Computer Aided Rehabilitation for Patients with Rheumatoid Arthritis

Vangelis Metsis*, Pat Jangyodsuk*, Vassilis Athitsos*, Maura Iversen†, Fillia Makedon*
*Department of Computer Science and Engineering
University of Texas at Arlington, Arlington, Texas 76019
Email: vmetsis@uta.edu, pat.jangyodsuk@mavs.uta.edu, athitsos@uta.edu, makedon@uta.edu
†Department of Physical Therapy
Northeastern University, Boston, MA 02115
Email: M.Iversen@neu.edu

***Abstract*—Rheumatoid Arthritis (RA) is a chronic disease that leads to inflammation of joints and the surrounding tissues and is a major cause of reduced quality of life and disability. RA can also cause major organ damage and is an independent risk factor for cardiovascular disease. Physical Therapy (PT) and Physical Activity (PA) have been shown to mitigate the effects of the disease, however, lack of motivation and low adherence of the patients reduce the benefits of PT and PA. In this paper, we present a cyberphysical system intended to preserve the functional Range-Of-Motion (ROM) and cardiovascular health in persons with RA, by promoting their physical activity levels and enhancing their physical therapy routines. The system uses game-based activities to increase user motivation and vision-based motion tracking to ensure patient compliance with the prescribed physical therapy routines and activity levels.**

I. Introduction

Rheumatoid Arthritis (RA) is a chronic systemic inflammatory disease typically involving joints on both sides of the body (hands, wrists, feet, knees). The prevalence of RA is 1%, with women affected three to five times as often as men [1]. RA pathology often leads to articular cartilage destruction and may result in joint ankylosis (stiffening). RA can also cause major organ damage; it produces diffuse inflammation in the lungs, the membrane around the heart (pericardium) and is an independent risk factor for cardiovascular disease [2].

It is estimated that RA reduces lifespan by 5 to 10 years and significantly increases morbidity [3]. Work disability among people with RA is markedly higher than in the general population resulting in huge loses for employees & employers; two-thirds of people with RA lose an average of 39 working days each year [4]. In the US, RA-associated costs translate to an annual cost of $19.3 billion. Adding the intangible costs of quality-of-life deterioration ($10.3 billion) and premature mortality ($9.6 billion), total annual societal costs of RA (direct, indirect, and intangible) increases to $39.2 billion [5].

Preserving functional range of motion (ROM) and enhancing cardiovascular health in persons with RA is a primary goal of physical therapy (PT) and is essential to ensuring maximal independence and quality of life. Physical therapy (PT) involves prescription of appropriate exercises to maximize joint ROM, muscle force production and enhance aerobic conditioning. Patient education in self-management techniques is a key component of physical therapy and essential to promoting Physical Activity (PA) and reducing deleterious effects of physical inactivity. Therefore, physical therapists address both exercise and physical activity.

The positive effects of PT are widely acknowledged in the literature (e.g., [6], [7]). Persons with RA who exercise regularly show improvements in muscle strength, ROM, pain, physical function, blood pressure, blood sugar, and aerobic capacity [8]. Further, exercise is associated with improvements in disease activity, bone mineral density and reduced mortality [9]. However, a major obstacle and key motivation for this project is that long-term engagement in exercise among patients with chronic conditions, such as RA, is poor and does not exceed 50% when patients are not supervised [10].

Systems that measure the three-dimensional (3-D) body motion are of great importance, because they provide clinicians with early and quantitative evidence for improved clinical decision making. However, these systems are not routinely integrated into daily clinical interactions because they exact high costs, require the subject to wear obtrusive body markers and require the engagement of the user. As a result, the vast majority of clinical movement analysis is conducted through manual and unreliable direct observation, and only at designated times. The challenges listed below form the basis upon which we build the hypotheses for the proposed work.

- How to develop affordable yet accurate systems to measure body motion.
- How to engage the patients in the rehabilitation process without the presence of physical therapy experts.
- How to adapt any RA physical therapy system to the specific needs of the patient.
- How to monitor and evaluate the exercises that are clinically useful to patients with RA.

In this paper, we present our preliminary results in our effort to create RPLAY, a cyberphysical system which, through the tight collaboration between human and computer, aspires to enhance traditional physical therapy. The system uses game-like activities to motivate and engage the patients in performing the prescribed routines and vision-based motion tracking and recognition to monitor the physical activity of the patient to ensure compliance.

