A Survey - Human Movement Tracking and Stroke Rehabilitation

TECHNICAL REPORT: CSM-420
ISSN 1744 - 8050

Huiyu Zhou and Huosheng Hu

8 December 2004

Department of Computer Sciences
University of Essex
United Kingdom

Email: zhou@essex.ac.uk, hhu@essex.ac.uk

Contents

Abstract

This technical report reviews recent progress in human movement tracking systems in general, and patient rehabilitation in particular. Major achievements in previous working systems are summarized. Meanwhile, problems in motion tracking that remain open are highlighted along with possible solutions. Finally, discussion is made regarding challenges which remain and a design specification is proposed for a potential tracking system.

Figure 1. A rehabilitation system at the Massachusetts Institute of Technology (MIT), USA.

1 Introduction

Evidence shows that, in 2001-02, 130,000 people in the UK experienced a stroke [62] and required admission to hospital. More than 75% of these people were elderly, who required locally based multi-disciplinary assessments and appropriate rehabilitative treatments after they were dismissed from hospital [35], [48]. As a consequence, this increased greatly the demand on healthcare services, and expense in the national health service. To enhance the health service, people intend to use intelligently devised equipment to conduct patient rehabilitation in the patient's home rather than in a hospital that may be geographically remote. These systems are expected to reduce the requirement for face-to-face therapy between therapy experts providing vision and audio supports, and patients.

The goal of rehabilitation is to enable a person who has experienced a stroke to regain the highest possible level of independence so that they can be as productive as possible. Since stroke patients often have complex rehabilitation needs, progress and recovery characteristics are unique for each person. Although a majority of functional abilities may be restored soon after a stroke, recovery is an ongoing process. Therefore, home-based rehabilitation systems are expected to have adaptive settings designed to meet the requirements of individuals, automatic operation, an open human-machine interface, rich database for later evaluation, and compactness and portability. In fact, rehabilitation is a dynamic process which uses available facilities to correct any undesired motion behavior in order to reach an expectation (e.g. ideal position).

During the rehabilitation process, the movement of stroke patients needs to be localized and learned so that incorrect movements can be instantly modified or tuned. Therefore, tracking these movements becomes vital and necessary during the course of rehabilitation. This report details a survey of technologies deployed by human movement tracking systems that consistently update the spatiotemporal information of patients. Previous systems (one of them shown in Figure 1) have proved that, to some extent, properly conducted designs are capable of improving the quality of human movement, but many challenges still remain due to complexity and uncertainty in movement. In the following sections, a comprehensive review of this type of systems is provided.

The rest of this report is organized as follows. Section 2 outlines the four main types of technologies used in human movement tracking. Section 3 presents non-vision based human movement tracking systems, which have been commercialized. Marker-based visual tracking systems are introduced in Section 4, and markerless visual systems described in Section 5. Section 6 provides robot-guided tracking system concepts and a description of their application in the rehabilitation procedure. A research proposal based on previous work at the University of Essex, and lit-

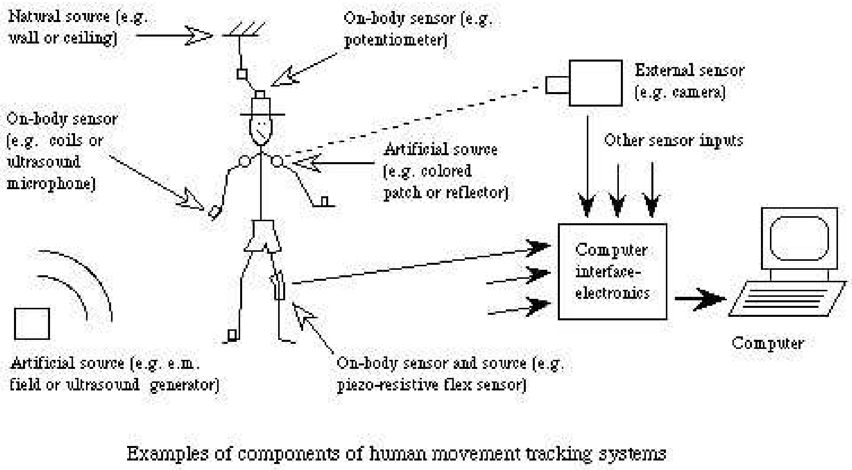


Figure 2. Illustration of a real human movement tracking system (courtesy of Axel Mulder, Simon Fraser University).

erature is provied in Section 8. Finally, conclusions are provided in Section 9.

2 Sensor technologies

Human movement tracking systems generate real-time data that represents measured human movement [80], based on different sensor technologies. For example, Figure 2 illustrates a hybrid human movement tracking system. Retrieving such sensing information allows a system to efficiently describe human movement, e.g. arm motion. However, it is recognized that sensor data encoded with noise or error due to relative movement between the sensor and the objects to which it is attached. It is therefore essential to understand the structure and characteristics of sensors before they are applied to a tracking system. According to sensor location on a human body, tracking systems can be classified as non-vision based, vision based with markers, vision based without markers, and robot assisted systems. These systems are described one at a time in the following sections.

2.1 Non-vision based tracking

In non-vision based systems, sensors are attached to the human body to collect movement information. Their sensors are commonly classified as mechanical, inertia, acoustic, radio or microwave and magnetic sensing. Some of them have a small sensing footprint that they can monitor small amplitudes such as finger or toe movement. Each kind of sensor has advantages and limitations. Limitations include modality-specific, measurement-specific and circumstance-specific limitations that accordingly affect the use of the sensor in different environments [108].

For example, as part of inertia sensors accelerometer sensors (Figure 3) convert linear acceleration, angular acceleration or a combination of both into an output signal [31]. There are three common types of accelerometers: piezoelectric which exploit the piezoelectric effect whereby a naturally occurring quartz crystal is used to produce an electric charge between two terminals; piezoresistive operating by measuring the resistance of a fine wire when it is mechanically deformed by a proof mass [71]; and variable capacitive where the change in capacitance is proportional to acceleration or deceleration [110]. An

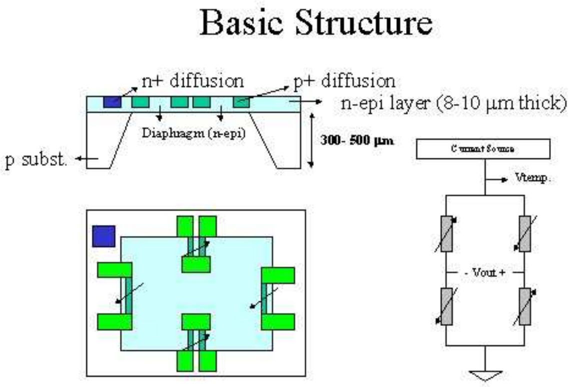


Figure 3. Illustration of a piezoresistive sensor.

example of accelerometers is given in Figure 4. Unfortunately, these sensors demand some computing power, which possibly increases response latency. Furthermore, resolution and signal bandwith are normally limited by the interface circuitry [28].

2.2 Vision based tracking with markers

This is a technique that uses optical sensors, e.g. cameras, to track human movements, which are captured by placing identifiers upon the human body. As human skeleton is a highly articulated structure, twists and rotations make the movement fully three-dimensional. As a consequence, each body part continuously moves in and out of occlusion from the view of the cameras, leading to inconsistent and unreliable tracking of the human body. As a good solution to this situation, marker-based vision systems have attracted the attention of researchers in medical science, sports science and engineering.

One major drawback of using optical sensors and markers, however, is that they are difficult to use to accurately sense joint rotation, leading to the infeasibility of representing a real 3-D model for the sensed objects [102].

Figure 4. Entran's family of miniature accelerometers.

2.3 Vision based tracking without markers

This technique exploits external sensors like cameras to track the movement of the human body. It is motivated by facts addressed in marker based vision systems [1]: (1) Identification of standard bony landmarks can be unreliable. (2) The soft tissue overlying bony landmarks can move, giving rise to noisy data. (3) The marker itself can wobble due to its own inertia. (4) Markers can even come adrift completely.

A camera can be of a resolution of a million pixels. This is one of the main reasons that such an optical sensor s attracted people's attention. However, such vision based techniques require intensive computational power to achieve efficiently and to reduce the latency of data [32]. Moreover, high speed camera's are also required, as conventionally less than sixty frames a second provides an insufficient bandwith for accurate data representation [24].

2.4 Robot assisted tracking

Recently, voluntary repetitive exercises administered with the mechanical assistance of robotic rehabilitators has proven effective in improving arm movement ability in post-stroke populations. During the course of rehabilitation, human movement is reflected by using sensors attached to the body, which consist of electromechanical and electromagnetic sensors. Electromechanical systems prohibit free movements and involve disconnecting sensors from the human body. The electromagnetic approach provides more freedom for human movement, but is seriously affected by directional sensors.

3 Human movement tracking: non-vision based systems

Understanding and interpreting human behavior has attracted attention of therapists and biometric researchers due to its impact on the recovery of patient post disease. So, people need to learn dynamic characteristics about the actions of certain parts of the body, e.g. hand-gestures and gait analysis. Tracking actions is an effective means that consistently and reliably represents human dynamics against time. This purpose can be reached through the use of electromechanical or electromagnetic sensors. This is so-called “non-vision based tracking”. Among the sensors and systems to be introduced below, MT9 based, G-link based and MotionStar systems have wireless properties, indicating that they are not limited in space.

3.1 MT9 based

The MT9 [64] of Xsens Motion Tech, is a digital measurement unit that measures 3-D rate-of-turn, acceleration and earth-magnetic field, as referred to in Figure 5. Combined with the MT9 Software it provides real-time 3-D orientation data in the form of Euler angles and Quaternions, at frequencies up to 512 Hz and with an accuracy better than 1 degree root-mean-square (RMS).

The algorithm of the MT9 system is equivalent to a sensor fusion process where the measures of gravity through accelerometers and magnetic north via magnetometers compensate for increasing errors from the integration of the rate of turn data. Hence, this drift compensation is attitude and heading referenced. In a homogeneous earth-magnetic field, the MT9 system has 0.05 degrees RMS angular resolution; $<$ 1.0 degrees static accuracy; and 3 degrees RMS dynamic accuracy.

Due to its compact size and reliable performance, the MT9 has easily been integrated into the field of biomechanics, robotics, animation, and virtual reality, etc. However, a MT9-based tracker with six MT9 units costs about 16,000 euros.

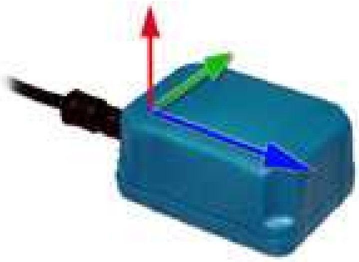


Figure 5. Illustration of MT9.

3.2 G-link

G-Link of MicroStrain is a high speed, triaxial accelerometer node, designed to operate as part of an integrated wireless sensor network system [2], as shown in Figure 6. The Base Station transceiver may trigger data logging (from 30 meters), or request previously logged data to be transmitted to the host PC for data acquisition/display/analysis. Featuring 2 KHz sweep rates, combined with 2 Mbytes flash memory, these little nodes pack a lot of power in a small package. Every node in the wireless network is assigned a unique 16 bit address, so a single host transceiver can address thousands of multichannel sensor nodes. The Base Station can trigger all the nodes simultaneously, and timing data is sent by the Base Station along with the trigger. This timing data is logged by the sensor nodes along with sensor data.

