated and reserved, the images may nevertheless include too much noise. In that case, we would conduct morphological close processing on the resultant feature maps.

- Largest Connected Object

Morphological close makes the object itself more connected and also filters out noise. The operation uses the connected object to find the largest connected object in the picture, and then records the location of its center point. The center of the position is the head of the skeleton depicted in Figure 10.

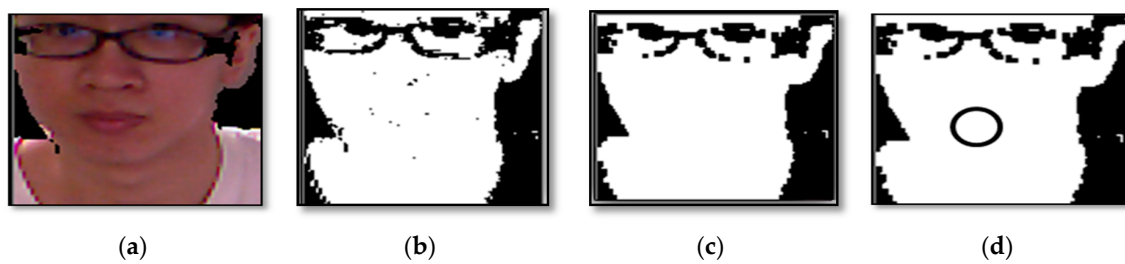


Figure 10. Skin detection process: (a) RGB image, (b) Elliptical skin model, (c) Morphological close, and (d) Largest connected object.

2.2.4. Skeleton Point Establishment

After obtaining the curve-shaped human skeleton from the body contour image and skin color detection, the skeleton is not yet quite consistent with the actual upper extremity structure of the linear human skeleton with joints. Therefore, to accurately detect body movements, it is necessary to establish the accurate position of each joint where movements are produced. In this study, we present a rule-based process for establishing the skeleton point, which is as follows.

Head Skeleton Point Determination

To rapidly determine the skeleton points of the human head, we assume that when the rehabilitation user is sufficiently near the screen center, then the head skeleton can be detected in the central region of the screen, denoted by S_C . We adopted the skeleton feature map and the detected skin region, denoted by SK and SR, respectively, to locate the head skeleton point. Because the user's face should be sufficiently close to the camera, the area of the face's skin region should be sufficiently large. The search process for the head skeleton point is performed through the following steps.

Step 1: Scan the skeleton feature map within the circular region whose center is located at the central coordinate of the input image with a radius of $\max(W/4, H/4)$, where W and H denote the width and height of the image, respectively.

Step 2: If the skeleton feature map can be obtained in Step 1, then we validate the corresponding connected-component area of the skin color region covering the skeleton feature map that is sufficiently large to be a human face—its area should be larger than a given threshold T_s , where the threshold is set at 800 pixels in our experiments. As a result, we can obtain the location of the head skeleton point, denoted by S_H and depicted in Figure 11.

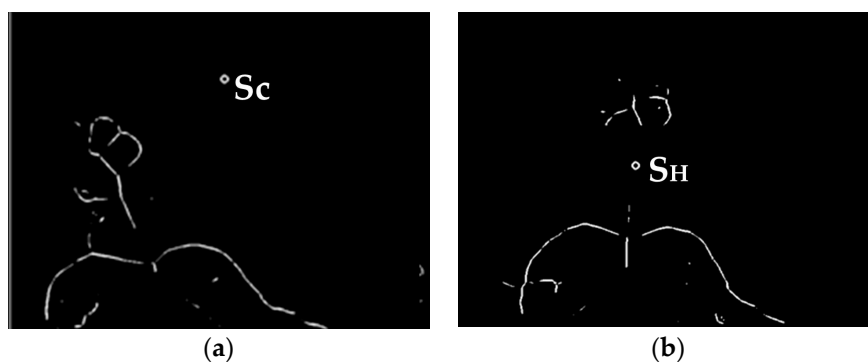


Figure 11. Establishing the skeleton point of the head: (a) The user approaches the preset point, which triggers detection. (b) The skeleton feature map combined with YCbCr elliptical skin color detection establishes the head skeleton point.