II. Related Work

Lately, gaming has attracted considerable attention due to the potential to provide a promising alternative or enhancement to the traditional rehabilitation therapies e.g., [11], [12]. In many cases, traditional rehabilitation exercises do not yield the expected results due to poor patient adherence with the prescribed routines [10]. People with kinetic disabilities report that traditional rehabilitation tasks can be mundane and boring, due to their repetitive nature. In addition, the lack of direct patient feedback regarding their progress diminishes their motivation to continue. Traditional rehabilitation is also often too hard to do during periods of RA flare-ups, leading to inactivity even when the pain phase is gone. Finally, lack of computational sensing and measurement in traditional therapy may result in errors when interpreting evaluation data.

Virtual reality (VR) shows great promise in creating systematic human testing and treatment environments where virtual representations of real environments can be precisely controlled and guided according to therapy needs. VR provides live feedback to a person doing PT and can act as a motivator for many situations. Furthermore, the gaming factor helps the subject forget about their problem and surroundings and focus directly on the task. Previous studies [13] demonstrate increased motivation in adults when using a virtual environment integrated with gaming.

Although the traditional conception of games is aimed at entertainment, for example driving games or first-person-shooters, lately, the markets have shifted attention towards commercial console games targeted at the "keep-fit" segment. Such games are called "exergames" [14]. In these games, users usually see a simulated virtual representation of themselves (Avatar) or part of themselves (tracked limbs) and they are asked to perform a task by moving their body. Such games are intuitive and suitable for uses with no previous gaming experience. However, commercial games are not designed to meet the needs of patients with disabilities and give feedback to the therapists.

Although such VR systems have been experimentally tested in many situations where physical therapy is required, for example stroke rehabilitation, or cerebral palsy patients [15], [16], they have not found their way towards every day, at-home use. A major prohibiting factor is the large cost of most of the existing VR systems [17]. In addition, they are cumbersome to use: most existing systems require users to carry electromagnetic sensors or attach special markers on their body for tracking purposes, limiting their usability and user friendliness. More simple, camera-based systems, e.g., [18], have either smaller user requirements or no special equipment, e.g. colored gloves. However, they are not accurate enough for the needs of RA rehabilitation.

Currently, most video capture systems track only single plane movement. However in RA, tracking of the exact Range-Of-Motion (ROM) is often necessary to assess disability, improvement or deterioration over time. Capturing 3D human motion from video, without using artificial markers, remains a challenge. Numerous approaches have been proposed, but the accuracy of the current state of the art remains far behind the accuracy attained using special-purpose motion capture systems such as VICON [1].

To address this problem, a common approach is to estimate the pose from multiple calibrated static cameras rather than from a single camera [19], [20], [21], [22]. One disadvantage of this approach is that camera calibration can be a time-consuming process and must precede data collection. Furthermore, every time a camera changes position, whether by plan or by accident, recalibration is required. Consequently, the capturing space is limited, and confined to a studio where the cameras have been set up.

In an at-home setting, where the system must be easy to set up and use by non-technical users, calibrated multi-camera systems are not an option. Hasler et al. [23] describe a video-based motion capture method from moving cameras that are calibrated on the fly, thus matching the setup that we aim to create. However, the method requires a 3D body scan of the person, something which is also impractical in our proposed setting. Many approaches for articulated pose estimation study the tracking version of the problem. The goal here is, given a set of estimates for the previous frame, to update those estimates given the observations in the current frame [24], [20], [21], [22], [25]. However, such methods do not address the challenging problem of how to initialize pose estimates in the first frame, and how to recover from errors. A key goal of our system is initialize and recover from errors fully automatically.

Estimating articulated pose from a single image is an approach that can, in principle, lead to automatic tracker initialization and recovery. Such methods have been proposed in the literature, e.g., [26], [27]. However, such methods suffer from significantly lower accuracy than tracking methods, which is to be expected as searching over the entire space of poses is more difficult than searching over poses similar to the pose in the previous frame. While the above-mentioned methods constitute significant theoretical contributions to the state of the art, they have demonstrated levels of accuracy that are far from sufficient for use in real-world applications. In contrast, our goal is to have a real-world system that actually works robustly and produces useful and meaningful measurements.