G-Link may also be wirelessly commanded to transmit data continually, at 1 KHz sweep rates, for a pre-programmed time period. The continuous, fast wireless transmission mode allows for real time data acquisition and display from a single multichannel sensor node at a time. G-link has two acceleration ranges: +/- 2 G’s and +/- 10 G’s, whilst its battery lifespan can be 273 hours. Furthermore, this product has a small transceiver size: $25\times25\times5\ mm^2$. A G-Link starter kit, inlcuding two G-Link data-logging transcevers (+/- 10 G full scale range), one basestation, all nec-

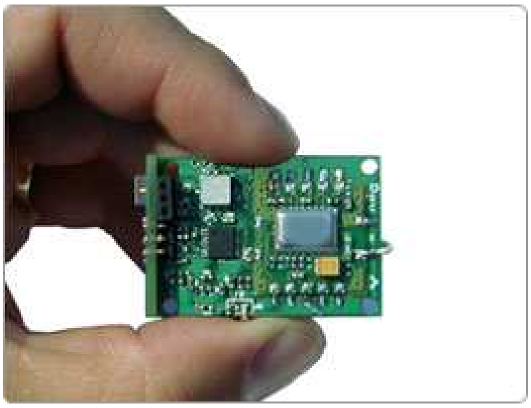


Figure 6. A G-link unit.

essary software and cables, costs about 2,000 US dollars.

As a wireless sensor, 3DM-G combines angular rate gyros with three orthogonal DC accelerometers, three orthogonal magnetometers to output its orientation. This product can be operated over the full 360 degrees of angular motion on all three axes with +/- 300 degrees/sec angular velocity range, 0.1 degrees repeatability and +/- 5 degrees accuracy. A gyro enhanced 3-axis orientation system starter kit 3DM-G-485-SK, consisting of one 3DM-G-485-M orientation module, one 3DM-G-485-CBL-PWR communication cable and power supply, a 3MG-G Software Suite for Win 95/98/2000/XP and a user manual, costs 1,500 US dollars (approx.).

3.3 MotionStar

MotionStar is a magnetic motion capture system produced by the Ascension Technology Corporation in the USA. This system applies DC magnetic tracking technologies, which are significantly less susceptible to metallic distortion than AC electronmagnetic tracking technologies. It provides real-time data output, capturing significant amounts of motion data in short order. Regardless of the number of sensors tracked, one can get up to 120 measurements per sensor per second. This system achieves six degree-of-freedom measurements, where each sensor calculates both position (x, y, z) and orientation (azimuth, elevation, roll) for a full 360 degrees coverage without the “line of sight” blocking problems of optical systems. There are 6 data points sampled by each sensor so fewer sensors are demanded. The communication between the console and the sensors is wireless.

Figure 7. Motionstar Wireless 2.

Figure 8. InterSense IS-300 Pro.

MotionStar Wireless 2 (Figure 7) is a magnetic tracker for capturing the motion of one or more performers. Data is sent via a wireless communications link to a base-station. It holds such good performance as: (1) translation range: +/- 3.05 m; (2) angular range: all attitude - +/- 180 deg for Azimuth and Roll, +/- 90 deg for Elevation; (3) static resolution (position): 0.08 cm at 1.52 m range; (4) static resolution (orientation): 0.1 RMS at 1.52 m range. Unfortunately, the communication range is only 12 feet (radius).

A vital drawback is that this system with six sensors costs around 56,000 US dollars.

3.4 InterSense

InterSense has its updated product IS-300 Pro Precision Motion Tracker shown in Figure 8. This system virtually eliminated the jitter common to other systems. It is featured with update rates of up to 500 Hz, steady response in metal-cluttered environments. The signal processor was small enough to wear on a belt for tetherless application. Furthermore, this system was the only one which predicted motion up to 50 ms and compensated for graphics rendering delays and further contributed to eliminating simulator lag. Therefore, it has been used successfully to implement feed-forward motion prediction strategies.

3.5 Polhemus

Polhemus [3] is the number one global provider of 3-D position/orientation tracking systems, digitizing technology solutions, eye-tracking systems and handheld three-dimensional scanners. The company was founded in 1969 by Bill Polhemus in Grand Rapids, MI. In early 1971 Polhemus moved to the Burlington area. Polhemus provided several novel fast and easy digital tracking systems: LIBERTY, FASTRAK and PATRIOT.

3.5.1 LIBERTY

This was the forerunner in electromagnetic tracking technology (Figure 9). LIBERTY computed at an extraordinary rate of 240 updates per second per sensor with the ability to be upgraded from four sensor channels to eight, by the addition of a single circuit board. Also, it had a latency of 3.5 milliseconds, a resolution of .00015 in (0.038 mm) at 12 in. (30 cm) range; and a 0.0012 orientation. The system provided an easy, intuitive user interface. Application uses were boundless, from biomechanical, and sports analysis, to virtual reality.

Figure 9. Illustration of LIBERTY by Polhemus.

3.5.2 FASTRAK

FASTRAK was a solution for accurately computing position and orientation through space (Figure 10). With real time, six-degree-of-freedom tracking and virtually no latency, this award-winning system was ideal for head, hand, and instrument tracking, as well as biomedical motion and limb rotation, graphic and cursor control, stereotaxic localization, telerobotics, digitizing, and pointing.

Figure 10. Illustration of FASTRAK by Polhemus.

3.5.3 PATRIOT

PATRIOT was a cost effective solution for six-degree-of-freedom tracking and 3-D digitizing. A good answer for the position/orientation sensing requirements of 3-D applications and environments where cost is a primary concern, it was ideal for head tracking, biomechanical analysis, computing graphics, cursor control, and sterotaxic localization. See Figure 11.

3.6 HASDMS-I

Human Performance Measurement, Inc. provided the HASDMS-I Human Activity State Detection and Monitoring System [4]. The Model HASDMS-I is a system designed to detect and log

Figure 11. Illustration of PATRIOT by Polhemus.

Figure 12. HASDMS-I from Human Performance Measurement, Inc.

Figure 13. Illustration of a glove-based prototype (image courtesy of KITTY TECH).

selected human activity states over prolonged periods (up to 7 days). It consists of a Sensing and Logging Unit (SLU) and Windows-based Host Software that runs on a user supplied PC. The system is based on the observation that while humans engage in activities which are often quite complex dynamically and kinematically, there are distinct patterns that lead us to identify these activities with specific words such as standing, walking, etc. Such words are referred to as “activity states”.

The HASDMS-I was designed to provide the greatest activity discrimination with the smallest possible sensor array (i.e., one sensing site on the body). The SLU is a compact, battery powered instrument with special sensors and a microprocessor that is mounted to the lateral aspect of the monitored subject’s thigh. It detects four unique activity states: (1) lying-sitting (grouped), (2) standing, (3) walking, and (4) running. A fifth state (”unknown”) is also provided to discriminate unusual patterns from those which the system is designed to detect.

The SLU is first connected to a Host PC (via a simple serial port connection) for initialization and start-up. It is then attached to a subject for an unsupervised monitoring session. When the session is complete, the SLU is again connected to the Host PC and logged data is uploaded to the host software for databasing, display, and analysis. Several standard activity summaries are provided including (1) the percent time spent in different states and (2) the total amount of time (hours, minutes, seconds) spent in different states. In addition, an Activity State History Graph depicts the type and duration of each state in a time-sequenced, scrollable window. Activity state data can be printed or exported in the form of an ASCII text file for any other user-specified analyses. The HASDMS-1 system is shown in Figure 12.

3.7 Glove-based analysis

Since the late 1970s people have studied glove-based devices for the analysis of hand gestures. Glove-based devices adopt sensors attached to a glove that transduces finger flexion and abduction into eletrical signals to determine the hand pose (Figure 13).

The Dataglove (originally developed by VPL Research) was a neoprene fabric glove with two fiber optic loops on each finger. Each loop was dedicated to one knuckle and this can be a problem. If a user has extra large or small hands, the loops will not correspond very well to the actual knuckle position and the user will not be able to produce very accurate gestures. At one end of each loop is an LED and at the other end is a photosensor. The fiber optic cable has small cuts along its length. When the user bends a finger, light escapes from the fiber optic cable through these cuts. The amount of light reaching the photosensor is measured and converted into a measure of how much the finger is bent. The Dataglove requires recalibration for each user [117].

The CyberGlove system included one CyberGlove [5], an instrumentation unit, a serial cable to connect to your host computer, and an executable version of the VirtualHand graphic hand model display and calibration software. Many applications require measurement of the position and orientation of the forearm in space. To accomplish this, mounting provisions for Polhemus and Ascension 6 degrees of freedom tracking sensors are available for the glove wristband. Tracking sensors are not included in the basic CyberGlove system. The CyberGlove had a software programmable switch and LED on the wristband to permit the system software developer to provide the CyberGlove wearer with additional input/output capability. The instrumentation unit provided a variety of convenient functions and features including time-stamp, CyberGlove status, external sampling synchronization and analog sensor outputs.

Based on the design of the DataGlove, PowerGlove was developed by Abrams-Gentile Entertainment (AGE Inc.) for Mattel through a licensing agreement with VPL Research. PowerGlove consists of a sturdy Lycra glove with flat plastic strain gauge fibers coated with conductive ink running up each finger; which measure change in resistance during bending to measure the degree of flex for the finger as a whole. It employs an ultrasonic system (back of glove) to track the roll of the hand (reported in one of twelve possible roll positions), ultrasonic transmitters must be oriented toward the microphones to get an accurate reading; pitching or yawing hand changes orientation of transmitters and signal would be lost by the microphones; poor tracking mechanism. (4D - x, y, z, roll).

Similar technologies can be referred to 5DT DataGlove [6], PINCH Gloves [7], and Hand Master [8].

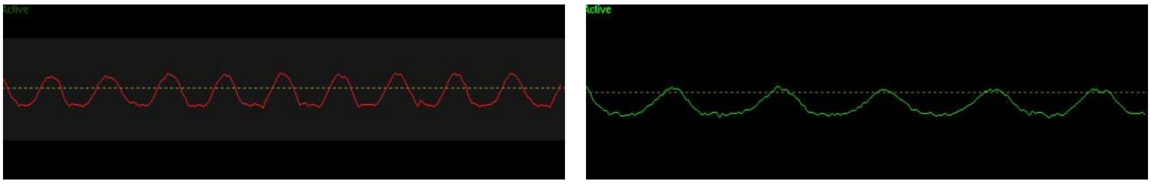
3.8 Non-commercial systems

The commercial systems described earlier accommodate stable and consistent technologies. Nevertheless, they are sold with high prices. This extremely limits the applications of these systems in the community. As a result, people intend to propose some affordable, compact and friendly systems instead. In this context, an example is given as follows.

Dukes [45] developed a compact system which comprised two parts, an embedded hand unit that encapsulated the necessary hardware for capturing human arm movement and a software interface implemented in a computer terminal for displaying the collected data.

Within the embedded hand unit a microcontroller gathered data from two accelerometers. The collected data was then transmitted to the computer terminal for the purpose of analysis. The software interface was implemented to collect data from the embedded hand unit. The data was presented to the user both statically and dynamically in the form of a three dimensional animation operation. The whole system successfully captured human movement. Moreover, the transference of data from the hand unit to the terminal was consistently achieved in an asynchronized mode. In the computer terminal, the collected data was clearly illustrated for representing the continuous sampling of the FM transmitter and receiver modules, demonstrated in Figure 14.