Shoulder Skeleton Point Determination

After the skeleton point of the head (S_H) is established, we can determine the pair of shoulder skeleton points according to the position of the head because the shoulders should appear under the two sides of the head:

Step 3: To find the initial shoulder skeleton points, we first search down the rectangular boundaries of the box formed by the head's skin color region to locate the pair of skeleton feature points that are first encountered under the head. Then, we set these two skeleton points as the initial left and right shoulder skeleton points, denoted by S_{SL} and S_{SR} respectively.

Step 4: As depicted in Figure 12a,b, the initial shoulder points are possibly determined on the basis of the clavicle positions under the head's boundaries, but the actual shoulder points should be closer to the lateral ends of the clavicles. Therefore, we set the initial shoulder points S_{SL} and S_{SR} as the centers of the corresponding semicircular regions formed by the movement regions of the two arms.

Step 5: Because the real shoulder points should be at the rotational centers of the shoulder joints, we set the initial shoulder points S_{SL} and S_{SR} as the centers, and set a radius r with the angles θ ranging from 90° to -90° of the semicircular regions of the left and right shoulders. Then, we determined whether skeleton feature points were present over the two semicircular regions through Equation (2).

$$\begin{aligned} S_X &= r \times \cos \theta + S_x \\ S_Y &= -r \times \sin \theta + S_y \end{aligned} \quad (2)$$

where S_x and S_y denote the x and y coordinates of the initial shoulder points (i.e., S_{SL} and S_{SR}), respectively, and S_X and S_Y represent the x and y coordinates of the possible shoulder points we evaluated along the semicircular regions of the left and right shoulders, respectively. Because the real shoulder point will be the rotational center of the shoulder joint, the candidate shoulder point should accord with this characteristic. Thus, if a pair of skeleton points is located at the search regions of the left and right shoulders, denoted by S_L and S_R respectively, then we set these two skeleton points as the actual left and right shoulder points. The search process for the shoulder skeleton points is depicted in Figure 12.

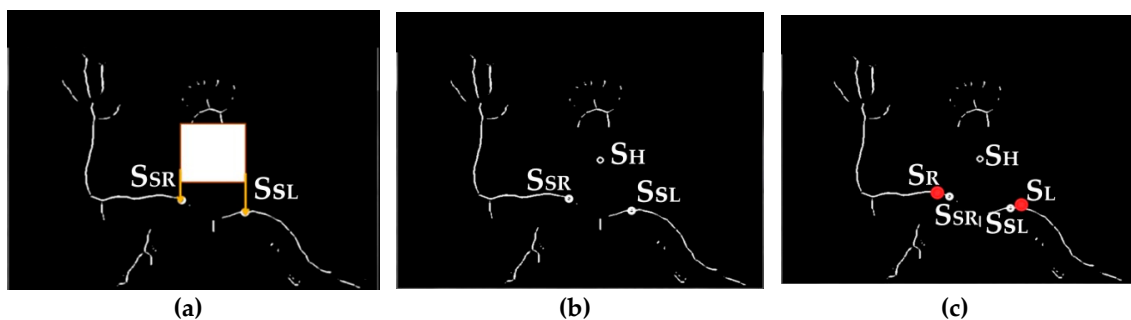


Figure 12. The shoulder point search process: (a) The initial shoulder point is sought under the square region (b) The initial shoulder points (c) The actual shoulder points.

Elbow and Wrist Skeleton Point Determination

After establishing actual shoulder skeleton points (S_L , S_R), we can determine the elbow and wrist skeleton points.

Step 6: We set the obtained shoulder skeleton points (i.e., S_L , S_R) from Step 5 as the centers, then set a radius r with the angles θ ranging from 45° to 315° to search a $3/4$ -circular region to find the arm skeleton points as depicted in Equation (3). The search is completed when the maximum value is found, and all the found points constitute arm skeleton points, as depicted in Figure 13a. We set the

end of points as the pair of wrist skeleton points, denoted by S_{WL} and S_{WR} , respectively, and depicted in Figure 13b.

$$\begin{aligned} S_X &= r \times \cos \theta + S_x(S_L), & S_X &= r \times \cos \theta + S_x(S_R) \\ S_Y &= -r \times \sin \theta + S_x(S_L), & S_Y &= -r \times \sin \theta + S_x(S_R) \end{aligned} \tag{3}$$

where the radius of the search region r is determined to be twice the width of the shoulder because the length of the human arm is typically within twice the shoulder width. It can be determined using the following equation:

$$r = S_x(S_L) - S_x(S_R) \tag{4}$$

Step 7: Next, we can place the left and right elbow skeleton points halfway between the whole arm skeleton points. They are denoted as S_{EL} and S_{ER} , respectively, and are depicted in Figure 13c.

