A Digital Video System for the Automated Measurement of Repetitive Joint Motion

ChunMei Lu and Nicola J. Ferrier, *Member, IEEE*

***Abstract*—Automated measurement and analysis of human motion during performance of workplace tasks are desirable for ergonomic studies. While numerous technologies exist for accurate measurement of biomechanical data, their use is often not feasible in the *workplace* environment. We present a digital-video based system suitable for measuring human motion of repetitive workplace tasks. Due to practical considerations, a single-camera solution is exploited by adding some control over the environment. We present an analysis of experiments demonstrating the accuracy of our system.**

***Index Terms*—Ergonomics, joint motion, telemonitoring, video analysis.**

I. Introduction

QUANTIFICATION of the relationship between posture, repetitiveness, and risk of injury while performing a workspace task requires accurate measurement of human motion. Other medical and entertainment applications also require highly accurate measurement of human motion. Most of the existing technologies for measuring human motion use specialized hardware, often making the process invasive (e.g., mechanical armatures), difficult to operate (e.g., electromagnetic trackers which are sensitive to their environment), and expensive (e.g., multiple synchronized high-speed infrared cameras). For example, electromagnetic tracking devices function by measuring the strength of the magnetic fields generated by transmitters. This technology does not suffer from occlusion and can offer real-time performance, so it has been successful for animation applications [9] where its cost, small working volumes, and sensitivity to the environment do not preclude its use. However, the sensitivity to the environment makes this technology impractical for the application considered here, i.e., workplace task analysis. Similarly, electrogoniometers are widely used for human motion acquisition in ergonomic assessment. They are relatively expensive and their precision is easily perturbed by metallic objects and other magnetic fields. Another example system, armatures, use a light-weight mechanical linkage with encoders placed at joints to measure the change in position and orientation with respect to a reference point. Such mechanical devices are obtrusive and potentially cumbersome to wear and will interfere with worker's motions. Additionally, the above data acquisition systems lack the ability to link the data with the visual activities and this makes the postprocessing (data extraction and reduction) hard. Simultaneous video acquisition of the activity being performed is used [1], [8], [11], [14], [17], [20], [21] and trained human analysts review the tape to identify specific activities. Manual identification of events in the video is generally tedious and time consuming. Using an efficient multimedia video annotation system required an average of 6.8 min for an analyst to analyze 1 min of tape [22].

The desirable properties for a human motion measurement system for our risk assessment application include the following: 1) it is practical for "in the field" use; 2) it requires minimal interference with the task; 3) it is accurate; 4) it is capable of handling complex human motion; 5) it is able to cope with cycle times on the order of seconds up to tens of minutes; 6) it is affordable; and, ideally, 7) it requires no human intervention. Digital video can potentially meet these needs: It is noninvasive (does not interfere with the subject), easy to operate, mobile, and inexpensive. Because the motion data and video information ("events") are both extracted from the same medium, synchronization of analog sensors and video is avoided. Data can be recorded at 30 frames/s for any length (given sufficient video storage capabilities). Of primary concern when using video measurement is the accuracy of the recovered motion and the complexity of motion that can be measured.

In the remainder of this paper, we present a brief overview of a simple digital video system and our evaluation of its performance. The accuracy limitations of the system are explored: We model the error sources to obtain bounds, then demonstrate experimentally that our error modeling holds in practice. We exploit the ability to permit minor workplace modifications, i.e., the subject must wear special clothing and/or markers, and we require the camera person to film in positions to ensure optimal viewing of a specific motion. These assumptions simplify the image analysis task and can be relaxed as superior tracking algorithms are developed. Using motion data obtained with the system described here, we have developed techniques for automated analysis of the video data [6], [7] and for the detection and classification of motions [7].

II. Recovery of Joint Motion From Video

Ergonomics practitioners require information on subject posture, typically recorded as joint angles. Many studies have considered only a single joint (e.g., [15] and [16]); however, multijoint motion is desired for many studies of workplace activities. Video data must be processed to obtain joint angle information

Manuscript received November 19, 2001; revised September 17, 2003 and May 3, 2004. This work was supported by the National Science Foundation (IRI-9703352) and by a University of Michigan Center for Occupational Health and Safety Engineering Pilot Research Award.