However, the animation shown in the terminal failed to correct display human movement in terms of distance travelled and speed of movement. This is due to the direct output of the data from the accelerometers without any calculation with respect to the distance. To perform a correct demonstration, this data needs to be resampled and further processed in the terminal based on the travelled distance and its corresponding time.



(a) (b)

Figure 14. Demo of Dukes's approach: (a) collected data on x-axis, and (b) collected data on y-axis.

3.9 Other techniques

Acoustic systems collect information by transmitting and sensing sound waves, where the flight duration of a brief ultrasonic pulse is timed and calculated. These systems are being used in medical applications, [46], [83], [91], but have not been used in motion tracking. This is due to inherent drawbacks corresponding to the ultrasonic systems: (1) the efficiency of an acoustic transducer is proportional to the active surface area so large devices are demanding; (2) to improve the detected range the frequency of ultrasonic waves must be low (e.g. 10Hz) but this affects system latency in continuous measurement; (3) acoustic systems require a line of sight between the emitters and the receivers.

Radio and microwaves are normally used in navigation systems and airports landing aids [108] although they have no application in the human motion tracking. Electromagnetic wave-based tracking approaches can provide range information by calculating the radiated energy dissipated in a form of radius r as $1/r^2$. For example, using a delay-locked loop (DL) the Global Positioning System (GPS) can achieve a resolution of 1 meter. Obviously, this is not enough for the human motion that is usually of 40-50 cm displacements per sec. The only radio frequency-based precision motion tracker can be of a surprisingly good resolution of a few millimeters, but it used large racks of microwave equipment and was demonstrated in an empty room. That is to say, a hybrid system is potential to obtain higher resolution but incurs integration difficulties.

4 Vision based tracking systems with markers

In 1973 Johansson explored his famous Moving Light Display (MLD) psychological experiment to perceive biological motion [69]. He attached small reflective markers to the joints of human subjects, which allow these markers to be monitored during trajectories. This experiment became the milestone of human movement tracking. Although Johansson's work established a solidate theory for human movement tracking, it still faces the challenges such as errors, non-robustness and expensive computation due to environmental constraints, mutual occlusion and complicated processing. However, tracking systems with markers minimize uncertainty of subject movements due to the unique appearance of the markers. Consequently, plenty of marker-based tracking systems are nowadays available in the market. Study of these systems allows their advantages to be exploited in a further developed platform.

4.1 Qualisys

A Qualisys motion capture system depicted in Figure 15 consists of 1 to 16 cameras, each emitting a beam of infrared light [9]. Small reflective markers are placed on the object or person to be measured. The camera flash infrared light and the markers reflect it back to the camera. The camera then measures a 2-dimensional position of the reflective target by combining the 2-D data from several cameras a 3D position is calculated. The data can be analyzed in Qualisys Motion Manager

Figure 15. An operating Qualisys system.

Figure 16. Reflective markers used in a real-time VICON system.

(QMM) or is exported in several external formats.

This system can be combined with Visual3D, an advanced analysis package for managing and reporting optical 3-D data, to track each segment of the model. The pose (position and orientation) of each segment is determined by 3 or more non-collinear points attached to the segment. For better accuracy, a cluster of targets can be rigidly attached to a shell. This prevents the targets from moving relative to each other. This shell is then affixed to the segment. The kinematics model is calculated by determining the transformation from the tracking targets recorded to a calibration pose.

4.2 VICON

VICON, a 3-D optical tracking system, was specifically designed for use in virtual and immersive environments [63]. By combining Vicons high-speed, high-resolution cameras with new automated Tracker software, the system delivers immediate and precision manipulation of graphics for first person immersive environments for military, automotive, and aerospace visualizations. Precise, low-latency and jitter free motion tracking, though key to creating a realistic sense of immersion in visualizations and simulations, has not been possible previously due to lag, inaccuracies, unpredictability and unresponsiveness in electromagnetic, inertial and ultrasonic tracking options.

The VICON Tracker, which offers wireless, extreme low-latency performance with six degrees of freedom (DOF) and zero environmental interference, outclasses these obsolescent systems, yet is the simplest to set up and calibrate. Targets are tracked by proprietary CMOS VICON cameras ranging in resolution from 640x480 to 1280x1024 and operating between 200-1000 Hz. The entire range of cameras are designed, developed and built specifically for motion tracking.

At the heart of the system, the VICON Tracker software automatically calculates the center of every marker, reconstructs its 3-D position, identifies each marker and object, and outputs 6 DOF information typically in less than 7 milliseconds. The strength of Vicon Tracker software lies in its automation. The very first Tracker installation requires about an hour of system set-up; each following session requires only that the PC running the software be switched on. Objects with three or more markers will automatically output motion data that can be applied to 3-D objects in real time, and integrated into a variety of immersive 3-D visualization applications, including EDS Jack, Dassault Delmia, VRCOM, Fakespace, VRCO Track D and others. Figure 16 shows that reflective markers within a real-time VICON system are applied to two subjects.

Figure 17. CODA system.

4.3 CODA

CODA is an acronym of Cartesian Opto-electronic Dynamic Anthropometer, a name first coined in 1974 to give a working title to an early research instrument developed at Loughborough University, United Kingdom by David Mitchelson and funded by the UK Science Research Council [10], illustrated in Figure 17.

The system was pre-calibrated for 3-D measurement, which means that the lightweight sensor can be set up at a new location in a matter of minutes, without the need to recalibrate using a space-frame. Up to six sensor units can be used together and placed around a capture volume to give extra sets of eyes and maximum redundancy of viewpoint. This enables the Codamotion system to track 360 degree movements which often occur in animation and sports applications. The active markers were always intrinsically identified by virtue of their position in a time multiplexed sequence. Confused or swapped trajectories can never happen with the Codamotion system, no matter how many markers are used or how close they are to each other.

The calculation of the 3-D coordinates of markers was done in real-time with an extremely low delay of 5 milliseconds. Special versions of the system were available with latency shorter than 1 millisecond. This opens up many applications that require real-time feedback such as research in neuro-physiology and high quality virtual reality systems as well as tightly coupled real-time animation. It was also possible to trigger external equipment using the real-time Codamotion data. At a three metre distance, this system has such good accurate parameters as follows: +/-1.5 mm in X and Z axes, +/- 2.5 mm in Y axis for peak-to-peak deviations from actual position.

Figure 18. ReActor2 system.

4.4 ReActor2

As products of Ascension Tech. Corporation ReActor2 digital active-optical tracking systems shown in Figure 18 capture the movements of an untethered performer C free to move in a capture area bordered by modular bars that fasten together. The digital detectors embedded in the frame provide full coverage of performers while minimizing blocked markers. The Instant Marker Recognition instantly reacquires blocked markers for clean data. This means less post processing and a more efficient motion capture pipeline [11].

Up to 544 new and improved digital detectors embedded in a 12-bar frame and over 800 active LEDs flashing per measurement cycle for complete tracking coverage. A sturdy, ruggedized frame eliminiates repetitive camera calibration and tripod alignment headaches. Set up the system once and start capturing data immediately.

4.5 ELITE Biomech

ELITE Biomech from BTS of Italy is based on the latest generation of ELITE systems:

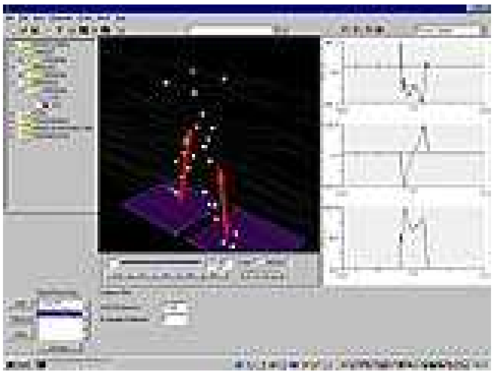


Figure 19. Demo of ELITE Biomech's outcomes.

ELITE2002. ELITE2002 performs a highly accurate reconstruction of any type of movement, on the basis of the principle of shape recognition of passive markers.

3D reconstruction and tracking of markers starting from pre-defined models of protocols are widely validated by the international scientific community. Tracking of markers based on the principle of shape recognition allows the use of the system in extreme conditions of lighting. This system is capable of managing up to 4 force platforms of various brands, and up to 32 electromyographic channels. It also runs in real time recognition of markers with on-monitor-display during the acquisition, and real time processing of cinematic and analog data, demonstrated in Figure 19.

4.6 APAS

The Ariel Performance Analysis System (APAS) [12] is the premier products designed, manufactured, and marketed, by Ariel Dynamics, Inc. It is an advanced video-based system operating from the Windows 95/98/NT/2000 environments. Specific points of interest are digitized with user intervention or automatically using contrasting markers. Additionally, analog data (i.e. force platform, EMG, goniometers etc.) can be collected and synchronized with the kinematic data. Although the system has primarily been used for quantification of human activities, it has also been utilized in many industrial, non-human applications. Optional software modules include real-time 3D (6 degree of freedom) rendering capabilities and full gait pattern analysis utilizing all industry standard marker sets.

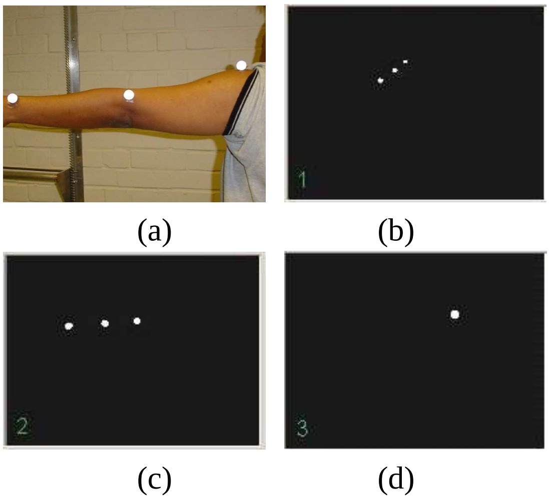
Figure 20. The Polaris system.

4.7 Polaris

The Polaris system (Northern Digital Inc.) [13] is of real-time tracking flexibility for comprehensive purposes, including academic and industrial environments. This system optimally combines simultaneous tracking of both wired and wireless tools (Figure 20).

The whole system can be divided into two parts: the position sensors and passive or active markers. The former consist of a couple of cameras that are only sensitive to infrared light. This design is particularly useful when the background lighting is varying and unpredictive. Passive markers are covered by reflective materials, which are activated by the arrays of infrared light-emitting diodes surrounding the position sensor lenses. In the meantime, active markers can emit infrared light themselves. The Polaris system is able to provide 6 degrees of freedom motion information. With proper calibration, this system may achieve 0.35 mm RMS accuracy in position measures. A basic Polaris with a software development kit (SDK) costs about $2000.

However, similar to other marker-based techniques, the Polaris system cannot sort out the occlusion problem due to the existence of the line of sight. Adding extra position sensors possibly mitigates the trouble but also increases computa-



(a) (b)

(c) (d)

Figure 21. Demo of Tao and Hu's approach: (a) markers attached to the joints; (b), (c) and (d) marker points captured from three cameras.

tional cost and operational complexity.

4.8 others

Other commercial marker-based systems are given in [14], [15], [16].