Using the Kinect sensor has led to reasonable accuracy to allow deployment as part of the Xbox, which is a mass-market consumer product [28]. A method for pose estimation using Kinect is described in [29], where a randomized decision forest is built from a lot of training image pairs under a fully-supervised training (depth image and body part image). The previous method, as well as the Microsoft SDK in general, are proprietary and not open source, and thus they do not lend themselves to domain-specific improvements and customizations by others. In contrast, the code developed for this project will be open source and publicly available online.

[1] http://www.vicon.com/

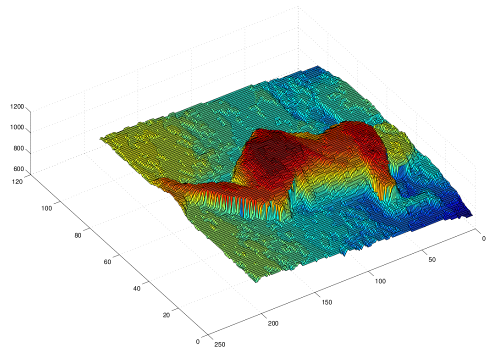

Fig. 1. A heat-map visualization of the depth sensing input of a person lying in bed as obtained by the Kinect depth sensor.

Fig. 2. An example of Kinect based apple picking game where the user tries to catch the apples from a tree by moving their hands in the 3-D space. On the left image we can see a user playing the game and on the right image we see a closer view of screen during the game. The left part of the screen shows the skeletal representation of the player - used for debugging purposes.

III. Framework

Two major goals of our RPLAY system are the development of rehabilitation games that are created in accordance to the prescribed exercise routines and the accurate tracking and recognition of the performed motion activities. The rehabilitation games will guide the patients to perform the types of motion that will benefit their condition in a similar way as the traditional exercises. Subsequently, the motion tracking and recognition component will identify which motion activities/exercises have been performed and how.

Video-based tracking is ideal in this case, because it is minimally invasive and can be adapted to track the required motion types. The Kinect sensor can provide 3-dimensional information regarding the tracked subject and does not require stereo calibration. In addition it is relatively inexpensive compared to other existing solutions. For that reason we use it as our video capturing device.

Kinect is a motion sensing input device designed by Microsoft for the Xbox 360 video game console [28]. Kinect outputs 3 different data streams, RGB video stream, depth sensing video stream and audio. The video output frame rate is 30 Hz. The RGB video stream uses 8-bit VGA resolution (640×480 pixels), while the monochrome depth sensing video stream is in VGA resolution (640×480 pixels) with 11-bit depth, which provides 2,048 levels of sensitivity. In our experiments we used only the depth sensing video stream. The depth sensor consists of an infrared laser projector combined with a monochrome CMOS sensor, which captures video data in 3D under any ambient light conditions. The 3D input that we get regarding the subject's body posture is more informative compared to the 2D information that we could get from the RGB video. The value of each pixel in a depth video stream frame is the distance, in millimeters, of the corresponding surface part of the object from the sensor. Figure 1 shows an example of how the depth sensing data obtained by Kinect would look like using a heat-map visualization.

A. Rehabilitation Games

Using rehabilitation games as opposed to the traditional exercise routines introduces a number of advantages: (1) The user gets real-time performance feedback, which increases motivation. (2) By simulating real-world activities, the performed exercises have greater validity, which means that the degree of relevance or similarity that a test or training system has to the "real" world is high. (3) Certain interface modifications contingent on user's impairment are possible. (4) The use of virtual humans can showcase the exercises that the users need to perform and immediate feedback can be given to the user in case of poor adherence. (5) pacing and difficulty level can be adapted to meet the patient's needs.

Since the designed games will be used for rehabilitation purposes, instead of plain entertainment, and the main target group of users will be individuals of a certain age, the game behavior has to be easily parameterizable by the therapists, and games should be intuitive in their use by the patients so that they do not pose additional challenges which may discourage them from engaging.

In our previous work [30], in collaboration with occupational therapists, we have successfully developed a library of touch screen-based games for rehabilitation of children with Cerebral Palsy (CP). Such games measure the reaction speed, the arm control accuracy, the improvement over time and other parameters of importance for patients with CP.

However, touch screen-based games are not suitable for rehabilitation of RA patients, since the purpose of RA exercises is to increase or maintain the range-of-motion of the patients and increase physical activity rather than improve arm/finger control. In this case, games that involve the whole body (arms, torso, and legs) are preferable. To achieve that purpose we are creating a new library of games based on video tracking using the Kinect sensor. One such example is our apple-picking game shown in Figure 2, where the user has to extend his/her arms to certain directions in order to catch the apples on an apple tree by guiding a virtual hand appearing on the screen. The apples can be placed in such positions that force the user to move their hands towards the required directions and with the required speed according to the prescribed exercise routines.