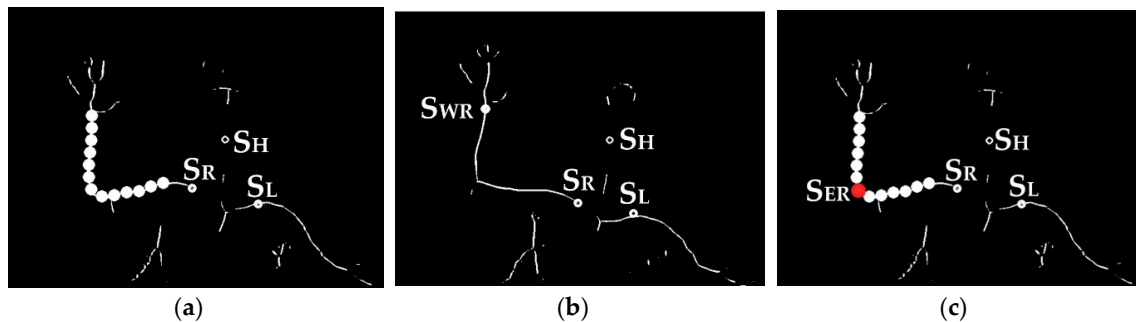


Figure 13. The established skeleton points of the wrist and elbow: (a) All the points we found constitute arm skeleton points. (b) We set the end points at the wrist skeleton points. (c) We set the skeleton point of the elbow halfway between the whole arm skeleton points.

The Overall Skeleton Points Correction Process

If the user does not sit in an upright position or if the user’s body rotates, these situations may cause some errors in the skeleton point setting, as indicated in Figure 14a. This study used depth information to correct the positions of skeletons. When the depth information varies for the right side and left side, we can shrink the radius r for searching on the far side and extend the radius r for searching on the proximal side, as depicted in Figure 14b.

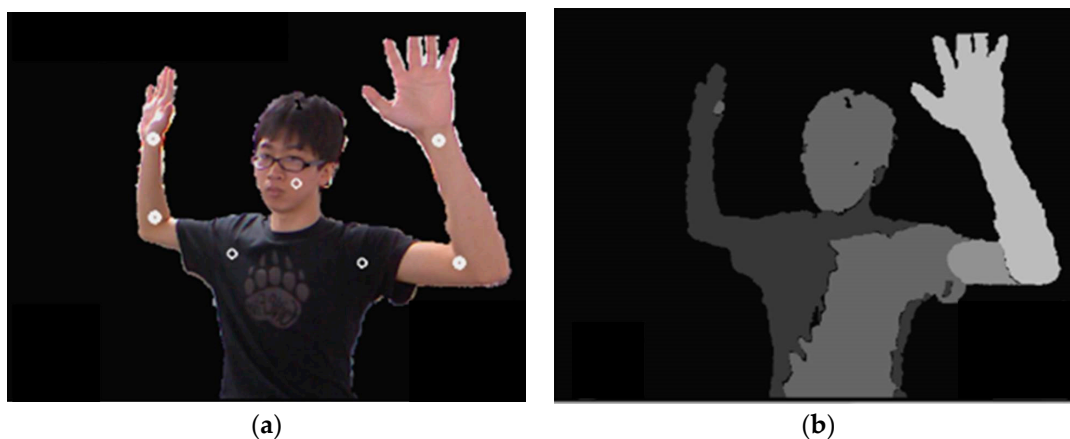


Figure 14. Results before and after arm point correction: (a) before arm point correction and (b) after arm skeleton point correction.