By combining with the commercial marker-based systems, people have developed some hybrid techniques to implement human motion tracking. These systems, although still in the experimental stage, already demonstrate encouraging performance. For example, Tao and Hu [103] built a visual tracking system, which exploited both marker-based and marker-free tracking methods. The proposed system consisted of three parts: a patient, video cameras and a PC. The patient's motion was filmed by video cameras and the captured image sequences were input to the PC. The software in the PC comprised of three modules: motion tracking module, database module and decision module. The motion tracking module was formulated in an analysis-by-synthesis framework, which was similar to the strategy introduced by O'Rourke and Badler [82]. In order to enhance the prediction component, a marker-based motion learning method was adopted: small retro-reflective ball markers were attached to the performer's joints, which reflected infrared light so that cameras picked up the bright points indicating the cameras' positions. The 3-D position of each marker was calculated by corresponding a 2-D marker point in an image plane via the epipolar constraint. The skeleton motion of the performer was then deduced [37]. By using the perspective camera models, the 3-D model recovered previously was projected into 2-D image planes, which behaved as the prediction of the matching framework. The performance of this approach is demonstrated in Figure 21.

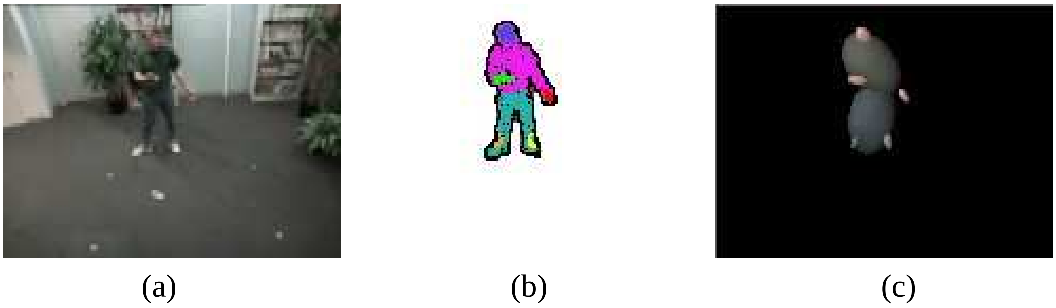
5 Vision based tracking systems without markers

In the previous section, we described the features of the marker-based tracking systems, which are restritive to some degree due to the mounted markers. As a less restritive motion capture technique, markerless based systems are capable of overcoming the mutual occlusion problem as they are only concerned about boundaries or features on human bodies. This is an active and promising but also challenging research area in the last decade. The research with respect to this area is still ongoing due to unsolved technical problems.

From a review's point of view, Aggarwal and Cai [17] classified human motion analysis as: body structure analysis (model and non-model based), camera configuration (single and multiple), and correlation platform (state-space and template matching). Gavrila [51] claimed that the dimensionality of the tracking space, e.g. 2-D or 3-D, be mainly focused. To be coincident with these exiting definitions, we suggest to contain all these issues in this context.

5.1 2-D approaches

As a commonly used framework, 2-D motion tracking only concerns the human movement in an image plane, although sometimes people intend to project a 3-D structure into its image plane for processing purposes. This approach can be catalogued with and without explicit shape mod-



(a) (b) (c)

Figure 22. Demonstration of Pfinder by Wren, et al.

els.

5.1.1 2-D approches with explicit shape models

Due to the arbitrary movements of humans self-occlusion exists during human trajetories. To sort out this problem, one normally uses a priori knowledge about human movements in 2-D by segmenting the human body. For example, Wren et al. [109] presented a region-based approach, where they regarded the human body as a set of “blobs” which can be described by a spatial and color Gaussian distribution. To initialize the process, a foreground region can be extracted given the background model. The blobs, representing human hands, head, etc., are then placed over the foreground region instead. A 2-D contour shape analysis was undertaken to identify various body parts. The working flowchart is referred to Figure 22.

Akita [18] explored an approach to segment and track human body parts in common circumstances. To prevent the body tracking from collapsing, he presumed the human movements are known *a priori* in some kind of “key frames”. He followed the tracking order, legs, head, arms, and trunk, to detect the body parts. However, due to simplification his model works in some special situations.

Long and Yang [75] advocated that the limbs of a human silhouette could be tracked based on the shapes of the antiparallel lines. They also conducted experimental work to cope with occlusion, i.e. disappearance, merging and splitting. Kurakake and Nevatta [74] attempted to obtain joint locations in images of walking humans by establishing correspondence between extracted ribbons. Their work assumed small motion between two consecutive frames, and feature correspondence was conducted using various geometric constraints.

Shimada *et al.* [95] suggested to achieve rapid and precise estimation of human hand postures by combining 2-D appearance and 3-D model-based fitting. First, a rough posture estimate was obtained by image indexing. Each possible hand appearance generated from a given 3-D shape model was labeled by an index obtained by PCA compression and registered with its 3-D model parameters in advance. By retrieving the index of the input image, the method obatined the matched appearance image and its 3-D parameters rapidly. Then, starting from the obtained rough estimate, it estimated the posture and moreover refined the given initial 3-D model by model-fitting.

Figure 24. Computer game on-chip by Freeman, W. et al.

Figure 23. Human tracking in the approach of Baumberg and Hogg.

5.1.2 2-D approaches without explicit shape models

This is a more often addressed topic. Since human movements are non-rigid and arbitrary, boundaries or silhouettes of human body are viable and deformable, leading to difficult description for them. Tracking human body, e.g. hands, is normally achieved by means of background substraction or color detection. Furthermore, due to the unavailability of models one has to attend low level image processing such as feature extraction.

Baumberg and Hogg [21] considered using Active Shape Model (ASM) for tracking pedestrians (Figure 23). B-splines were used to represent different shapes. The foreground region was first extracted by substracting the background. A Kalman filter was then applied to accomplish the spatio-temporal operation, which is similar to the work of Blake et al [26]. Their work was then extended by automatically generating an improved physically based model using a training set of examples of the object deforming, tuning the elastic properties of the object to reflect how the object actually deforms. The resulting model provides a low dimensional shape description that allows accurate temporal extrapolation at low computational cost based on the training motions [22].

Freeman et al. [49] developed a special detector for computer games on-chip (Figure 24), which is to infer useful information about the position, size, orientation, or configuration of the human body parts. Two algorithms were used, one of which used image moments to calculate an equivalent rectangle for the current image, and the other used orientation histograms to select the body pose from a menu of templates.

Cordea *et al.* [39] discussed a 2.5 dimensional tracking method allowing real-time recovery of the 3-D position and orientation of a head moving in its image plane. This method used a 2-D elliptical head model, a region- and edge-based matching algorithms, and a Linear Kalman Filter estimator. The tracking system worked in a realistic situation without makeup on the face, with an uncalibrated camera, and unknown lighting conditions and background.

Fablet and Black [47] proposed a solution for the automatic detection and tracking of human motion using 2-D optical flow information, which provided rich descriptive cues, while being independent of object and background appearance. To represent the optical flow patterns of people from arbitrary viewpoints, they developed a novel representation of human motion using low-dimensional spatio-temporal models that were learned using motion capture data of human subjects. In addition to human motion (the foreground) they modelled the motion of generic scenes (the background); these statistical models were defined as Gibbsian fields specified from the first-order derivative of motion observations. Detection and tracking were posed in a principled Bayesian framework which involved the computation of a posterior probability distribution over the model parameters. A particle filter was then used to represent and predict this non-Gaussian posterior distribution over time.The model parameters of samples from this distribution were related to the pose parameters of a 3-D articulated model.

5.2 3-D approaches

These approaches attempted to recover 3-D articulated poses over time [51]. People usually project a 3-D model into a 2-D image for substantial processing. This is due to the application of image appearance and dimensional reduction.

5.2.1 3-D modelling

Modelling human movements *a priori* allows the tracking problem to be minimized: the future movements of the human body can be predicted regardless of self-occlusion or self-collision. O'Rourke and Badler [82] discovered that the prediction in state space seemed more stable than that in image space due to the incorporated semantic knowledge in the former. In their tracking framework, four components were inlcuded: prediction, synthesis, image analysis, and state estimation. This strategy has been applied to most of the existing tracking systems.

Model-based approaches contain stick figures, volumetric and a mixture of models.

5.2.2 Stick figure

The stick figure is the representation of the skeletal structure, which is normally regarded as a collection of segments and joint angles (Figure 25). Bharatkumar *et al* [23] used stick figures to model the lower limbs, e.g. hip, knees, and ankles. They applied a medial-axis transformation to extract 2-D stick figures of the lower limbs.

Chen and Lee [37] first applied geometric projection theory to obtain a set of feasible postures from a single image, then made use of the given dimensions of the human stick figure, physiological and motion-specific knowledge to constrain the feasible postures in both the single-frame analysis and the multi-frame analysis. Finally a unique gait interpretation was selected by an optimization algorithm.

Huber's human model [65] was a refined version of the stick figure representation. Joints were connected by line segments with a certain degree of constraint that could be relaxed using "virtual springs". This model behaved as a mass-spring-damper system. Proximity space (PS) was used to confined the motion and stereo measurements of joints, which started from the human head and extended to arms and torso through the expansion of PS.

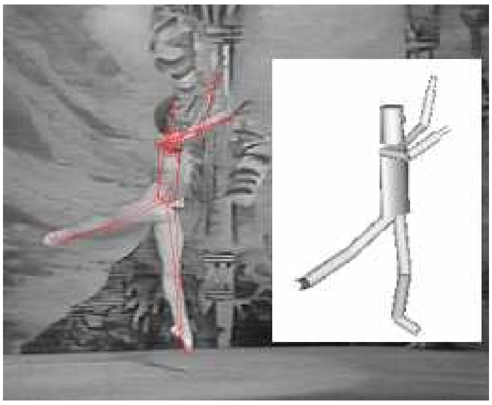


Figure 25. Stick figure of human body (image courtesy of Freeman, W.T.).

By modelling a human body with 14 joints and 15 body parts, Ronfard *et al.* [93] attempted to find people in static video frames using learned models of both the appearance of body parts (head, limbs, hands), and of the geometry of their assemblies. They built on Forsyth and Fleck's general 'body plan' methodology and Felzenszwalb and Huttenlocher's dynamic programming approach for efficiently assembling candidate parts into 'pictorial structures'. However they replaced the rather simple part detectors used in these works with dedicated detectors learned for each body part using Support Vector Machines (SVMs) or Relevance Vector Machines (RVMs). RVMs are SVM-like classifiers that offer a well-founded probabilistic interpretation and improved sparsity for reduced computation. Their benefits were demonstrated experimentally in a series of results showing great promise for learning detectors in more general situations.

Further technical reports are given in [50], [61], [67], [81], [86].

5.2.3 Volumetric modeling

Elliptical cylinders are one of the volumetric models that model human body. Hogg [60] and Rohr [92] extended the work of Marr and Nishihara [78], which used elliptical cylinders for representing the human body. Each cylinder consisted of three parameters: the length of the axis, the major and minor axes of the ellipse cross section. The coordinate system originated from the center of the torso. The difference between the two approaches is that Rohr used eigenvector line fitting to project the 2-D image onto the 3-D human model.

