International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:9, No:3, 2015

Human Motion Capture: New Innovations in the Field of Computer Vision

Najm Alotaibi

***Abstract*—**Human motion capture has become one of the major area of interest in the field of computer vision. Some of the major application areas that have been rapidly evolving include the advanced human interfaces, virtual reality and security/surveillance systems. This study provides a brief overview of the techniques and applications used for the markerless human motion capture, which deals with analyzing the human motion in the form of mathematical formulations. The major contribution of this research is that it classifies the computer vision based techniques of human motion capture based on the taxonomy, and then breaks its down into four systematically different categories of tracking, initialization, pose estimation and recognition. The detailed descriptions and the relationships descriptions are given for the techniques of tracking and pose estimation. The subcategories of each process are further described. Various hypotheses have been used by the researchers in this domain are surveyed and the evolution of these techniques have been explained. It has been concluded in the survey that most researchers have focused on using the mathematical body models for the markerless motion capture.

***Keywords*—**Human Motion Capture, Computer Vision, Vision based, Tracking.

I. Introduction

THE term motion capture can formally be defined as capturing the process of live motion even and transforming it into a mathematical formulation, somehow so that it can be used for further mathematical analysis. The idea of human motion capture was first introduced in 1887 by Eadweard Muybridge, during the experiments done for studying the animal locomotion. A new step was taken in the field in 1973 when Johansson, the psychologist conducted famous Moving Light Display (MLD) experiments. [1]

Considerable amount of research has been done till today in the area of human motion capture and analysis techniques. The increasing interest of the research community as well as the industry in the areas of motion capture is due to the several reasons, such as the advancements in the technology and techniques of Silicon in the modern world, lowered costs of video capture and processing application, and definitely the rapidly evolving applications of human motion capture; some of these application include advanced user interfaces, model based encoding, virtual reality, motion analysis, and smart surveillance systems [1].

Broadly, the research done in the field of human motion capture can be divided into 3 categories; face, body and gesture analysis. Each of these categories can further be divided into subcategories; for example, the face recognition involves recognizing the face in the scene, and analyzing the face so that it becomes possible to build the face models. Gesture analysis involves the gestures which a person gives by using hands, arms or both. Finally, body analysis deals with estimating the pose of the human. The scale at which the motion capture is to be done largely depends on the level of detail; while capturing the motion of the body, the minor details of face and gestures can be ignored and vice versa [2].

Some of the major technologies that are in use for human motion capture today include the use of optical tracking systems, or electromechanical and electromagnetic systems. The body suits with the tracing devices fixed at some points are used in the electromechanical approach. The sliding rods and potentiometers are used in the measuring devices which can detect the small motion of human. Although, these electrochemical suits provide with accurate results about the human motion, but they are inconvenient due to the restrictions on movement of the subject, and also due to their weight. Thus, the electrochemical approach fails to capture the motion data fully, as it restricts the human subject from walking or performing large scale motions. On the other hand, the electromagnetic approach of motion capture enlarges the range of motion that can be captured; the electromagnetic sensors are mounted at different parts of the body, to capture the orientation as well as the movement made by the specific body part [3]. The electromagnetic suits are clearly easier to use as compared to the electrochemical suits, but still there is disadvantage in this approach due to the use of wires. Finally, the optical tracking system provides a chance of most feasible human motion capture. This system consists of markers attached to various locations on the human body and a set of multiple cameras to capture the motion. The limb and joint locations of human are tracked using infrared reflective balls. Despite the convenience of using optical systems for human motion capture, the erroneous results may be obtained due to individual markers' misclassification. This error can be reduced by using pulsating LED's that vary frequency [4].