C. Lu is with the Center for Mathematical Sciences, University of Wisconsin, Madison, WI 53706 USA.

N. J. Ferrier is with the Departments of Mechanical Engineering and Biomedical Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: ferrier@engr.wisc.edu).

Digital Object Identifier 10.1109/TITB.2004.832544

Fig. 1. Example of the recovered skeleton shown with the corresponding video frame.

from the recorded video data. There has been significant effort in the computer vision community in developing systems to track humans (see, e.g., [10]), thus, the focus of this paper is *not* on improved tracking of humans, but on the design of a system for automated assessment of human motion, assuming a tracking system exists.

One method of extracting joint angles from video data is by reconstructing the subject pose, then extracting joint information from the pose. Multiple cameras (stereo or trinocular systems), or structure from motion techniques can provide three-dimensional (3-D) information. Because of the pragmatic difficulties in precisely synchronizing multiple cameras and due to constraints imposed by our application, practitioners felt that multi-camera systems were not practical. Restriction to a single video camera means that the motion data from our system is inherently two-dimensional and occlusion is possible. For the factory setting, we facilitated the task by requiring 1) the camera operator to film in such a way that the image plane is roughly parallel to the plane of motion of the links and 2) the subject to wear colored markers.

Colored markers are easily tracked in images. The image marker locations are used to fit a skeletal model. In video systems, the tracked marker locations can be used to recover joint angles and a variety of approaches have been used (e.g., [2]–[23]). For our single-camera system, a skeletal model[1] can be used to recover the human pose from the marker locations. A nonlinear least-squares optimization process to fit the skeleton to the marker data is presented in [9]. Using the skeletal model as a constraint, linear algorithms can be used [3]. When using two or more markers per body segment, the orientation of each segment can be calculated from the corresponding marker data and a global skeleton fitting technique can be applied to recover the joint angle data [18]. Once the skeleton has been constructed, the joint angles are determined trivially from the reconstructed skeleton. A human analyst can verify the reconstructed motion using a video animation of the body segments under observation (Fig. 1). The recovered joint angles in each frame can be saved as a time-indexed motion sequence and used directly by ergonomics practitioners, or, the data can be transformed into representations more appropriate for analysis [6].

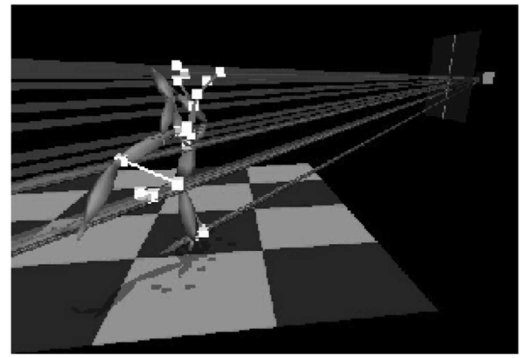

Fig. 2. Skeleton reconstruction based on image observations and link length constraints can lead to false poses. The reconstructed leg on the subject (shown as a white stick figure) takes on the *inverted* pose from the actual configuration (gray figure). This pose is consistent with the observed image marker locations and link lengths.

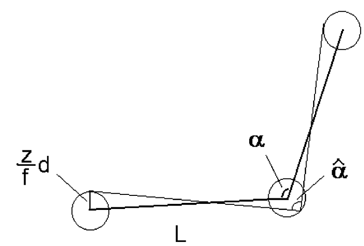

Fig. 3. Because of image quantization and noise, image observations constrain the joint position within a volume and the ideal projection of the estimated joint position is constrained within a circle around it, which gives a joint angle measurement $\hat{\alpha}$ instead of the real angle α. The real joints are depicted by the solid lines.