Rehg *et al.* [87] represented two occluded fingers using several cylinders, and the center axes of cylinders were projected into the center line segments of 2-D finger images. Goncalves *et al.* [52] modelled both the upper and lower arm as truncated circular cones, and the shoulder and elbow joints were presumably spherical joints. A 3-D arm model was projected to an image plane and then fitted to the blurred image of a real arm. The maching was acheived by minimizing the error between the model projection and the real image adapting the size and the orientation of the model.

Chung and Ohnishi [38] proposed a 3-D model-based motion analysis which used *cue circles* (CC) and *cue sphere* (CS). Stereo matching for reconnstructing the body model was performed by finding pairs of CC between the pair of contour images investigated. A CS needed to be projected back onto two image planes with its corresponding CC.

Theobalt *et al.* [105] suggested to combine efficient real-time optical feature tracking with the reconstruction of the volume of a moving subject to fit a sophisticated humanoid skeleton to the video footage. The scene is observed with 4 video cameras, two connected to one PC (Athlon 1GHz). The system consisted of two parts: a distributed tracking and visual hull reconstruction system (online component), and a skeleton fitting application that took recorded sequences as input. For each view, a moving person was separated from background by a statistical background subtraction. In the initial frame, the silhouette of the person seen from the 2 front view cameras was separated into distinct regions using a Generalized Voronoi Diagram Decomposition. The locations of its hands, head and feet could now be identified. In the front camera view for all video frames after initialization the locations of these body parts could be tracked and their 3-D location reconstructed. In addition a voxel-based approximation to the visual hull was computed for each time step. The experimental volumetric data was given in Figure 26.

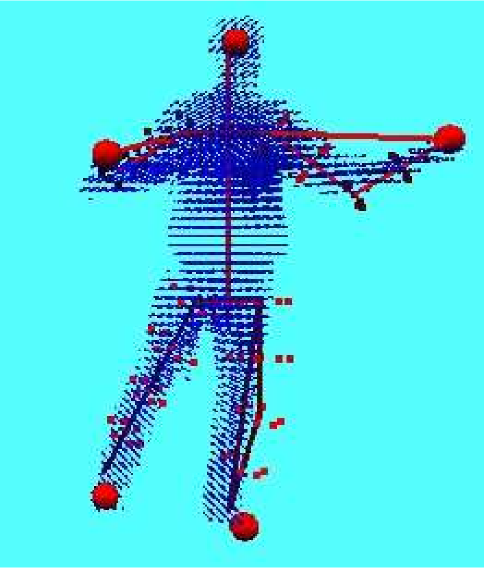


Figure 26. Volumetric modelling by Theobalt, C.

5.3 Camera configuration

The tracking problem can be tackled by proper camera setup. Literature has been linked with a single camera and a distributed-camera configuration. Using multiple cameras does require a common spatial reference to be employed, and a single camera does not have such a requirement. However, a single camera from time to time suffers from the occlusion of the human body due to its fixed viewing angle. Thus, a distributed-camera strategy is a better option of minimizing such a risk.

5.3.1 Single camera tracking

Polana and Nelson [84] observed that the movements of arms and legs converge to that of the torso. Each walking person image was bounded by a rectangular box, and the centroid of the bounding box was treated as the feature to track. Positions of the center point in the previous frames were used to estimate the current position. As such, correct tracking was conducted when the two subjects were occluded to each other even in the middle of the image sequences.

Sminchisescu and Triggs [98] present a method for recovering 3-D human body motion from monocular video sequences using robust image matching, joint limits and non-self-intersection constraints, and a new sample-and-refine search strategy guided by rescaled cost-function covariances. Monocular 3-D body tracking is challenging: for reliable tracking at least 30 joint parameters need to be estimated, subject to highly nonlinear physical constraints; the problem is chronically ill-conditioned as about 1/3 of the d.o.f. (the depth-related ones) are almost unobservable in any given monocular image; and matching an imperfect, highly flexible, self-occluding model to cluttered image features is intrinsically hard. To reduce correspondence ambiguities they used a carefully designed robust matching-cost metric that combined robust optical flow, edge energy, and motion boundaries. Even so, the ambiguity, nonlinearity and non-observability made the parameter-space cost surface be multi-modal, unpredictable and ill-conditioned, so minimizing it is difficult. They discussed the limitations of CONDENSATION-like samplers, and introduced a novel hybrid search algorithm that combined inflated-covariance-scaled sampling and continuous optimization subject to physical constraints. Experiments on some challenging monocular sequences showed that robust cost modelling, joint and self-intersection constraints, and informed sampling were all essential for reliable monocular 3-D body tracking.

Bowden *et al.* [29] advocated a model based approach to human body tracking in which the 2-D silhouette of a moving human and the corresponding 3-D skeletal structure were encapsulated within a non-linear Point Distribution Model. This statistical model allowed a direct mapping to be achieved between the external boundary of a human and the anatomical position. It showed that this information, along with the position of lanmark features, e.g. hands and head, could be used to reconstruct information about the pose and structure of the human body from a monoscopic view of a scene.

Barron and Kakadiaris [20] present a simple, efficient, and robust method for recovering 3-D human motion capture from an image sequence obtained using an uncalibrated camera. The proposed algorithm included an anthropometry initialization step, assuming that the similarity of appearance of the subject over the time of acquisition led to the minimum of a convex function on the degree of freedom of a Virtual Human Model (VHM). The method searched for the best pose in each image by minimizing discrepancies between the image under consideration and a synthetic image of an appropriate VHM. By including on the objective function penalty factors from the image segmentation step, the search focused on regions that belong to the subject. These penalty factors converted the objective function to a convex function, which guaranteed that the minimization converged to a global minimum.

To reduce side-effects of hard kinematic constraints, Dockstader *et al.* [44] proposed a new model-based approach toward three-dimensional (3-D) tracking and extraction of gait and human motion. They suggested the use of a hierarchical, structural model of the human body that introduced the concept of soft kinematic constraints. These constraints took the form of a priori, stochastic distributions learned from previous configurations of the body exhibited during specific activities; they were used to supplement an existing motion model limited by hard kinematic constraints. Time-varying parameters of the structural model were also used to measure gait velocity, stance width, stride length, stance times, and other gait variables with multiple degrees of accuracy and robustness. To characterize tracking performance, a novel geometric model of ex-

Figure 27. Human motion tracking by Black, M.J. et al.

pected tracking failures was then introduced.

Yeasin and Chaudhuri [113] proposed a simple, inexpensive, portable and real-time image processing system for kinematic analysis of human gait. They viewed this as a feature based multi-target tracking problem. They tracked the artificially induced features appearing in the image sequence due to the non-impeding contrast markers attached at different anatomical landmarks of the subject under analysis. The paper described a real-time algorithm for detecting and tracking feature points simultaneously. By applying a Kalman filter, they recursively predicted tentative features location and retained the predicted point in case of occlusion. A path coherence score was used for disambiguation along with tracking for establishing feature correspondences. Experimentations on normal and pathological subjects in different gait was performed and results illustrated the efficacy of the algorithm. Similar algorithms to this one can be found in [115] and [114].

Further to the application of optical flow in the motion learning, Black *et al.* [25] proposed a framework for learning parameterized models of optical flow from image sequences. A class of motion is represented by a set of orthogonal basis flow fields that were computed from a training set using principal component analysis. Many complex image motion sequences can be represented by a linear combination of a small number of these basis flows. The leared motion models may be used for optical flow estimation and for model-based recognition. They described a robust, multi-resolution scheme for directly computing the parameters of the learned flow models from image derivatives. As examples they included learning motion discontinuities, non-rigid motion of humans, and articulated human motion. Later, Sidenbladh [96] *et al.*, also in [97], extended the work of [25] to a generative probabilistic method for tracking 3-D articulated human figures in monocular image sequences (see the example shown in Figure 27). These ideas similar to [66] that obtained further extension in [111], [112].

5.3.2 Multiple camera tracking

To enlarge the monitored area and to avoid the disappearance of subjects, a distributed-camera strategy is set up to solve the ambiguity of mactching when subjects are occluded to each other. Cai and Aggarwal [34] used multiple points belonging to the medial axis of the human upper body as the feature to track. These points were sparsely sampled and assumed to be independent of each other. Location and average intensity of feature points were integrated to find the most likely match between two neighboring frames. Multivariate Gaussian distributions were presumably addressed in the lcass-conditional probability density function of features of candidate subject images. It was shown that using such a system with three cameras indoors led to real time operation.

Sato *et al.* [94] represented a moving person as a combination of blobs of its body parts. All the cameras were calibrated in advance regarding the CAD model of an indoor environment. Blobs were corresponded using their area, brightness, and 3-D position in the world coordinates. The 3-D position of a 2-D blob was estimated on the basis of its height retrieved from the distance between the weight center of the blob and the floor.

Ringer and Lasenby [90] proposed to use markers placed at the joints of the arm(s) or leg(s) being analyzed, referred to Figure 28. The location

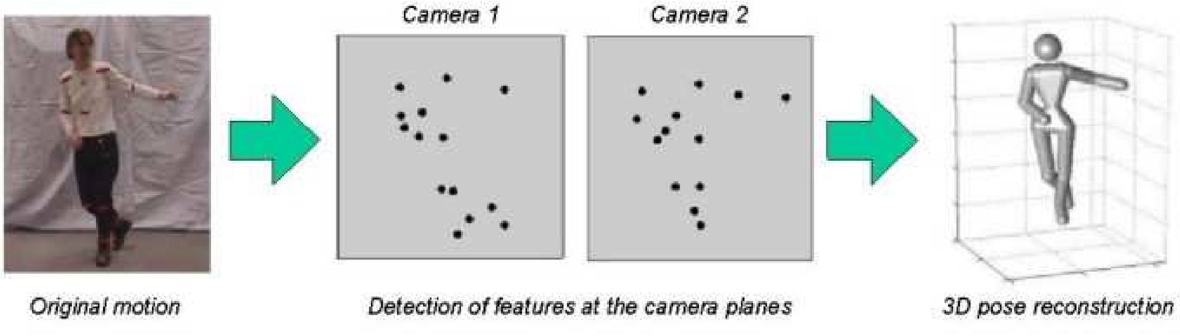


Figure 28. Applications of multiple cameras in human motion tracking by Ringer and Lasenby.

of these markers on a camera's image plane provided the input to the tracking systems with the result that the required parameters of the body could be estimated to far greater accuracy that one could obtain in the markerless case. This scheme used a number of video cameras to obtain complete and accurate information on the 3-D location and motion of bodies over time. Based on the extracted kinematics and measurement models, the extended Kalman filter (EKF) and particle filter tracking strategies were compared in their applications to update state estimates. The results justified that the EKF was preferred due to its less computational demands.

Rodor *et al.* [27] introduced a method for employing image-based rendering to extend the range of use of human motion recognition systems. Input views orthogonal to the direction of motion were created automatically to construct the proper view from a combination of non-orthogonal views taken from several cameras. Image-based rendering was utilized in two ways: (1) to generate additional training sets for these systems containing a large number of non-orthogonal views, and (2) to genrate orthogonal views from a combination of non-orthogonal views from several cameras.

Multiple cameras are needed to completely cover an environment for monitoring activity. To track people successfully in multiple perspective imagery, one has to establish correspondence between objects captured in multiple cameras. Javed *et al.* [68] presented a system for tracking people in multiple uncalibrated cameras. The system was able to discover spatial relationships between the camera field of views and uses this information to correspond between different perspective views of the same person. They explored the novel approach of finding limits of field of view of a camera as visible in other cameras. This helped disambiguate between possible candidates of correspondence.