B. The RPLAY System

RPLAY will monitor patients, as they perform a daily rehabilitation routine or play a rehabilitation game. We employ real-time computer vision technologies that capture the 3D motion of the patients, and recognize patients' gestures and activities. On the application level, the system has the following functionalities. The system can determine:

1) What rehabilitation activities or exercises the patient engages in and how much time is spent on each. These data are then reported to the physical therapist, so that he/she can quickly determine if the patient follows the prescribed rehabilitation program and calibrate exercise levels based on observed symptoms during exercise.
2) How well each activity is performed, especially the estimated ROM exhibited in the activity. Such measurements are then summarized, visualized and reported to the therapist who uses this information on correct performance of exercise to prevent possible deleterious effects (e.g., bursitis, joint pain). The ROM improvements over time are then stored into the database.
3) The metabolic equivalents computed from gross body movements while the user performs the game activities using gestures and body motion. This is important feedback that will determine the amount of physical activity performed and whether a patient is meeting CDC recommendations for physical activity to maintain cardiovascular health, given the elevated risk of cardiovascular disease in this population.

C. Motion Tracking and Recognition

While there is a significant body of literature studying the problem of capturing and analyzing 3D human motion, most existing approaches require multiple calibrated cameras or other specialized equipment that do not satisfy the requirements of cost effectiveness and easy set up by non-technical people. On the other hand, methods using a single camera typically are not accurate enough for 3D motion capture. To solve the problem of 3D human motion analysis, we use our recently developed, Dynamic Space-Time Warping (DSTW) algorithm [31]. Its key idea is that inaccurate estimation of a person's 3D motion can be sufficient for accurate recognition of the patient's activity. At the same time, recognizing the specific activity the person is performing provides a powerful additional constraint, that can refine the estimation of the person's 3D motion to a satisfactory of level accuracy.

In DSTW, two modules, a tracking module, and a recognition module, communicate with each other. The tracking module produces estimates of the person's movements in each frame, consisting of an estimated position for each body part in each frame. The recognition module uses the estimates obtained from the tracking module to recognize the kind of motion (e.g., the type of rehabilitation exercise, or the type of gesture that the person is making to control the game) the person is performing. It is unrealistic to expect that the tracking module will produce the correct answer at each frame. The key difference between our Dynamic Space-Time Warping (DSTW) method and competing approaches is that the recognition module does not require the tracking module to be highly accurate. Instead, in DSTW, we make the milder assumption that the tracking module will produce, at each frame, multiple candidate estimates (possibly several tens), and that in almost all frames the correct estimate will be included in those candidates. This level of accuracy is easier to satisfy with existing methods, compared to a requirement that in almost all frames the tracker can provide a single correct estimate.

Since there are many possible answers for how the person is tracked at each frame, there are even more candidate sequences of such answers. A candidate motion sequence is formed by choosing one candidate answer for a frame. Naturally, the number of candidate motion sequences is exponential, in the order of $O(K^T)$, where K is the number of candidate answers per frame and T is the number of frames. DSTW is based on the following observation: as long as the correct answer appears as a candidate answer in almost all frames, one of the exponentially many candidate sequences will actually include these correct answers. If we knew this sequence, it would be easy to recognize what type of activity (e.g., a specific rehabilitation exercise) it represents, as the number of such activities is relatively small. The DSTW algorithm is exactly designed to find, given a specific activity model, the candidate sequence that optimally matches that activity model. Although, as mentioned above, the number of candidate sequences is exponential, DSTW identifies the optimal sequence in polynomial time, and in practice the algorithm is fast enough for real-time recognition [31].

Once the current activity has been identified, the proposed system will measure the range of motion. The specialist will specify in advance, for each activity, the moments in that activity that correspond to motion extrema that need to be measured. The DSTW algorithm will automatically identify those moments, as well as the correct tracking estimate for those moments, and the range of motion can be identified from those tracking estimates.

D. Motion Tracking Using Kinect

Using the DSTW algorithm frees us from the need to design a highly accurate tracking module (which would be unrealistic given the current state of the art). At the same time, a tracking module that works as well as possible, is of great advantage. Microsoft's Kinect SDK implements its own tracking system and provides skeletal information by specifying the location of each of the main joints of the human body in the 3D space. Figure 3 shows an example of the tracking information extracted by Kinect SDK.