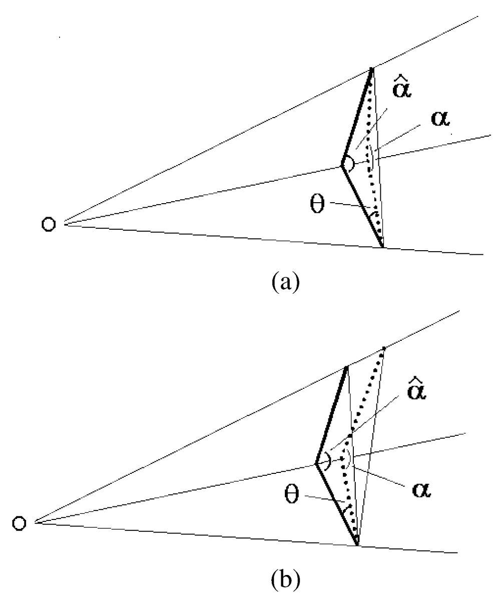

Fig. 4. Two situations when (a) fronto-parallel viewpoint and (b) planar motion assumptions fail. In both figures, α is the true joint angle while $\hat{\alpha}$ is the estimated joint angles.

III. Performance Assessment

To formally analyze the error in the reconstructed motion, one must consider the constraints used in obtaining the motion. These constraints take different forms (constraints on the skeleton structure, i.e., link lengths, constraints on the skeleton pose, i.e., joint limits) making formal quantification of error complex. We have performed experiments to empirically demonstrate accuracy. Experiments were performed both with *known* controlled motion of a serial linkage (a robot), and using a commercially available active marker tracking system. The results show that motion is accurately captured (Section III-B).

[1]In practice, this could be obtained by measurement.