5.4 Segmentation of human motion

Spatio-temporal segmentation, illustrated in Figure 29, is vital in vision related analysis due to the required reconstruction of dynamic scenes. Spatial segmentation attempts to extract moving objects from their backgorund, and divide a complicated motion stream into a set of simple and stable motions [55]. In order to fully depict human motion in constraint-free environments, people have explored a variety of motion segmentation strategies which consisted of both model-based and appearence-based approaches [55], [116], [77], [54], [99]. Nevertheless, it does not mean that segmentation is independently achieved. Instead, motion segmentation is normally encoded within the tracking procedure and performs like an assistive tool and descriptor.

Gonzalez et al. [53] estimated motion flows of features on human body using a standard tracker. Given a pair of subsequent images, an affine fundamental matrix was estimated by four pairs of corresponding feature points such that number of

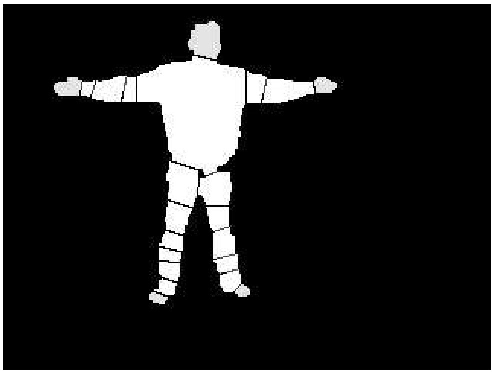


Figure 29. Segmentation of human body by Theobalt, C.

other feature points undergoing the affine motion modelled by the matrix should be maximized [118]. In the remaining subsequent frames, feature points corresponding to those used for the first fundamental matrix continued to estimate an affine motion model. At the last pair of frames, it led to a set of feature points identified as those belonging to a same limb and hence undergoing a same motion over the whole sequence. By repeating this estimate-and-sortout process over the remaining feature points, different limb motions were finally segmented.

To obtain a high level interpretation of human motion in a vedio stream one has to first detect body parts. Hilti et al. [59] proposed to combine both pixel-based skin color segmentation and motion-based segmentation for human motion tracking. The motivation of using skin color was raised due to its orientation invariant and fast detection. The human face normally presents a large skin surface in a flesh-tone, which is quite similar from person to person and even across various races [100]. Using hue and saturation (HS) as inputs, a color map was changed to a filtered image, where each pixel is associated with a likelihood of being [85]. For compensating the impacts of lighting changes, the motion-based segmentation was implemented and adaptive to exogenous changes.

Bradski and Davies [30] present a fast and simple method using a timed motion history image (tMHI) for representing motion from gradients in successively layered silhouettes. The segmented regions were not “motion blobs” but motion regions that were naturally connected to parts of moving objects. This movivated by the fact that segmentation by collecting “blobs” of similar direction motion frame to frame from optical flow [41] did not gurantee the correspondence of the motion over time. By labeling motion regions connected to the current silhouette using a downward stepping floodfill, areas of motion were directly attached to parts of the object of interest.

Moeslund and Granum [79] suggested to use colour information to segment the hand and head. To the sensitivity of orignal RGB-based colours to the intensity of lightling, they used chromatic colours which were normalised according to the intensity. In order to determine dance motion, human observers were shown video and 3-D motion capture sequences on a video display [70]. Observers were asked to define gesture boundaries within each microdance, which was analyzed to compute the local minima in the force of the body. At the moment of each of these local minima, the force, momentum, and kinetic energy parameters were examined for each of lower body segments. For each segment a binary triple was computed. It provided a complete characterization of all the body segments at each instant when body acceleration was at a local minimum.

6 Robot-guided tracking systems

In this section, one can find a rich variety of rehabiliation systems that are driven by electromechanical or electromagnetic tracking strategies. These systems, namely robot-guided systems hereafter, incorporate sensor technologies to conduct “move-measure-feedback” training strategies.

6.1 Discriminating static and dynamic activities

To distinguish static and dynamic activities (standing, sitting, lying, walking, ascending stairs, descending stairs, cycling), Veltink et al. [107] presented a new approach to monitoring ambulatory activities for use in the domestic environment, which uses two or three uniaxial accelerometers mounted on the body. They achieved a set of experiments with respect to the static or dynamic characteristics. First, the discrimination between static or dynamic activities was studied. It was illustrated that static activities could be distinguished from dynamic activities. Second, the distinction between static activities was investigated. Standing, sitting and lying could be distinguished by the output of two accelerometers, one mounted tangentially on a thigh, and the other sited on the sternum. Third, the distinction between a few cyclical dynamic activities was conducted.

As a result, it was concluded that the "discrimination of dynamic activities on the basis of the combined evaluation of the mean signal value and signal morphology is therefore proposed". This ruled out the standard deviation of the signal and the cycle time as the indexes of discriminating activities. The performance of the dynamic activity classification on the basis of signal morphology needs to be improved in the future work. The authors revealed that averaging adjacent motion cycles might reduce standard deviations of signal correlation so as to improve measure performance. As a futher study, Uiterwaal et al. [106] developed a measurement system using accelerometry to assess a patient's functional physical mobility in non-laboratory situations. Every second the video recording was compared to the measurment from the proposed system, and it showed that the validity of the system was up to 93%.

6.2 Typical working systems

6.2.1 Cozens

To justify whether motion tracking techniques can assist simple active upper limb exercises for patients recovered from neurological diseases, i.e. stroke, Cozens [40] reported a pilot study of using torque attached to an individual joint, combined with EMG measurement that indicated the pattern of arm movement in exercises. Evidence depicted that greater assistance tended to be given to patients with more limited exercises capacity. However, this work was only able to demonstrate the principle of assisting single limb exercise using 2-D based technique. Therefore, a real system was expected to be developed for realistic therapeutic exercises, which may contain "three degrees of freedom at the shoulder and two degrees of freedom at the elbow".

Figure 30. The MANUS in MIT.

6.2.2 MANUS

To find out whether exercise therapy influences plasticity and recovery of the brain following a stroke, a tool is demanded to control the amount of therapy delievered to a patient, where appropriate, objectively measuring the patient's performance. In other words, a system is required to "move smoothly and rapidly to comply with the patients' actions" [73]. Furthermore, abnormally low or high muscle tone may misguide a therapy expert to apply wrong forces to achieve the desired motion of limb segments. To address these problems, a novel automatic system, named MIT-MANUS (Figure 30), was designed to move, guide, or perturb the movement of a patient's upper limb, whilst recording motion-related quantities, e.g. position, velocity, or forces applied [73] (Figure 31). The experimental results were so promising that the commercializing of the established system were under construction. However, it was described that the biological basis of recovery and individual patients' needs should be further studied in order to improve the performance of the system in different circumstances. These findings were also justified in [72].

Figure 31. Image courtesy of Krebs, H.I.

6.2.3 Taylor and improved systems

Taylor [104] described an initial investigation where a simple two DOFs arm support was built to allow movements of shoulder and elbow in a horizontal plane. Based on this simple device, he then suggested a five exoskeletal system to allow activities of daily living (ADL) to be performed in a natural way. The design was validated by tests which showed that “configuration interfaces properly with the human arm”, resulting in the trivial addition of goniometric measurment sensors for identification of arm position and pose.

Another good example was shown in [89], where a device was designed to assist elbow movements. This elbow exerciser was strapped to a lever, which rotated in a horizontal plane. A servomotor driven through a current amplifier was applied to drive the lever, where a potentiometer indicated the position of the motor. Obtaining the position of the lever was achieved by using a semi-circular array of light emitting diodes (LEDs) around the lever. However, this system required a physiotherapist to activate the arm movement and to use a force handle to measure forces applied. This system was meanless to patients as realistic physiotherapy exercises normally occur in three dimensions. As a suggestion, a three DOF prototype was rather advised.

To cope with the problem arisen from individuals with spinal cord injuries Harwin and Rahman [56] explored the design of a head controlled force-reflecting master-slave telemanipulators for rehabilitation applications. This approach was further expanded for a similar class of assistive devices that may support and move the person’s arm in a programmed way. Enclosed within the system, a test-bed power assisted orthosis consisted of a six DOF master with the end effector replaced by a six axis force/torque sensor. A splint assembly was mounted on the force torque sensor and supported the person’s arm. The base level control system first substract the weight of the person’s arm from the whole measurement. Control algorithms were established to relate the estimation of the patient’s residual force to system position, velocity and acceleration [101]. These characteristic parameters are desired in regular movement analysis. Similar to this technique, Chen et al. [36] provided a comprehensive justification for their proposal and testing protocols.

6.2.4 MIME

Burgar et al. [33] and [76] summarized systems for post-stroke therapy conducted at the Department of Veterans Affairs Palo Alto in collaboration with Stanford University. The original principle had been established with two or three DOF elbow/forearm manipulators. Amongst these systems, the MIME shown in Figure 32 was more attractive due to its ability of fully supporting the limb during 3-D movements, and self-guided modes of therapy. Subjects were seated in a wheelchair close to an adjustable height table. A PUMA-560 automation was mounted beside the table that was attached to a wrist-forearm orthosis (splint) via six-axis force transducer. These position digitizer quntified movement kinematics. Clinical trials justified that the better improvements occurred in the elbow measures by the biomechanical measures than the clinical ones. The disvantage of this system is that it could not allow the subject to freely move his/her body.

6.2.5 ARM-Guide

A rehabilitator namely the “ARM guide” [88] was presented to diagnose and treat arm movement impairment following stroke and other brain injuries. Some vital motor impairment, such as abnormal tone, incoordination, and weakness, could

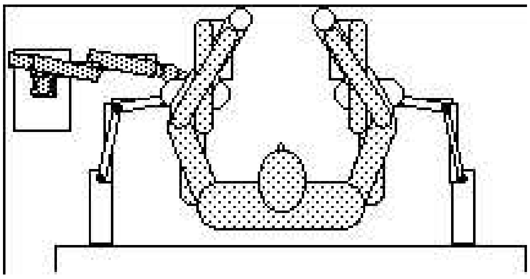


Figure 32. The MIME in MIT.

be evaluated. Pre-clinical results showed that this therapy produced quantifiable benefits in the chronic hemiparetic arm. In the design, the subject's forearm was strapped to a specially designed splint that "slides along the linear constraint". A motor drove a chain drive attached to the splint. An optical encoder mounted on the motor indicated the arm position. The forces produced by the arm were measured by a 6-axis load cell addressing between the splint and the linear constraint. The system needs to be further developed in efficacy and practicality although it achieved a great success.

6.3 Other relevant techniques

Although the following example might not be relevant to the arm training systems, it still provides some hints for constructing a motion tracking system. Hesse and Uhlenbrock [58] introduced a newly developed gait trainer allowing wheelchair-bound subjects to take repitive practice of a gait-like movement without overstressing therapists. It consisted of two footplates positioned on two bars, two rockers, and two cranks that provided the propulsion. The system generated a different movement of the tip and of the rear of the footplate during the swing. Else, the crank propulsion was controlled by a planetary symtem to provide a ratio of 60 percent to 40 percent between stance and swing pahses. Two cases of non-ambulatory patients who regained their walking ability after 4 weeks of daily training on the gait trainer were positively reported.