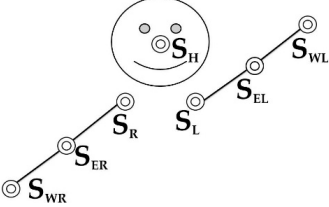

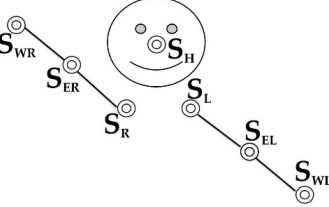

2.2.5. Action Classifier

When the seven $\{S_H, S_L, S_R, S_{EL}, S_{ER}, S_{WL}, S_{WR}\}$ major human skeletal points are established, then the actions can be determined. To determine whether a continuous action is a predefined rehabilitation action, it is necessary to compare the actions of the rehabilitation patient with the rehabilitation actions listed and defined in Table 1. We have preset seven types of rehabilitation movements for specific muscle groups in the system.

Table 1. Seven types of rehabilitation movement.

Position No.	Position of Skeleton Points	Demo of Movement
position 0 (initial position)		
position 1		
position 2		
position 3		
position 4		

Table 1. Cont.

Position No.	Position of Skeleton Points	Demo of Movement
position 5		
position 6		

After the user is trained, the screen will display the movement starting point and end point to guide the user to conduct the appropriate action. However, the time to complete each action may vary by user. Euclidean distance can compare two sequences, but it might not be able to compare two sequential movements of different durations. The traditional Euclidean distance calculation method easily leads to the expansion of distance between two similar but different vectors because the vectors contain action and time and different lengths. The action classifier process is equal to comparing two action sequences of different lengths. Therefore, we must adopt a suitable methodology to handle the two action vectors of different lengths on the time axis, and still achieve the best correspondence. The nonlinear dynamic time warped alignment allows a more intuitive measure to be calculated. The dynamic time warping (DTW) algorithm [31] can solve the problem of various lengths of time without performing additional computations during the training process. The DTW algorithm can achieve the best nonlinear point-to-point correspondence between two motion trajectories, and can still compare the trajectories of various lengths. The DTW algorithm is widely used in audio matching, voice recognition, gesture recognition, and limb motion recognition [32]. Many studies use the DTW algorithm to compare motion trajectories.

In this study, we adopted the DTW algorithm to compare actions and determine the similarity between the defined and the user-performed action, and to avoid the influence of various times spent on similar rehabilitation actions on the action determination. The DTW algorithm identified the features of continuous action sequences, and tolerated the effect of deviation of two vectors of different lengths on the time series in the movement comparison. On the basis of the dynamic programming optimization results, correlations between the two nonlinear trajectories were determined to be the corresponding point-to-point matching relations. According to the optimization results of the dynamic programming, the best point-to-point corresponding relationship between the two nonlinear trajectories was found, and different movement trajectories with different time lengths for the same action could be accurately identified.

3. Results

The hardware of the upper extremity rehabilitation system comprised a personal computer with the Intel Core2 CPU and a Kinect camera. The software comprised the WIN7 operating system, a QT2.4.1 platform, and OpenNI, as presented in Table 2.

Table 2. Hardware and software of the proposed system.

Hardware	Software
CPU: Intel Core(TM)2 Quad 2.33 GHz 2.34 GHz	OS: WIN7
RAM: 4 GB	Platform: QT 2.4.1
Depth camera: Kinect	Library: QT 4.7.4, OpenCV-2.4.3
RGB camera: LogitechC920	OpenNI1.5.2.23(only for RGB image and depth information)
	/

Using Kinect depth information, we performed experiments at the resolutions of 640×480 and 320×240 as displayed in Table 3. To improve execution speed, we adopted the resolution of 320×240 for further examination.

Table 3. Comparison of system execution speed at various resolutions.

Dpi	Release	
	ms	fps
640×480	27	37
320×240	8	125

We examined 20 samples, and each of them had at least 500 frames. The total number of frames was 11,134.

We used skin color detection to determine the presence of the human body and to orient the head. The skin color detection process yielded an average execution speed of 0.26 ms, equivalent to 3889 frames per second (FPS) in a release model. The results from testing all of the system processing functions, including skin color detection and skeletonizing, indicated that each frame only required 8.02 ms, equivalent to 125 FPS in a release model; results are presented in Table 4. Results of test samples are depicted in Figures 15–17.