As seen from the above approaches, each motion capture approach offers restrictions and can only be used in certain application areas, the idea of markerless motion capture emerged to fit to a wide variety of applications. As the name suggests, this approach does not make use of any marker for tracking. Thus, the markerless approach of human motion capture refers to the capture of full body motion without the need of using specialized suits or markers. The major processes for which this research presents a comprehensive survey include tracking, initialization, pose recognition and

Dr. Najm Alotaibi is with Institute of Diplomatic Studies, Saudi Arabia (Phone: +9660540534336; e-mail: Alotaibinajm@gmail.com).

estimation. The tracking and pose estimation has been given more attention [5].

Next, the use of models for performing the human motion capture has been briefly discussed in Section II, and then the categories of tracking and pose estimation techniques have been Analyzed in the Sections III and IV. The work has been concluded in section V.

II. Using Models

Majority of researchers have used body models at some level during the process of human motion capture. The use of model facilitates the motion capture through providing a priori information of human motion capture. The plausible states of the body model makes the guidance of pose estimation or tracking simpler, based on the concept that the data of motion is only the description of locomotion of the body which can be defined either by the movement of a single human model over time, or joint positions over time. It is to be noted that most applications of human motion capture rely on the detailed description of the motion itself.

The pose estimation can thus be classified into 3 sub categories; the techniques which deploy the a priori models for constraints, the techniques which deploy the a priori models for constraints, and finally the techniques of pose estimation which do not deploy the a priori model at any level. Active and passive are the two types of human body models that deploy a priori information in different ways. The passive models make use of the a priori information for guidance of the tracking and pose estimation stages, but the information is not confined within the system [6]. On the other hand, the active models define the postures for representing the individual states of the model; thus, the information is extracted from the tracking and image processes.

Stick figure has been considered as one of the earliest representation of the human body [7]. In this model, each node represents the position of stylized joint, while each bone in the human body is assumed to be represented by the line between nodes. Most commercial motion capture system use Stick figure as the underlying model. 3D points in Euclidean space can be used to represent the position of nodes in the system, but mostly relative transformations from parent nodes are used for representations. In-build hierarchy has been found in this system; the subject has to adopt an initialization or starting pose and then transforming the joint positions to represent the global motion is required, in order to perform a motion simulation. The major disadvantage of this approach lies in the underlying difficulty of mapping the model configuration and the features of image. Thus, the techniques of flashing out the flesh and bones have been proposed by the authors to reduce this issue, for example the 2-D contours have been used by [8].

III. Tracking

A sequence of frames has to be correctly analyzed in order to track the human motion. Tracking human motion becomes challenging because of the non-rigid articulated motion due to which the appearance of the object keeps on changing with time. Also, there is the probability that the object may wholly or partially occlude within the sequence of frames [9]. Generally, three approaches are being used for tracking the humans:

1. The human is segmented from the background.
2. The complexity of the data is reduced by bringing it into more convenient form.
3. At the last step of tracking, some model is used for deploying the motion between the frames. It is generally assumed that the motion between the frames is small; this assumption helps in using the algorithms (such as Kalman Filter) for predicting the new feature position.

Each of these steps is briefly described next:

A. Segmentation

Tracking firstly requires identifying the target from the scene. In the terms of computer vision, identification of a target refers to the segmentation of target from the clutter. A large number of techniques have been proposed in the literature for achieving efficient segmentation. After segmentation is done, the morphological operators are often used for reducing the noise and improving the tracking results [10].

B. Image Differencing

This is the simplest possible technique used for detecting a change between the sequences of frames. For this approach, the two consecutive frames of a picture are taken, while the camera angle has not been changed significantly. The later frame is then subtracted from the former one and the resultant frame only contains the information that reveals the difference between the two frames. The only difference between the two frames is assumed to be the target that has moved, because the background of both the frames is already assumed to be fixed. However, with this difference, the white Gaussian noise also remains in the resultant frame, and it needs to be reduced. The lighting conditions are also essentially important for this segmentation technique.

One approach of segmentation has been proposed by Lee. The study used color and light for segmenting the face regions in the real scenes [11]. A velocity light vector of the face is extracted initially and then the clear motion is showed by thresholding; the hue space thresholding is used in conjunction with the extracted information obtained with the help of previously mentioned technique. Another researcher Davis has worked on the same approach as Lee, but he has also assumed that the target person is moving while the background is either stationary or moving very slow [12].