A number of projects have been undertaken for human arm trajectories. However, to make the proposed systems feasible to non-trained users, further studies need to be performed for the development of a patient interface and therapist workspace. For example, to improve the performance of haptic interfaces, many researchers exhibited their successful prototype systems, e.g. [19], [57]. Hawkins et al. [57] set up an experimental apparatus consisting of a frame with one chair, a wrist connection mechanism, two embedded computers, a large computer screen, and exercise table, a keypad and a 3 DOF haptic interface arm. The user "was seated on the chair with their wrist connected the haptic interface through the wrist connection mechanism. The device end-effector consisted of a gimbal wich provides an extra three DOF to facilitate wrist movement." These tests encourage a novel system to be explored so that a patient can move his/her arm consistantly, smoothly, and correctly. Also, a friendly and human-like interface between the system and the user can be obtained afterwards.

Comprehensive reviews on rehabilitation systems are given in the literature [42] and [43].

7 Discussion

7.1 Remaining challenges

The characters of the previous tracking systems have been summarized earlier. It is demanding to understand the key problems addressed in these systems. Identifying the remaining challenges in the previous systems allows people to specify the aims of further development in the future work.

All the previous systems required therapists to attend during training courses. Without the help of a therapist, any of these systems either was unable to run successfully or just lost controlling commands. The developed systems performed as supervised machines that simply followed orders from the on-site therapists. Therefore, they did not feasibly achieve patient-guided manipulation therapy so they can not be directly used in homes yet.

The second challenge is cost. People intended to build up complicated systems in order to achieve multi-purposes. This leads to expensive components applied to the designed systems. Some of these systems also consisted of particularly designed movement sensors, which limit the further development and broad application of the designed systems.

Inconvenience is another obvious challenge addressed in the previous systems. Most systems demanded people sit in front of a table or in a chair. This configuration constrains people in mobility so they are not helpful at enhancing the overall training of human limbs.

Due to the large working space requested for these systems patients had to prepare spacious recovery rooms for setting up these syetems. As a consequence, this prevents people, who have less accommodation space, from using such systems to regain their mobility. Alternatively, a telemetric and compact system coping with the space problem shall be instead proposed.

Poor performance of human-computer interface (HCI) designed for these syetems has been recognized. Unfortunately, people seldom touch this issue as the other main technical problems had not been solved yet. However, a bad HCI might stop post-stroke patients actively using any training system.

Generally speaking, when one considers a recovery system, such six issues need to be taken into account: cost, size, weight, functional performance, easy operation, and automation.

7.2 Design specification for a proposed system

Consider a system that looks at the limb rehabilitation training for the stroke-suffered patients during their recovery. The designer has to be mainly concerned with such specified issues as follows:

Real time operation of the tracking system is required in order that arm movement can be recorded simultaneously;

Human movement must not be limited in a particular workspace so telemetry is considered for transmitting data from the mounted sensors to the workstation;

The proposed system shall not bring any cumbersome tasks to a user;

Human movement parameters shall be properly and accurately represented in the computer terminal;

A friendly graphical interface between the system and the user is vital due to its application in home-based situations.

The whole system needs to be flexibly attached or installed in a domestic site.

8 Conclusions

A number of applications have already been developed to support various health and social care delivery. It has been justified that the rehabilitation systems are able to assist or replace face to face therapy. Unfortunately, evidence also shows that human movement has a very complicated physiological nature, which prevents futher development of the existing systems. People hence need to have an insight into the formulation of human movement. Our proposed project will cope with this technical issue by attempting to grasp human motion at each moment. Achieving such an accurate localization of the arm may lead to efficient, convenient and cheap kinetic and kinematic modelling for movement analysis.

Acknowledgements

We are grateful for the provision of partial literature sources from Miss Nargis Islam in the University of Bath, and Dr Huiru Zheng in the University of Ulster.

References

[1] In *http://www.polyu.edu.hk/cga/faq/*.
[2] In *http://www.microstrain.com/*.
[3] In *http://www.polhemus.com/*.
[4] In *http://home.flash.net/ hpm/*.
[5] In *http://www.vrealities.com/cyber.html*.
[6] In *http://www.5dt.com/*.
[7] In *http://www.fakespace.com*.
[8] In *http://www.exos.com*.
[9] In *http://www.qualisys.se/*.
[10] In *http://www.charndyn.com/*.
[11] In *http://www.ascension-tech.com/*.
[12] In *http://www.arielnet.com/*.

[13] In *http://www.ndigital.com/polaris.php*.

[14] In *http://www.ptiphoenix.com/*.

[15] In *http://www.simtechniques.com/*.

[16] In *http://www.peakperform.com/*.

[17] J. Aggarwal and Q. Cai. Human motion analysis: A review. *Computer Vision and Image Understanding: CVIU*, 73(3):428–440, 1999.

[18] K. Akita. Image sequence analysis of real world human motion. *Pattern Recognition*, 17:73–83, 1984.

[19] F. Amirabdollahian, R. Louerio, and W. Harwin. A case study on the effects of a haptic interface on human arm movements with implications for rehabilitation robotics. In *Proc. of the 1st Cambridge Workshop on Universal Access an d Assistive Technology (CWUAAT)*, 25-27th March, University of Cambridge 2002.

[20] C. Barron and I. Kakadiaris. A convex penalty method for optical human motion tracking. In *IWVS'03*, volume Nov, Berkeley, pages 1–10, 2003.

[21] A. Baumberg and D. Hogg. An efficient method for contour tracking using active shape models. In *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pages 194–199, 1994.

[22] A. Baumberg and D. Hogg. Generating spatiotemporal models from examples. *IVC*, 14:525–532, 1996.

[23] A. Bharatkumar, K. Daigle, M. Pandy, Q. Cai, and J. Aggarwal. Lower limb kinematics of human walking with the medial axis transformation. In *Proc. of IEEE Workshop on Non-Rigid Motion*, pages 70–76, 1994.

[24] D. Bhatnagar. Position trackers for head mounted display systems. In *Technical Report TR93-010, Deapartment of Computer Sciences, University of North Carolina*, 1993.

[25] M. Black, Y. Yaccob, A. Jepson, and D. Fleet. Learning parameterized models of image motion. In *Proc. of CVPR*, pages 561–567, 1997.

[26] A. Blake, R. Curwen, and A. Zisserman. A framework for spatio-temporal control in the tracking of visual contour. *Int. J. Computer Vision*, pages 127–145, 1993.

[27] R. Bodor, B. Jackson, O. Masoud, and N. Ppanikolopoulos. Image-based reconstruction for view-independent human motion recognition. In *Int. Conf. on Intel. Robots and Sys.*, 27-31 Oct 2003.

[28] C. Bouten, K. Koekkoek, M. Verduim, R. Kodde, and J. Janssen. A triaxial accelerometer and portable processing unit for the assessment daily physical activity. *IEEE Trans. on Biomedical Eng.*, 44(3):136–147, 1997.

[29] R. Bowden, T. Mitchell, and M. Sarhadi. Reconstructing 3d pose and motion from a single camera view. In *BMVC*, pages 904–913, Southampton 1998.

[30] G. Bradski and J. Davies. Motion segmentation and pose recognition with motion history gradient. 13:174–184, 2002.

[31] T. Brosnihan, A. Pisano, and R. Howe. Surface micromachined angular accelerometer with force feedback. In *Digest ASME Inter. Conf. and Exp.*, Nov 1995.

[32] S. Bryson. Virtual reality hardware. In *Implementating Virtual Reality, ACM SIGGRAPH 93*, pages 1.3.16–1.3.24, New York 1993.

[33] C. Burgar, P. Lum, P. Shor, and H. Machiel Van der Loos. Development of robots for rehabilitation therapy: The palo alto va/stanford experience. *Journal of Rehab. Res. and Devlop.*, 37(6):663–673, 2000.

[34] Q. Cai and J. K. Aggarwal. Tracking human motion using multiple cameras. In *ICPR96*, pages 68–72, 1996.

[35] J. Cauraugh and S. kim. Two coupled motor recovery protocols are better than one electromyogram-triggered neuromuscular stimulation and bilateral movements. *Stroke*, 33:1589–1594, 2002.

[36] S. Chen, T. Rahman, and W. Harwin. Performance statistics of a head-operated force-reflecting rehabilitation robot system. *IEEE Trans. on Rehab. Eng.*, 6:406–414, Dec 1998.

[37] Z. Chen and H. J. Lee. Knowledge-guided visual perception of 3d human gait from a single image sequence. *IEEE Trans. On Systems, Man, and Cybernetics*, pages 336–342, 1992.

[38] J. Chung and N. Ohnishi. Cue circles: Image feature for measuring 3-d motion of articulated objects using sequential image pair. In *AFGR98*, pages 474–479, 1998.

[39] M. Cordea, E. Petriu, N. Georganas, D. Petriu, and T. Whalen. Real-time 21/2d head pose recovery for model-based video-coding. In *IEEE Instrum. and Measure. Tech. Conf.*, Baltimore, MD, May 2000.

[40] J. Cozens. Robotic assistance of an active upper limb exercise in neurologically impaired

patients. *IEEE Transactions on Rehabilitation Engineering*, 7(2):254–256, 1999.

[41] R. Cutler and M. Turk. View-based interpretation of real-time optical flow for gesture recognition. In *Int. Conf. on Auto. Face and Gest. Recog.*, pages 416–421, 1998.

[42] J. Dallaway, R. Jackson, and P. Timmers. Rehabilitation robotics in europe. *IEEE Trans. on Rehab. Eng.*, 3:35–45, 1995.

[43] K. Dautenhahn and I. Werry. Issues of robot-human interaction dynamics in the rehabilitation of children with autism. In *Proc. of FROM ANIMALS TO ANIMATS, The Sixth International Conference on the Simulation of Adaptive Behavior (SAB2000)*, 11-15 Sep, Paris 2000.

[44] S. Dockstader, M. Berg, and A. Tekalp. Stochastic kinematic modeling and feature extraction for gait analysis. *IEEE Transactions on Image Processing*, 12(8):962–976, Aug 2003.

[45] I. Dukes. Compact motion tracking system for human movement. *MSc dissertation, University of Essex*, Sep 2003.

[46] F. Escolano, M. Cazorla, D. Gallardo, and R. Rizo. Deformable templates for tracking and analysis of intravascular ultrasound sequences. In *1st International Workshop of Energy Minimization Methods in CVPR*, Venecia, Mayo, 1997.

[47] R. Fablet and M. J. Black. Automatic detection and tracking of human motion with a view-based representation. In *ECCV02*, pages 476–491, 2002.

[48] H. Feys, W. De Weerdt, B. Selz, C. Steck, R. Spichiger, L. Vereeck, K. Putman, and G. Van Hoydonck. Effect of a therapeutic intervention for the hemiplegic upper limb in the acute phase after stroke: a single-blind, randomized, controlled multicenter trial. *Stroke*, 29:785–792, 1998.

[49] W. Freeman, K. Tanaka, J. Ohta, and K. Kyuma. Computer vision for computer games. In *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pages 100–105, 1996.

[50] H. Fujiyoshi and A. Lipton. Real-time human motion analysis by image skeletonisation. In *Proc. of the Workshop on Application of Computer Vision*, Oct. 1998.

[51] D. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding: CVIU*, 73(1):82–98, 1999.

[52] L. Goncalves, E. Bernardo, E. Ursella, and P. Perona. Monocular tracking of the human arm in 3d. In *ICCV95*, pages 764–770, 1995.