Table 4. Experimental data of skin color detection speed and system execution speed.

Test Sequence	No. of Frames	Skin Color Detection Speed		System Execution Speed	
		Release Model		Release Model	
		ms	fps	ms	fps
Test Sample 1	618	0.27	3704	8.46	118
Test Sample 2	555	0.25	4000	8.29	121
Test Sample 3	554	0.26	3846	8.04	124
Test Sample 4	509	0.26	3846	8.14	123
Test Sample 5	541	0.25	4000	8.09	124
Test Sample 6	547	0.27	3704	8.20	122
Test Sample 7	539	0.25	4000	8.02	125
Test Sample 8	640	0.26	3846	7.97	125
Test Sample 9	612	0.26	3846	8.08	124
Test Sample 10	540	0.26	3846	7.93	126
Test Sample 11	629	0.25	4000	7.86	127
Test Sample 12	534	0.27	3704	7.87	127
Test Sample 13	517	0.28	3571	8.58	117
Test Sample 14	517	0.25	4000	7.65	131
Test Sample 15	604	0.25	4000	7.94	126
Test Sample 16	528	0.25	4000	8.07	124
Test Sample 17	535	0.25	4000	7.67	130
Test Sample 18	553	0.27	3704	7.90	127
Test Sample 19	524	0.24	4167	7.75	129
Test Sample 20	538	0.25	4000	7.90	127
average	557	0.26	3889	8.02	125



Figure 15. Results of test samples 1–7.



Figure 16. Results of test samples 8–14.



Figure 17. Results of test samples 15–20.

Because the action identification system proposed in this study is aimed at assisting patients who must complete rehabilitation actions, calculation of its accuracy is based on whether the designated user has performed an action correctly or not; a correct performance receives an OK system response, whereas an improperly performed action receives a NO system response.

In addition to quantitatively evaluating the performance of the proposed action identification system, this study evaluated the identification accuracy of rehabilitation actions. Table 5 displays the quantitative identification accuracy data for rehabilitation actions performed using the proposed system. The table data indicate that the proposed system can achieve high accuracy in identifying rehabilitation actions. The average action identification accuracy was 98.1%. Results of rehabilitation action identification are depicted in Figure 18. Accordingly, the high accuracy of the system's identification of rehabilitation actions enables it to effectively assist patients who perform rehabilitation actions of the upper extremity at home. The computational speed of the proposed system can reach

125 FPS, which equates to a processing time per frame of less than 8 ms. This low computation cost ensures that the proposed system can effectively satisfy the demands of real-time processing. Such computational efficiency allows for extensibility with respect to future system developments that account for complex ambient environments or implementation in embedded and pervasive systems.



Figure 18. The results of rehabilitation action identification: position 0 to 6.

Table 5. Accuracy of identification of rehabilitation actions.

Test Sequence	No. of Actual Determined Actions	No. of Correctly Determined Actions	Accuracy Rate of Identifying Rehabilitation Action
position 1	151	149	98.7%
position 2	150	146	97.3%
position 3	156	155	99.4%
position 4	190	184	96.8%
position 5	120	119	99.2%
position 6	186	182	97.8%
Total	953	935	98.1%

4. Discussion

In this study, we replaced dual cameras with a Kinect depth camera. Compared with the study using dual cameras [33], we spent more time to detect skin color—approximately 12 ms. However, this disadvantage can be overcome and palm position can be determined if the user wears long sleeves that reveal only the palm. In addition, the rapid processing speed for the upper extremity tracking ensures the accuracy of the proposed system because we can make a direct judgment of each position. We also check the key points of the skeleton to ensure movements are correctly enacted.

Compared with the study that used HSV for skin color detection [27], we selected to use the YCbCr and elliptical skin models for skin color detection to hasten the skin color detection process. The skeletonizing process of the proposed system optimizes skeletonizing and achieves a high degree of accuracy. It can perform well at a high dpi (640×480) as well as at a low dpi (320×240), and achieves a high processing speed.