However, the accuracy achieved by the tracking module provided by the Kinect sensor is far from perfect and in addition to that, the underlying software is proprietary and closed source. That does not allow us to adapt it to our needs and possibly extend or improve its abilities. To overcome these problems, we have created our own tracking module. Figure

Fig. 3. Example of full body tracking provided by Kinect SDK. The left image shows the RGB color frame of a subject performing a RA exercise. The middle image shows the depth sensing information of the same frame where the pixels corresponding to the tracked subject have been marked in red. The right image shows the skeletal information extracted by Microsoft Kinect SDK.

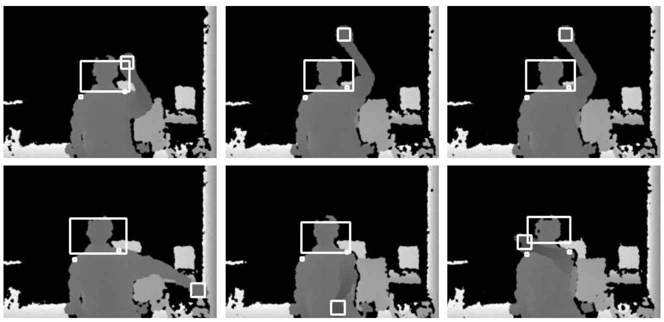

Fig. 4. The images show the qualitative result of body parts detection of our tracking module using the depth sensing information obtained by the Kinect sensor. The tracked body parts (head, shoulders and hands) are surrounded by bounding boxes.

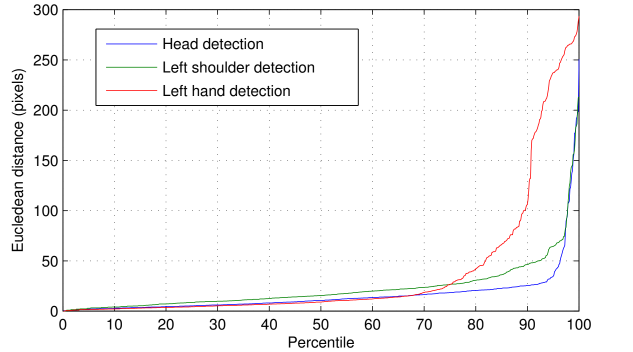

Fig. 5. Quantitative evaluation of body parts detection. The accuracy is measured in terms of Euclidean distance in pixels of predicted position and actual position. The X axis shows the percentage of the test images on which the predicted position has lower or equal distance from the actual position than the value indicated in the Y axis.

4 shows a qualitative result of the tracking achieved by our tracking module using the depth sensing data coming from Kinect. At its current status it can detect and track the head, the shoulders and the hands of the tracked subject. The detected parts are surrounded by bounding boxes.

Our method starts by detecting the head using template search to find the face of a person. Taking the postion of the head into consideration, it detects the position of the body and then it uses that information to improve the originally detected position of the head and subsequently detect the position of the shoulders and the position of the hands. Algorithm 1 describes the high level procedure we followed to detect the head, shoulders and hands. Figure 5 demonstrates our experimental quantitative evaluation of the accuracy of our method in detecting the position of the body parts of interest with respect to their actual location in a dataset of 800 manually annotated images randomly selected from ChaLearn gesture challenge dataset [32]. The graph shows the accuracy of detecting the head, the left shoulder and the left hand (accuracy for right shoulder and hand is similar), in terms of the percentage of images on which the distance of the predicted position is smaller than a number of pixels (value of Y axis) from the actual position as annotated by humans.

IV. Conclusion

In this paper, we presented a cyberphysical system (named RPLAY) we are developing to enhance the rehabilitation of patients with RA by, first, making the physical therapy routines more interesting and engaging for the patients and providing them with real-time feedback of their progress, and second, by providing information to the therapists regarding the patient's adherence to the rehabilitation program. We demonstrated how video games can be used as part of the rehabilitation process and how our motion tracking method can track the human body parts and recognize different types of motion which correspond to prescribed exercises.

Acknowledgment

This work is supported in part by the National Science Foundation under award numbers NSF-1035913, NSF-0923494, NSF-1128296, NSF-1059235, NSF-1055062, NSF-0812601. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] V. Majithia and S. A. Geraci, "Rheumatoid arthritis: diagnosis and management," *The American journal of medicine*, vol. 120, no. 11, pp. 936–939, 2007.