[53] J. Gonzalez, I. Lim, P. Fua, and D. Thalmann. Robust tracking and segmentation of human motion in an image sequence. In *ICASSP03*, April, Hong Kong 2003.

[54] R. Green. Spatial and temporal segmentation of continuous human motion from monocular video images. In *Proc. of Image and Vision Computing New Zealand*, pages 163–169, 2003.

[55] G. Haisong, Y. Shiral, and M. Asada. Mdl-based segmentation and motion modeling in a long image sequence of scene with multiple independently moving objects. 18:58–64, 1996.

[56] W. Harwin and T. Rahman. Analysis of force-reflecting telerobotics systems for rebalitation applications. In *Proc. of the 1st European Conf. on Dis., Virt. Real. and Assoc. Tech.*, pages 171–178, Maidenhead 1996.

[57] P. Hawkins, J. Smith, S. Alcock, M. Topping, W. Harwin, R. Loureiro, F. Amirabdollahian, J. Brooker, S. Coote, E. Stokes, G. Johnson, P. Mark, C. Collin, and B. Driessen. Gentle/s project: design and ergonomics of a stroke rehabilitation system. In *Proc. of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT)*, 25-27th March, University of Cambridge 2002.

[58] S. Hesse and D. Uhlenbrock. A mechanized gait trainer for restoration of gait. *Journal of Rehab. Res. and Devlop.*, 37(6):701–708, 2000.

[59] A. Hilti, I. Nourbakhsh, B. Jensen, and R. Siegwart. Narrative-level visual interpretation of human motion for human-robot interaction. In *Proceedings of IROS 2001. Maui, Hawaii*, 2001.

[60] D. Hogg. Model-based vision: A program to see a walking person. *Image and Vision Computing*, 1:5–20, 1983.

[61] T. Horprasert, I. Haritaoglu, D. Harwood, L. Davies, C. Wren, and A. Pentland. Real-time 3d motion capture. In *PUI Workshop98*, 1998.

[62] http://bmj.bmjjournals.com/cgi/reprint/325/.

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

[64] http://www.xsens.com/.

[65] E. Huber. 3d real-time gesture recognition using proximity space. In *Proc. of Intl. Conf. on Pattern Recognition*, pages 136–141, August 1996.

[66] M. Isard and A. Blake. Contour tracking by stochastic propagation of conditional density. In *ECCV*, pages 343–356, 1996.

[67] M. Ivana, M. Trivedi, E. Hunter, and P. Cosman. Human body model acquisition and tracking using voxel data. *Int. J. of Comp. Vis.*, 53(3):199–223, 2003.

[68] O. Javed, S. Khan, Z. Rasheed, and M. Shah. Camera handoff: tracking in multiple uncalibrated stationary cameras. In *IEEE Workshop on Human Motion, HUMO-2000*, pages Austin, TX, 2000.

[69] G. Johansson. Visual motion perception. *Scientific American*, 232:76–88, 1975.

[70] K. Kahol, P. Tripathi, and S. Panchanathan. Gesture segmentation in complex motion sequences. In *ICIP*, pages Barcelona, Spain, 2003.

[71] A. Kourepenis, A. Petrovich, and M. Meinberg. Development of a monotithic quartz resonator accelerometer. In *Proc. of 14th Biennial Guidance Test Symp., Hollman AFB, NM*, 2-4 Oct 1989.

[72] H. Krebs, N. Hogan, M. Aisen, and B. Volpe. Robot-aided neurorehabilitation. *IEEE Transactions on Rehabilitation Engineering*, 6(1):75–87, Mar 1998.

[73] H. Krebs, B. Volpe, M. Aisen, and N. Hogan. Increasing productivity and quality of care: robot-aided nero-rehabilitation. *Journal of Rehabilitation Research and Development*, 37(6):639–652, November/December 2000.

[74] S. Kurakaka and R. Nevatia. Description and tracking of moving articulated objects. In *ICPR92*, pages 491–495, 1992.

[75] W. Long and Y. Yang. Log-tracker: An attribute-based approach to tracking human body motion. *PRAI*, 5:439–458, 1991.

[76] P. Lum, D. Reinkensmeyer, R. Mahoney, W. Rymer, and C. Burgar. Robotic devices for movement therapy after stroke: current status and challenges to clinical acceptance. *Top Stroke Rehabil.*, 8(4):40–53, 2002.

[77] M. Machline, L. Zelnik-Manor, and M. Irani. Multi-body segmentation: Revisiting motion consistency. In *International Workshop on Vision and Modeling of Dynamic Scenes (with ECCV02)*, 2002.

[78] D. Marr and K. Nishihara. Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London*, 200:269–294, 1978.

[79] T. Moeslund and E. Granum. Multiple cues used in model-based human motion capture. In *FG'00*, pages Grenoble, France, 2002.

[80] A. Mulder. Human movement tracking technology. In *Technical Report 94-1, Simon Fraser University*, 1994.

[81] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in xyt. In *CVPR*, pages 469–474, 1994.

[82] J. O'Rourke and N. Badler. Model based image analysis of human motion using constraint propagation. *PAMI*, 2:522–536, 1980.

[83] X. Pennec, P. Cachier, and N. Ayache. Tracking brain deformations in time sequences of 3D US images. In *Proc. of IPMI'01*, pages 169–175, 2001.

[84] R. Polana and R. Nelson. Low level recognition of human motion. In *Proc. of Workshop on Non-rigid Motion*, pages 77–82, 1994.

[85] C. Poynton. *A technical introducton to digital vedio*. New York: Wiley, 1996.

[86] R. Qian and T. Huang. Estimating articulated motion by decomposition. In *Time-Varying Image Processing and Moving Object Recognition, 3-V. Cappellini (Ed.)*, pages 275–286, 1994.

[87] J. Rehg and T. Kanade. Model-based tracking of self-occluding articulated objects. In *ICCV*, pages 612–617, 1995.

[88] D. Reinkensmeyer, L. Kahn, M. Averbuch, A. McKenna-Cole, B. Schmit, and W. Rymer. Understanding and treating arm movement impairment after chronic brain injury: Progress with the arm guide. *Journal of Rehab. Res. and Devlop.*, 37(6):653–662, 2000.

[89] R. Richardson, M. Austin, and A. Plummer. Development of a physiotherapy robot. In *Proc. of the Intern. Biomech. Workshop, Enshede*, pages 116–120, April 1999.

[90] M. Ringer and J. Lasenby. Modelling and tracking articulated motion from multiple camera views. In *BMVC*, Sep, Bristol, UK 2000.

[91] A. Roche, X. Pennec, G. Malandain, and N. Ayache. Rigid registration of 3D ultrasound with MR images: a new approach combining intensity and gradient information. *IEEE Transactions on Medical Imaging*, 20(10):1038–1049, oct 2001.

[92] K. Rohr. Toward model-based recognition of human movements in image sequences. *Computer Vis., Graphics Image Process.*, 59:94–115, 1994.

[93] R. Ronfard, C. Schmid, and B. Triggs. Learning to parse pictures of people. In *European Conference on Computer Vision, LNCS 2553*, volume 4, pages 700–714, June 2002.

[94] K. Sato, T. Maeda, H. Kato, and S. Inokuchi. Cad-based object tracking with distributed monocular camera for security monitoring. In *Proc. 2nd CAD-Based Vision Workshop*, pages 291–297, 1994.

[95] N. Shimada, K. Kimura, Y. Shirai, and Y. Kuno. Hand posture estimation by combining 2-d appearance-based and 3-d model-based approaches. In *ICPR00*, 2000.

[96] H. Sidenbladh, M. J. Black, and D. Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *ECCV*, pages 702–718, Dublin, June 2000.

[97] H. Sidenbladh, M. J. Black, and L. Sigal. Implicit probabilistic models of human motion for synthesis and tracking. In *ECCV*, 2002.

[98] C. Sminchisescu and B. Triggs. Covariance scaled sampling for monocular 3d body tracking. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 447–454, Kauai, Hawaii 2001.

[99] Y. Song, X. Feng, and P. Perona. Towards detection of human motion. In *CVPR*, pages 810–817, 2000.

[100] M. Storring, H. Andersen, and E. Granum. Skin colour detection under chaning lighting conditions. In *Proc. of 7th Symposium on Intelligence Robotics System*, 1999.

[101] S. Stroud. A force controlled external powered arm orthosis. *Masters Thesis*, University of Delaware 1995.

[102] D. Sturman and D. Zeltzer. A survey of glove-based input. *IEEE Computer Graphics and Aplications*, pages 30–39, 1994.

[103] Y. Tao and H. Hu. Building a visual tracking system for home-based rehabilitation. In *Proc. of the 9th Chinese Automation and Computing Society Conf. In the UK*, England, Sep 2003.

[104] A. Taylor. Design of an exoskeletal arm for use in long term stroke rehabilitation. In *ICORR'97, University of Bath*, April 1997.

[105] C. Theobalt, M. Magnor, P. Schueler, and H. Seidel. Combining 2d feature tracking and volume reconstruction for online video-based human motion capture. In *Proceedings of Pacific Graphics 2002*, pages 96–103, Beijing 2002.

[106] M. Uiterwaal, E. Glerum, H. Busser, and R. Van Lummel. Ambulatory monitoring of physical activity in working situations, a validation study. *Journal of Medical Engineering & Technology*, 22:168–172, July/August 1998.

[107] P. Veltink, H. Bussmann, W. de Vries, W. Martens, and R. Van Lummel. Detection of static and dynamic activities using uniaxial accelerometers. *IEEE Transactions on Rehabilitation Engineering*, 4(4):375–385, Dec 1996.

[108] G. Welch and E. Foxlin. Motion tracking survey. *IEEE Computer Graphics and Applications*, pages 24–38, Nov/Dec 2002.

[109] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.

[110] H. Xie and G. Fedder. A cmos z-axis capacitive accelerometer with comb-finger sensing. In *Technical Report, The Robotics Institute, Carnegie Mellon University*, 2000.

[111] Y. Yaccob and M. Black. Parameterized modeling and recognition of activities. In *ICCV*, pages 120–127, 1998.

[112] Y. Yacoob and L. Davies. Learned models for estimation of rigid and articulated human motion from stationary or moving camera. *Int. Journal of Computer Vision*, 36(1):5–30, 2000.

[113] M. Yeasin and S. Chaudhuri. Development of an automated image processing system for kinematic analysis of human gait. *Real-Time Imaging*, 6:55–67, 2000.

[114] M. Yeasin and S. Chaudhuri. Towards automatic robot program: Learning human skill from perceptual data. *IEEE Trans. on Systems Man and Cybernetics-B*, 30(1):180–185, 2000.

[115] M. Yeasin and S. Chaudhuri. Visual understanding of dynamic hand gestures. *Pattern Recognition*, 33(11):1805–1817, 2000.

[116] L. Zelnik-Manor, M. Machline, and M. Irani. Multi-body segmentation: Revisiting motion consistency. In *International Workshop on Vision and Modeling of Dynamic Scenes (with ECCV02)*, 2002.

[117] T. Zimmerman and J. Lanier. Computer data entry and manipulation apparatus method. In *Patent Application 5,026,930*, VPL Research Inc. 1992.

[118] A. Zisserman, L. Shapiro, and M. Brady. 3d motion recovery via affine epipolar geometry. *Int. J. of Comput. Vis.*, 16:147–182, 1995.