The hand-pair gesture presented by Patlolla et al. [33] used two skeleton points (palms), achieved an accuracy of 93%, and operated nearly in real time. The body gesture method proposed by Gonzalez-Sanchez et al. [27], used three skeleton points and achieved 32 FPS with an accuracy of 98%. The method proposed in this study used seven skeleton points, and achieved 37 FPS at 640×480 dpi with an accuracy of 98%; at the lower dpi of 320×240 , the speed was 8 ms, equivalent to 125 FPS. The results of comparison with two studies are depicted in Table 6.

Table 6. Comparison with two studies.

Title	dpi	No. of Skeleton Points	ms	fps	Accuracy Rate
Real-time hand-pair gesture recognition using a stereo webcam [33]	640×480	2	40	25	93%
Real-time body gesture recognition using depth camera [27]	640×480	3	32	32	98%
proposed	640×480	7	27	37	98%
proposed	320×240	7	8	125	98%

5. Conclusions

This study proposed an action identification system for daily home upper extremity rehabilitation. In this section, we will discuss the key contributions of this study. First, in Skeleton Point Establishment phase, we set up those skeleton points not only according to the changes of image features, but also consider the real movements of joints and muscles. Hence, we set up and correct those skeleton points following the principles of anatomy of actions and principles of muscles testing, and those principles also be applied to set rehabilitation action programs. We have preset seven kinds of rehabilitation movements for specific muscle groups in the system to provide a guide for users to perform the actions. The selection of these rehabilitation actions was based on their importance for rehabilitating the ability to perform daily living activities that require the use of the upper extremity such as dressing oneself or reaching for things. Each rehabilitation action program corresponds to training of specific

muscle groups such as the biceps, triceps, or deltoid muscles. Second, the hardware used in the upper extremity rehabilitation system comprises a personal computer with a Kinect depth camera that can be set up similarly to a study desk or computer table at home; thus, the system provides a familiar and convenient environment for rehabilitation users, and the patient can perform the rehabilitation routine even in a limited space without extraneous equipment. Third, patients who cannot stand for a long periods of time, such as stroke, hemiplegia, muscular dystrophy, advanced age, etc., can perform rehabilitation actions in a sitting position, which reduces energy expenditure and enables the patient to focus on the motion control training of the upper extremities without worrying about falling. Fourth, the execution speed of the proposed system can reach 8 ms at a resolution of 320×240 , handle a frame rate equivalent to 125 FPS, and achieve 98% accuracy when identifying rehabilitation actions. Fifth, the proposed upper extremity rehabilitation system operates in real time, achieves efficient vision-based action identification, and consists of low-cost hardware and software. In light of these enumerated benefits, we contend that this system can be effectively used by rehabilitation patients to perform daily exercises, and can reduce the burden of transportation and the overall cost of rehabilitation.

The current limitations of the proposed system are: (1) We combined the skin color detection into skeletonizing process in order to allow the system to be more accurate in establishing the skeleton. That may cause some errors in skeletonizing process if user wears long-sleeved clothes; (2) the proposed system has to establish six rehabilitation actions in the training programs for specific joints such as shoulders, elbows and for specific muscles, such as biceps, triceps, deltoid, that might insufficient for a totally training program. We need to consider how the motion of each plane and axis of the human body translates into 3D or 2D pose estimation in order to develop a more comprehensive and more accurate action rehabilitation system.

In our further studies, the proposed rehabilitation system can also be improved and extended to fit the real human body skeleton to measure the range of motion of human action by image instead of manual of profession in the future based on the machine learning techniques. If set an accurate and real human body skeleton map in pose estimation or action identification system is possible, then we can know not only the pose changed but also the amount of movements and the maximal ranges of active motions. According to this base, a vision-based movement analysis system can be built in our future study. The system can analyze the image sequences of a patient who are performing given activities, then analytical results will determine whether if a patient uses a compensatory action or an error movement pattern violate biomechanical principles.

Author Contributions: Y.-L.C. and P.L. have investigated the ideas, designed the system architecture, algorithm and methodology of the proposed upper extremity rehabilitation system, and wrote the manuscript; C.-H.L. conceived of the presented ideas, implemented the proposed system, and wrote the manuscript with support from Y.-L.C.; C.-W.Y. and Y.-W.K. conducted the experiments, analyzed the experimental data, and provided the analytical results; All authors discussed the results and contributed to the final manuscript.