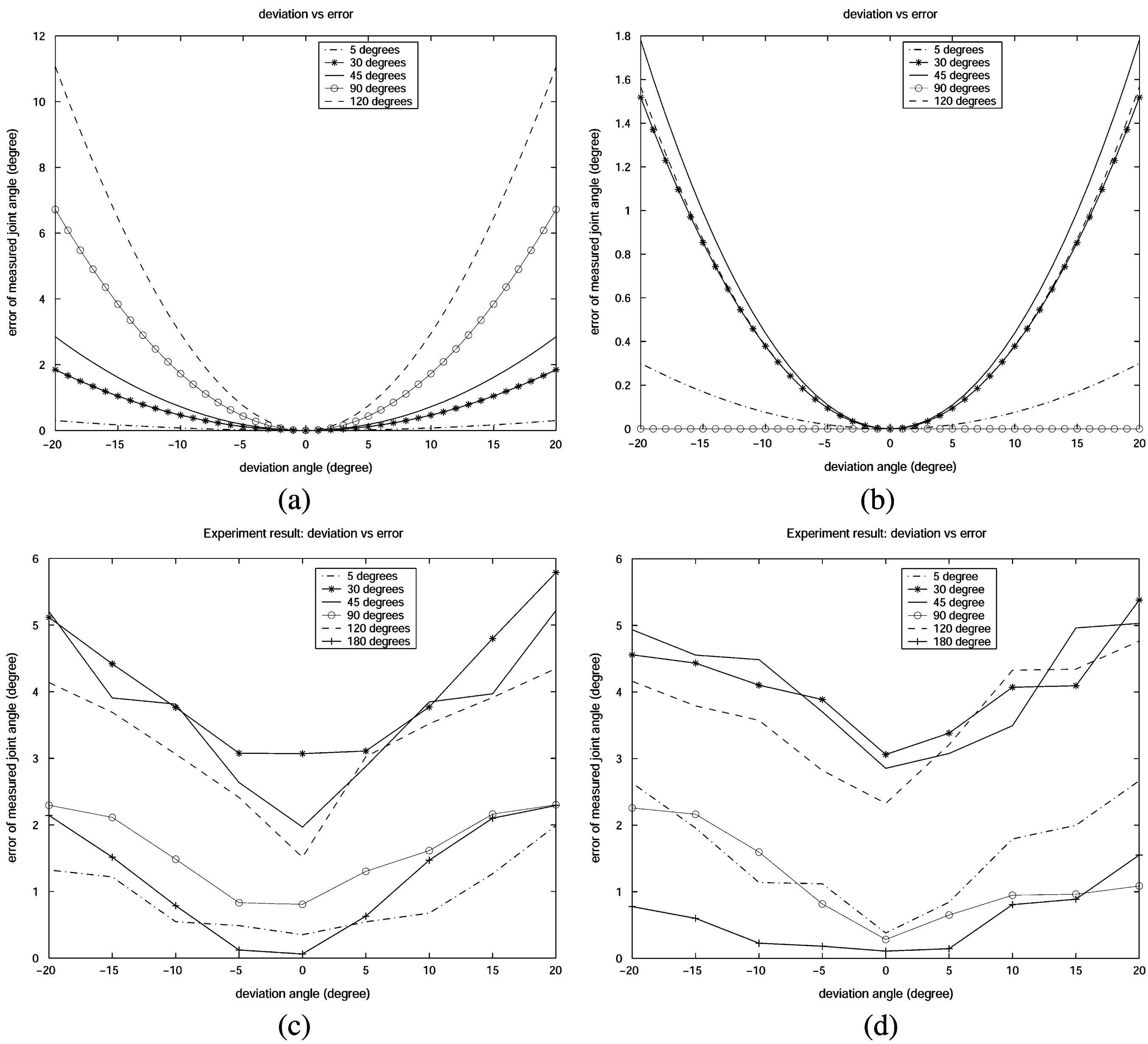

Fig. 5. Plots of the joint angle error versus angle of deviation for different joint angles. Theoretical lower bounds for deviation from (a) planar and (b) fronto-parallel view and, experimentally measured errors using a robot arm for deviation from (c) planar and (d) fronto-parallel view.

A. Accuracy

Some marker configurations and link length constraints can lead to multiple interpretations of the 3-D position of the joint. In other work [4], we have demonstrated that 3-D reconstruction based on noisy image observations and link length constraints can lead to false poses (see Fig. 2). Because erroneous poses are consistent with the data, such errors can only be detected if one has additional knowledge. Our current solution requires the camera operator to position the camera roughly fronto-parallel to the plane of motion of the joint under consideration (limiting the likely poses). Thus, errors in our reconstructed joint angles primarily come from errors in computed marker location, assumptions of planar motion with fronto-parallel view, and skeleton reconstruction error. Below we derive bounds on the first two error sources and, through comparison with an active 3-D marker tracking system, we empirically assess the third source of error.

1) Image Error: We assume an ideal pin-hole projection parameterized by the focal distance f and the principal point. When observing joints moving parallel to the image plane, the error in reconstructed joint angle depends on the position of the joint, the distance between the image plane and the plane of motion, the link length, and the camera parameters. Fig. 3 schematically represents a joint and its projection in the image. Due to tracking and/or finite resolution error, the image observation of the marker (here assumed to be the endpoints of the link) constrains the "true" location. If the image error is given by d (i.e., the distance between the actual projection and measured image location is within this distance), the actual joint angle is α, the estimated joint angle is $\hat{\alpha}$, the link lengths are L, and the distance to the plane of motion is Z, then using simple geometry, it