C. Appearance Data

This approach segments the target based on its appearance. This approach works under the assumption that there is clear difference between the appearance of the target and its background; a threshold of difference is often set for this approach. Croma key refers to the process of threshold in this approach. Generally, the color of background is turned to blue and then the non-blue pixels are taken to be as the target

required [13].

While using this approach, the separate parts of body may be colored differently for easy and quick segmentation. Similarly, the thermal images can also be segmented easily through this approach as the hot bodies or the shadows behind the subjects may easily be detected.

D.Temporal Tracking

Knowledge about the motion model has been used in the approach of temporal tracking in order to split a scene into different regions, known as blobs. The algorithms such as Kalman filter can then be used for modeling and predicting these blobs. The basic function of the Kalman filter is that it can predict the position of the feature, when given with some moving feature and a motion model to detect the motion. Initially, the prediction area is set to be large; then the local search is performed, and the prediction area is kept on updating and improving gradually [14].

E. Tracking through Multiple Hypotheses

It is to be noted that in many tracking application, the Kalman filtering approach does not suffice because it is based on Gaussian densities. Thus, the tracking algorithms based on these filters fail to track simultaneous multiple hypotheses. Various approaches have been developed for tracking multiple hypotheses simultaneously; CONDENSATION algorithm is one of them. Factors sampling was used in this algorithm, which was earlier used for interpretation of static images. The randomly generated sets are used for interpretation of probability distribution of the tracking. The random set is propagated over time using the visual observations as well as the learned dynamic models [15].

F. Low Level and High Level

The low level processing, in terms of computer vision refers to the concept where operations are performed at the grey scale. For example, there is no need for these operations to know whether a particular pixel belongs to the target or the background. On the other hand, the high level information requires the knowledge of lower levels as well. Thus, the data found at the lower levels is to be processed until some useful information is produced to be used at the higher levels; this approach is often referred to us bottom-up approach.

The low level data can generally be easily matched with the human models, because the low level data has high probability of being coinciding with the human data which the researchers maintain for their reference. A multilevel approach has also been proposed; this approach assumes that while in motion, most body parts of the humans move. The framework developed for detecting the human motion comprises of 4 levels and thus often delays the result.

IV. Pose Estimation

The 3D articulated motion approaches have been used for detecting the pose of human targets. All pose detectors use some model at various levels of their pose detection processes. However, clear difference exists between the algorithms that use a-priori algorithm and those which do not; difference mainly lies in how these algorithms establish the feature correspondence between the two consecutive frames [16].

The a-priori models base the correspondence upon the predictions that relate to texture, shape, velocity and position. The features are predicted using mathematical functions, and statistical heuristics are then applied for describing the human locomotion.

A. Passive a-Priori Models

The passive a-priori models have been used for the purpose of constraint guidance. In this approach, the information about any physical feature is used for the guidance of tracking system, for example the person's height [17].

B. Active Model Binding

Many authors used a-priori models which deploy each state to represent a valid pose. This helps in using the available 3-D knowledge to the maximum possible limit, and avoiding the 2-D knowledge. Avoidance of using 2-D knowledge clearly minimizes the probability of errors in the process of feature extraction and tracking.

C. Analysis by Synthesis

This approach is used to analyze a scene by comparing it with the model of the same scene. The major advantage of this approach is that it is computationally less expensive because it does not require the conversion of non-linear measurement equations.

The PDM (Point Distribution Model) is one of the best examples that deploy the analysis by synthesis approach for the purpose of tracking the human shape. This is done by placing a sequence of control points in the scene [18].

D.Model Dimensionality

This is one the standard techniques that have been developed by the research community in order to minimize the error probability or the cost function. This approach does not require any prior correspondence between the scene and model. For example, the global search strategies that are based on greedy techniques such as neural networks and greedy algorithms are based on the perturbation of individual state parameters [19].