Funding: This research was funded by Ministry of Science and Technology of Taiwan under the grant numbers MOST-106-2628-E-027-001-MY3 and MOST-106-2218-E-027-002.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kvedar, J.; Coye, M.J.; Everett, W. Connected health: A review of technologies and strategies to improve patient care with telemedicine and telehealth. *Health Aff.* **2014**, *33*, 194–199. [[CrossRef](#)] [[PubMed](#)]
2. Lindberg, B.; Nilsson, C.; Zotterman, D.; Soderberg, S.; Skar, L. Using Information and Communication Technology in Home Care for Communication between Patients, Family Members, and Healthcare Professionals: A Systematic Review. *Int. J. Telemed. Appl.* **2013**, *2013*, 461829. [[CrossRef](#)] [[PubMed](#)]
3. Bianciardi Valassina, M.F.; Bella, S.; Murgia, F.; Carestia, A.; Prosseda, E. Telemedicine in pediatric wound care. *Clin. Ther.* **2016**, *167*, e21–e23.

4. Gattu, R.; Teshome, G.; Lichenstein, R. Telemedicine Applications for the Pediatric Emergency Medicine: A Review of the Current Literature. *Pediatr. Emerg. Care* **2016**, *32*, 123–130. [[CrossRef](#)] [[PubMed](#)]
5. Burke, B.L., Jr.; Hall, R.W. Telemedicine: Pediatric Applications. *Pediatrics* **2015**, *136*, e293–e308. [[CrossRef](#)] [[PubMed](#)]
6. Grabowski, D.C.; O'Malley, A.J. Use of telemedicine can reduce hospitalizations of nursing home residents and generate savings for medicare. *Health Aff.* **2014**, *33*, 244–250. [[CrossRef](#)] [[PubMed](#)]
7. Isetta, V.; Lopez-Agustina, C.; Lopez-Bernal, E.; Amat, M.; Vila, M.; Valls, C.; Navajas, D.; Farre, R. Cost-effectiveness of a new internet-based monitoring tool for neonatal post-discharge home care. *J. Med. Internet Res.* **2013**, *15*, e38. [[CrossRef](#)] [[PubMed](#)]
8. Henderson, C.; Knapp, M.; Fernandez, J.L.; Beecham, J.; Hirani, S.P.; Cartwright, M.; Rixon, L.; Beynon, M.; Rogers, A.; Bower, P.; et al. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): Nested economic evaluation in a pragmatic, cluster randomised controlled trial. *BMJ* **2013**, *346*, f1035. [[CrossRef](#)] [[PubMed](#)]
9. Patel, S.; Park, H.; Bonato, P.; Chan, L.; Rodgers, M. A review of wearable sensors and systems with application in rehabilitation. *J. Neuroeng. Rehabil.* **2012**, *9*, 21. [[CrossRef](#)] [[PubMed](#)]
10. DeLisa, J.A.; Gans, B.M.; Walsh, N.E. *Physical Medicine and Rehabilitation: Principles and Practice*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2005; Volume 1.
11. Cameron, M.H.; Monroe, L. *Physical Rehabilitation for the Physical Therapist Assistant*; Elsevier: Amsterdam, The Netherlands, 2014.
12. Taylor, R.S.; Watt, A.; Dalal, H.M.; Evans, P.H.; Campbell, J.L.; Read, K.L.; Mourant, A.J.; Wingham, J.; Thompson, D.R.; Pereira Gray, D.J. Home-based cardiac rehabilitation versus hospital-based rehabilitation: A cost effectiveness analysis. *Int. J. Cardiol.* **2007**, *119*, 196–201. [[CrossRef](#)] [[PubMed](#)]
13. Lange, B.; Chang, C.Y.; Suma, E.; Newman, B.; Rizzo, A.S.; Bolas, M. Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 1831–1834.
14. Jorgensen, M.G. Assessment of postural balance in community-dwelling older adults—Methodological aspects and effects of biofeedback-based Nintendo Wii training. *Dan. Med. J.* **2014**, *61*, B4775. [[PubMed](#)]
15. Bartlett, H.L.; Ting, L.H.; Bingham, J.T. Accuracy of force and center of pressure measures of the Wii Balance Board. *Gait Posture* **2014**, *39*, 224–228. [[CrossRef](#)] [[PubMed](#)]
16. Clark, R.A.; Bryant, A.L.; Pua, Y.; McCrory, P.; Bennell, K.; Hunt, M. Validity and reliability of the Nintendo Wii Balance Board for assessment of standing balance. *Gait Posture* **2010**, *31*, 307–310. [[CrossRef](#)] [[PubMed](#)]
17. Seamon, B.; DeFranco, M.; Thigpen, M. Use of the Xbox Kinect virtual gaming system to improve gait, postural control and cognitive awareness in an individual with Progressive Supranuclear Palsy. *Disabil. Rehabil.* **2016**, *39*, 721–726. [[CrossRef](#)] [[PubMed](#)]
18. Ding, M.; Fan, G. Articulated and generalized gaussian kernel correlation for human pose estimation. *IEEE Trans. Image Process.* **2016**, *25*, 776–789. [[CrossRef](#)] [[PubMed](#)]
19. Ding, M.; Fan, G. Articulated gaussian kernel correlation for human pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 57–64.
20. Ding, M.; Fan, G. Generalized Sum of Gaussians for Real-Time Human Pose Tracking from a Single Depth Sensor. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 5–9 January 2015; pp. 47–54.
21. Ye, M.; Wang, X.; Yang, R.; Ren, L.; Pollefeys, M. Accurate 3d pose estimation from a single depth image. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 731–738.
22. Baak, A.; Müller, M.; Bharaj, G.; Seidel, H.P.; Theobalt, C. A data-driven approach for real-time full body pose reconstruction from a depth camera. In *Consumer Depth Cameras for Computer Vision*; Springer: London, UK, 2013; pp. 71–98.
23. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu Hawaii, 21–26 July 2017; pp. 7291–7299.