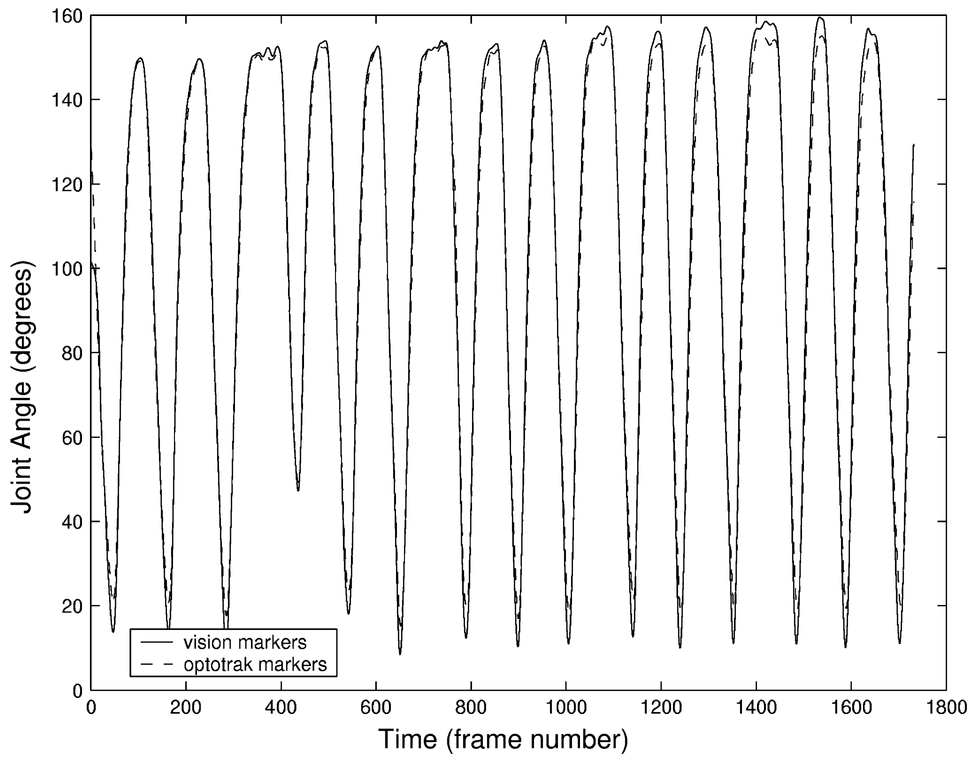

Fig. 6. Comparison of our measured joint angles versus angles recovered using the OptoTrak IR active marker tracking system for a reaching task.

can be shown that the error in the joint angle estimate is bounded by

$$|\alpha - \hat{\alpha}| \leq 2 \arcsin\left(\frac{2dZ}{fL}\right). \tag{1}$$

While this estimate depends on a number of assumptions (pinhole camera model, two equal link lengths, joint centered in the image) which will generally *not* be valid, it does give a lower bound on the error due to imaging. Under reasonable viewing assumptions (viewing a 25-cm link, e.g., a forearm, at a distance of 1.5 m) and assuming 1–2 pixel error in image observations, the angular variation is on the order of 5° or less. This analysis considered only the uncertainty in the image observations. If one considers uncertainty in the optical parameters (e.g., focal length) and in the link lengths, then the region of uncertainty grows.

2) Viewpoint/Motion Assumption Error: When constrained to using a single camera, we required the camera operator to view the motion such that the motion was planar with a fronto-parallel view. In practice, such precise placement of the camera is hard to achieve. The goal is to place the camera such that the depth variation along the link is small relative to the camera distance. This lack of precision will introduce an error into our recovered joint angles due to deviation from the assumptions of planar motion and fronto-parallel view. Fig. 4 depicts the two basic situations where our assumptions fail. In both plots, the solid linkage is shown in a plane that is parallel to the image plane (the ideal). We denote by θ the deviation angle between the fronto-parallel plane and the actual plane of motion (or viewing).

Let $\alpha, \hat{\alpha}$ be the true and estimated joint angles, respectively. In both cases one can derive the relationship between the deviation angle θ (between true and assumed plane) and the joint angles α and $\hat{\alpha}$. For the case in Fig. 4(a), the error relationship is

$$\cos\hat{\alpha} - \cos^2\theta \cos\alpha = 1 - \cos^2\theta \tag{2}$$

and for the case in Fig. 4(b),

$$\frac{\tan\hat{\alpha}}{\tan\alpha} = \cos\theta. \tag{3}$$

These theoretical bounds are plotted in Fig. 5. for various joint angles (where a joint angle of zero represents a link "folded back onto itself"). From Fig. 5(b), assuming the actual joint angle is 45° and the view is within 20° of fronto-parallel, then the measured joint angle error is within 1.8°. An alternate analysis of the nonplanar motion assumption is presented in [12].

To verify the theoretical bounds on the error, we measured the errors experimentally. A robot was used to precisely control the motion of a joint. The robot joint configuration (pose) was varied and the motion was varied from the ideal fronto-parallel/planar motion. While not identical, the plots demonstrate similitude. We caution, however, that human joints are not ideal revolute joints as modeled in the analysis (and as found on the robot subject).

B. Experiments

Our analysis shows that the error in the recovered joint angles due to imprecise camera positioning and image observation noise can be small for many joint configurations. To evaluate the system in practice, we compared our joint angle data with data obtained using an active marker tracking system (OptoTrak [13]) and found the results to be comparable. Fig. 6 shows a comparison of motion from the digital video system with the active marker tracking system. For some rapid motions, the active marker system was unable to track, while our digital video system succeeded.