[2] K. J. Warrington, P. D. Kent, R. L. Frye, J. F. Lymp, S. L. Kopecky, J. J. Goronzy, and C. M. Weyand, "Rheumatoid arthritis is an independent risk factor for multi-vessel coronary artery disease: a case control study," *Arthritis research & therapy*, vol. 7, no. 5, p. R984, 2005.

[3] C. Turesson, W. M. OFallon, C. S. Crowson, S. E. Gabriel, and E. L. Matteson, "Extra-articular disease manifestations in rheumatoid arthritis: incidence trends and risk factors over 46 years," *Annals of the rheumatic diseases*, vol. 62, no. 8, pp. 722–727, 2003.

[4] W. Burton, A. Morrison, R. Maclean, and E. Ruderman, "Systematic review of studies of productivity loss due to rheumatoid arthritis," *Occupational Medicine*, vol. 56, no. 1, pp. 18–27, 2006.

[5] H. Birnbaum, C. Pike, R. Kaufman, M. Maynchenko, Y. Kidolezi, and M. Cifaldi, "Societal cost of rheumatoid arthritis patients in the US," *Current Medical Research & Opinion*, vol. 26, no. 1, pp. 77–90, 2009.

[6] A. Baillet, E. Payraud, V. A. Niderprim, M. J. Nissen, B. Allenet, P. Franois, L. Grange, P. Casez, R. Juvin, and P. Gaudin, "A dynamic exercise programme to improve patients disability in rheumatoid arthritis: a prospective randomized controlled trial," *Rheumatology*, vol. 48, no. 4, pp. 410–415, 2009.

Algorithm 1 Head, Shoulder and Hand detection using depth sensing data obtained by the Kinect sensor.

1) Detect Head:
   - Use template search to detect head.
   - Proceed to step 2. After detecting body position, correct initially detected head position with respect to body position.
2) Detect body:
   - Given given center head position, find largest connected component with depth closer to camera with respect to face.
3) Detect each shoulder:
   - Create Gaussian distribution of $P(Shoulder, Head)$ using human annotated images.
   - Find $P(Shoulder|Head)$ for each pixel of current image frame.
   - Refine detected shoulder by moving the expected position to the closest pixel in body.
4) Detect each hand:
   - Get score for every pixel in segmented body image. Use 3D Chamfer distance with respect to the neutral position image (1st frame of the same video).
   - Find the connected component with highest average score (this should be the arm containing the hand).
   - Heuristically find diameter of this connected component, one side should be hand while the other should be elbow or something else. This will give us 2 candidates for the hand position.
   - Use human annotated database of $P(elbow, hand)$ position to determine most probable hand position. Reduce to one candidate by using $min(1 - NN((candidate_1, candidate_2), database), 1 - NN((candidate_2, candidate_1), database))$ ▷ Where NN is Nearest Neighbor search and $database$ is a set of previously annotated images.

[7] M. D. Iversen, A. Finckh, and M. H. Liang, "Exercise prescriptions for the major inflammatory and non-inflammatory arthritides," *Exercise In Rehabilitation Medicine*, pp. 157–179, 2005.

[8] A. P. Cairns and J. G. McVeigh, "A systematic review of the effects of dynamic exercise in rheumatoid arthritis," *Rheumatology international*, vol. 30, no. 2, pp. 147–158, 2009.

[9] A. Hkkinen, T. Sokka, A. Kotaniemi, and P. Hannonen, "A randomized twoyear study of the effects of dynamic strength training on muscle strength, disease activity, functional capacity, and bone mineral density in early rheumatoid arthritis," *Arthritis & Rheumatism*, vol. 44, no. 3, pp. 515–522, 2001.

[10] R. Campbell, M. Evans, M. Tucker, B. Quilty, P. Dieppe, and J. L. Donovan, "Why don't patients do their exercises? understanding non-compliance with physiotherapy in patients with osteoarthritis of the knee," *Journal of epidemiology and community health*, vol. 55, no. 2, pp. 132–138, 2001.

[11] M. J. A. Jannink, G. J. Van Der Wilden, D. W. Navis, G. Visser, J. Gussinklo, and M. Ijzerman, "A low-cost video game applied for training of upper extremity function in children with cerebral palsy: a pilot study," *CyberPsychology & Behavior*, vol. 11, no. 1, pp. 27–32, 2008.