24. PCL/OpenNI Tutorial 1: Installing and Testing. Available online: http://robotica.unileon.es/index.php/PCL/OpenNI_tutorial_1:_Installing_and_testing (accessed on 17 July 2018).
25. Falahati, S. *OpenNI Cookbook*; Packt Publishing Ltd.: Birmingham, UK, 2013.
26. Hackenberg, G.; McCall, R.; Broll, W. Lightweight palm and finger tracking for real-time 3D gesture control. In Proceedings of the IEEE Virtual Reality Conference, Singapore, 19–23 March 2011; pp. 19–26.
27. Gonzalez-Sanchez, T.; Puig, D. Real-time body gesture recognition using depth camera. *Electron. Lett.* **2011**, *47*, 697–698. [[CrossRef](#)]
28. Rosenfeld, A.; Pfaltz, J.L. Sequential operations in digital picture processing. *J. ACM* **1966**, *13*, 471–494.
29. John, C.R. *The Image Processing Handbook*, 6th ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 654–659. ISBN 9781439840634.
30. Kakumanu, P.; Makrogiannis, S.; Bourbakis, N. A survey of skin-color modeling and detection methods. *Pattern Recognit.* **2007**, *40*, 1106–1122. [[CrossRef](#)]
31. Sempena, S.; Maulidevi, N.U.; Aryan, P.R. Human action recognition using dynamic time warping. In Proceedings of the IEEE International Conference on Electrical Engineering and Informatics (ICEEI), Bandung, Indonesia, 17–19 July 2011; pp. 1–5.
32. Muscillo, R.; Schmid, M.; Conforto, S.; D'Alessio, T. Early recognition of upper limb motor tasks through accelerometer: Real-time implementation of a DTW-based algorithm. *Comput. Biol. Med.* **2011**, *41*, 164–172. [[CrossRef](#)] [[PubMed](#)]
33. Patlolla, C.; Sidharth, M.; Nasser, K. Real-time hand-pair gesture recognition using a stereo webcam. In Proceedings of the IEEE International Conference on Emerging Signal Processing Applications (ESPA), Las Vegas, NV, USA, 12–14 January 2012; pp. 135–138.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).