IV. Evaluation

The system in its current form satisfies the initial criteria.

Practical: The system is easy to use in the field as long as the camera operator can gain a suitable vantage point. Color markers must be selected based on the task environment. Lighting must be sufficient.

Accuracy: Joint angles with errors on the order of $\pm 5^\circ$ can be obtained if the constraints (planar motion and fronto-parallel view) can be maintained. The length of the upper arm and forearm allow the placement of markers and for repetitive workplace tasks, where the frequency and magnitude of movements are of primary importance and the accuracy of the recovered motion will enable recovery of these motion statistics.

Complexity: Complexity of motion is limited by the visual tracking system employed. Here, we have restricted our studies to the upper extremities for seated employees. There are numerous tasks that require the worker to move around. In these situations, there may not be a single-camera viewpoint that is suitable to capture the motion.

Human Effort: Aside from the camera person, our system currently is capable of working without human intervention.

Due to the uncertainties from both skeleton reconstruction and image noise when a joint is connected by shorter linkages, the current system may not be appropriate for collecting motion data on joints such as fingers and wrist.

V. Discussion

We have analyzed the accuracy of a simple digital video system for the acquisition of human motion. The system is efficient, automatic, and noninvasive. The error in recovered motion can be estimated and experiments confirmed the estimated error bounds. The efficiency of the system makes it possible to collect and analyze a large amount of data for the further risk assessment.

The challenge of building a sufficiently accurate system for the automated measurement of human motion using a single-camera system involved tradeoffs between generality, practicality, and accuracy. Video naturally offers the advantages of minimal interference with task, cost effectiveness, and practicality. However, video posed problems in obtaining accurate 3-D geometry. Our solution required minor environmental modifications (markers) and placed constraints on the camera operator to maintain suitable viewpoints to obtain accurate joint motion data.

The constraint imposed on the camera position impacts the utility of this tool in two ways. First, because many motions are not planar, multiple video clips of the same motion may be required to obtain joint angle information for all the limbs under consideration. Thus, some motions cannot be analyzed for all joint motions simultaneously, and statistics must be collected for some joints separately. This is consistent with current practice for manual video annotation (where a reviewer may watch a segment of video repeatedly, and estimates the position of a particular joint on each pass). Second, certain motions cannot be measured under the fronto-parallel viewpoint constraint because the fronto-parallel viewpoint of the joint may move *with* the joint and, hence, a single-camera vantage point cannot capture such a motion. If the camera person does not maintain a fronto-parallel view, then one must deal with the possibility of recovering a plausible but erroneous pose (see Fig. 2). Recent work indicates that plausible poses are typically “far apart” and ambiguities can usually be resolved [19].

The use of a simplistic tracking methods limits application. As our goal was to prove feasibility and develop methods to analyze motion, tracking has not been our focus. Alternate tracking technologies can be readily incorporated into the system, and hence, we do not view this as a impediment to further system development.

Our initial experience demonstrates to practitioners that we are able, in certain situations, to capture human motion during repetitive workplace tasks and automatically convert the motion into joint angle time series data. In order to be truly useful, the current limitations indicated above must be addressed.

While the ability to measure motion for ergonomic studies was our focus, current trends in “telemedicine” may find such systems useful. The low cost of placing a digital recording system in a patient’s home may enable remote monitoring of exercise or other rehabilitation schemes.

References

[1] P. Rosende, C. H. Salvador, and M. A. Gonzalez, “Recording of physiological signals with domestic videotape storage,” *Med. Biol. Eng. Comput.*, vol. 29, pp. 155–158, 1991.

[2] M. J. Dysart and C. Woldstad, “Posture prediction for static sagittal-plane lifting,” *J. Biomechanics*, vol. 29, no. 10, pp. 1393–1397, 1996.

[3] M. Gleicher, Computer animation class notes, 2000.

[4] M. Gleicher and N. J. Ferrier, “Evaluating video based motion capture,” in *Making Motion Capture Useful—Siggraph Course Notes* Los Angeles, CA, Aug. 2001.