V. Conclusion

This paper initially provided an overview of the definitions, techniques and applications of human motion capture, as well as the models for these applications is then presented. Finally, the paper provided a description on the 4 techniques of marker less based motion capture including tracking, initialization, pose estimation and gesture recognition. The detailed analysis of pose estimation and tracking has been presented. Number of approaches involving the use of human boy models has been found in literature for completing these two stages of human motion capture.

REFERENCES

[1] Gall, J., Rosenhahn, B., Brox, T., & Seidel, H. P. (2010). Optimization and filtering for human motion capture. International journal of computer vision, 87(1-2), 75-92
[2] Pons-Moll, G., Baak, A., Helten, T., Muller, M., Seidel, H. P., & Rosenhahn, B. (2010). Multisensor-fusion for 3d full-body human motion capture. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (pp. 663-670). IEEE.
[3] Swaisaenyakorn, S., Kelly, S. W., Young, P. R., & Batchelor, J. C. (2012). Evaluation of 3D animated human model from 3D scanner and motion capture to be used in electromagnetic simulator for body-centric system. In Biomedical Engineering and Informatics (BMEI), 5th International Conference on (pp. 632-636). IEEE.
[4] Field, M., Pan, Z., Stirling, D., & Naghdy, F. (2011). Human motion capture sensors and analysis in robotics. Industrial Robot: An International Journal, 38(2), 163-171.
[5] Sigal, L., Balan, A. O., & Black, M. J. (2010). Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. International journal of computer vision, 87(1-2), 4-27.
[6] Keskin, C., Kıraç, F., Kara, Y. E., & Akarun, L. (2013). Real time hand pose estimation using depth sensors. In Consumer Depth Cameras for Computer Vision (pp. 119-137). Springer London.
[7] G. Johnasson, (1973). "Visual Perception of Biological Motion and a Model for Its Analysis." Perception Psychophysics 14(2): 201-211.
[8] M. K. Leung and Y.H. Yang (1995). "First Sight: A Human Body Outline Labeling System." IEEE Transactions on Pattern Analysis and Machine Intelligence 17(4): 359-377.
[9] Sapp, Benjamin, David Weiss, and Ben Taskar. "Parsing human motion with stretchable models." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.
[10] Shao, L., Ji, L., Liu, Y., & Zhang, J. (2012). Human action segmentation and recognition via motion and shape analysis. Pattern Recognition Letters, 33(4), 438-445.
[11] J. J. a. T. S. H. ". Kuch, "Vision Based Hand Modeling and Tracking for virtual teleconferencing and telecollaboration," in ICCV , 1995.
[12] J. W. a. A. B. Davis, "The Representation and Recognition of Action Using Temporal Templates," in International Conference on Computer Vision and, 1995.
[13] Jiang, Z., Lin, Z., & Davis, L. S. (2012). Recognizing human actions by learning and matching shape-motion prototype trees. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(3), 533-547.
[14] Niebles, J. C., Han, B., & Fei-Fei, L. (2010, June). Efficient extraction of human motion volumes by tracking. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (pp. 655-662). IEEE.
[15] Li, R., Tian, T. P., Sclaroff, S., & Yang, M. H. (2010). 3d human motion tracking with a coordinated mixture of factor analyzers. International Journal of Computer Vision, 87(1-2), 170-190.
[16] Vondrak, M., Sigal, L., & Jenkins, O. C. (2013). Dynamical simulation priors for human motion tracking. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(1), 52-65.
[17] Stone, E. E., & Skubic, M. (2011, May). Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. In Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on (pp. 71-77). IEEE.
[18] Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. ACM Computing Surveys (CSUR), 43(3), 16.
[19] Raskin, L., Rudzsky, M., & Rivlin, E. (2011). Dimensionality reduction using a Gaussian Process Annealed Particle Filter for tracking and classification of articulated body motions. Computer Vision and Image Understanding, 115(4), 503-519.