[12] F. Anderson, M. Annett, and W. F. Bischof, "Lean on wii: Physical rehabilitation with virtual reality and wii peripherals," *Stud Health Technol Inform*, vol. 154, pp. 229–34, 2010.

[13] D. Jack, R. Boian, A. S. Merians, M. Tremaine, G. C. Burdea, S. V. Adamovich, M. Recce, and H. Poizner, "Virtual reality-enhanced stroke rehabilitation," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 9, no. 3, pp. 308–318, 2001.

[14] I. Bogost, "The rhetoric of exergaming," in *Digital Arts and Cultures (DAC) Conference (December 2005), IT University Copenhagen*, 2005.

[15] K. Morrow, C. Docan, G. Burdea, and A. Merians, "Low-cost virtual rehabilitation of the hand for patients post-stroke," in *Virtual Rehabilitation, 2006 International Workshop on*. Ieee, 2006, pp. 6–10.

[16] Y. P. Chen, L. J. Kang, T. Y. Chuang, J. L. Doong, S. J. Lee, M. W. Tsai, S. F. Jeng, and W. H. Sung, "Use of virtual reality to improve upper-extremity control in children with cerebral palsy: a single-subject design," *Physical Therapy*, vol. 87, no. 11, pp. 1441–1457, 2007.

[17] A. S. Rizzo and G. J. Kim, "A SWOT analysis of the field of virtual reality rehabilitation and therapy," *Presence: Teleoperators & Virtual Environments*, vol. 14, no. 2, pp. 119–146, 2005.

[18] L. Geurts, V. Vanden Abeele, J. Husson, F. Windey, M. Van Overveldt, J. H. Annema, and S. Desmet, "Digital games for physical therapy: fulfilling the need for calibration and adaptation," in *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*. ACM, 2011, pp. 117–124.

[19] C. Chen, Y. Zhuang, and J. Xiao, "Silhouette representation and matching for 3D pose discrimination-a comparative study," *Image and Vision Computing*, vol. 28, no. 4, pp. 654–667, 2010.

[20] J. Darby, B. Li, and N. Costen, "Tracking human pose with multiple activity models," *Pattern Recognition*, vol. 43, no. 9, pp. 3042–3058, 2010.

[21] J. Gall, B. Rosenhahn, T. Brox, and H. P. Seidel, "Optimization and filtering for human motion capture," *International journal of computer vision*, vol. 87, no. 1, pp. 75–92, 2010.

[22] L. Sigal, A. O. Balan, and M. J. Black, "Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *International Journal of Computer Vision*, vol. 87, no. 1, pp. 4–27, 2010.

[23] N. Hasler, B. Rosenhahn, T. Thormahlen, M. Wand, J. Gall, and H. P. Seidel, "Markerless motion capture with unsynchronized moving cameras," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. Ieee, 2009, pp. 224–231.

[24] A. O. Balan, L. Sigal, M. J. Black, J. E. Davis, and H. W. Haussecker, "Detailed human shape and pose from images," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[25] H. Kjellstrom, D. Kragic, and M. J. Black, "Tracking people interacting with objects," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 747–754.

[26] A. Fossati, M. Dimitrijevic, V. Lepetit, and P. Fua, "From canonical poses to 3D motion capture using a single camera," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 7, pp. 1165–1181, 2010.

[27] C. W. Wang and A. Hunter, "Robust pose recognition of the obscured human body," *International journal of computer vision*, vol. 90, no. 3, pp. 313–330, 2010.

[28] "Kinect-xbox.com," http://www.xbox.com/en-US/kinect.

[29] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *CVPR*, vol. 2, 2011, p. 7.

[30] A. Papangelis, G. Mouchakis, A. Boisselle, D. Kosmopoulos, V. Karkaletsis, and F. Makedon, "A game system for remote rehabilitation of cerebral palsy patients," in *The 5th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA'12)*. ACM, 2012.

[31] J. Alon, V. Athitsos, Q. Yuan, and S. Sclaroff, "A unified framework for gesture recognition and spatiotemporal gesture segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 9, pp. 1685–1699, 2009.

[32] I. Guyon, V. Athitsos, P. Jangyodsuk, B. Hamner, and H. J. Escalante, "ChaLearn gesture challenge: Design and first results," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, 2012, pp. 1–6.