[5] M. Kawato, Y. Maeda, Y. Uno, and R. Suzuki, “Trajectory formation of arm movement by cascade neural network model based on minimum torque-change criterion,” *Biol. Cybern.*, vol. 62, pp. 275–288, 1990.

[6] C. M. Lu and N. J. Ferrier, “Automated analysis of repetitive joint motion,” *IEEE Trans. Inform. Technol. Biomed.*, vol. 7, pp. 263–273, Dec. 2003.

[7] ——, “Repetitive motion analysis: Segmentation and event classification,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, pp. 258–263, Feb. 2004.

[8] W. S. Marras and R. W. Schoenmarklin, “Wrist motions industrial,” *Ergonomics*, vol. 36, no. 4, pp. 341–351, 1993.

[9] A. Menache, *Understanding Motion Capture for Computer Animation and Video Games*. San Mateo, CA: Morgan Kaufman, 2000.

[10] T. Moeslund, “Summaries of 107 computer vision-based human motion capture papers,” Lab. for Image Analysis, Univ. Aalborg, Tech. Report LIA 99-01, 1999.

[11] A. Moore, R. Wells, and D. Ranney, “Quantifying exposure in occupational mannual tasks with cumulative trauma disorder potential,” *Ergonomics*, vol. 34, no. 12, pp. 1433–1453, 1991.

[12] D. D. Morris and J. M. Rehg, “Singularity analysis for articulated object tracking,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 1998, pp. 289–296.

[13] Northern Digital Optical Tracking [Online]. Available: http://www.ndigital.com

[14] N. D. Danieley, P. D. Wolf, R. E. Ideker, and W. M. Smith, "A digital tape recorder for electrophysiologic waveforms," in *Proc. 38th Annu. Conf. Engineering in Medicine and Biology*, vol. 124, Chicago, IL, Sept. 30–Oct. 2, 1985.

[15] R. G. Radwin and M. L. Lin, "An analytical method for characterizing repetitive motion and posture stress using spectral analysis," *Ergonomics*, vol. 36, no. 4, pp. 379–389, 1993.

[16] R. G. Radwin, M. L. Lin, and T. Y. Yen, "Exposure assessment of biomechanical stress in repetitive mannual work using frequency-weighted filters," *Ergonomics*, vol. 37, no. 12, pp. 1984–1998, 1994.

[17] R. G. Radwin and T. Y. Yen, "Development of an instrument for assessing exposure to physical stress in repetitive manual work," in *Proc. 1993 NSF Design and Manufacturing Systems Conf.*, 1993, pp. 1151–1158.

[18] M. C. Silaghi, R. Plänkers, R. Boulic, P. Fua, and D. Thalmann, "Local and global skeleton fitting techniques for optical motion capture," in *Modeling and Motion Capture Techniques for Virtual Environments: Lecture Notes in Artificial Intelligence*, N. Magnenat-Thalmann and D. Thalmann, Eds. New York: Springer-Verlag, 1998, vol. 1537, pp. 26–40.

[19] C. Sminchisescu and B. Triggs, "Kinematic jump processes for monocular 3D human tracking," in *Proc. Conf. Computer Vision and Pattern Recognition*, 2003, pp. 69–76.

[20] D. Smith and R. Propst, "Analog recording with a video cassette recorder," *IEEE Trans. Biomed. Eng.*, vol. BME-26, no. 1, pp. 166–168, 1979.

[21] T. Y. Yen and R. G. Radwin, "A video-based system for acquiring biomechanical data synchronized with arbitrary events and activities," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 9, pp. 944–948, Sept. 1995.

[22] T. Yen, "Physical stress measurements for work-related musculoskeletal disorders using video-based biomechanical data acquisition and interactive exposure analysis," Ph.D. Thesis, Dept. Industrial Eng., Univ. Wisconsin-Madison, Madison, WI, 1997.

[23] X. Zhang and D. B. Chaffin, "A three dimensional dynamic posture prediction model for simulating in-vehicle seated reaching movements: Development and validation," *Ergonomics*, vol. 43, no. 9, pp. 1314–1330, 2000.