A Study on Video of Motion Attitude for 3-D Human Computer Interaction

Yiou Liu

Xi'an International University, Shanxi,Xi'an710077 ,China

***Abstract* - Three dimensional human motion generation based on video is one of the most important processes in human motion analysis, but also it is a high level vision processing foundation. Based on previous work, we propose a reconstruction method based on motion database modeling, namely on 3D motion database collecting in fixed length data segment units for statistical learning and modeling, the establishment of metastasis model of spatial structure similar model and the first order Mkarov time probability. Thus, 3D movement in spatio temporal database model under the guidance of taking 3D human posture reconstruction and deformation refinement technique based on video content as the foundation, which can achieve 3D motion reconstruction of automation.**

Keywords - Histogram equalization; Image process; Three-dimensional motion; Pose estimation; Motion capture

I. INTRODUCTION

Computer vision is an important branch of computer science and artificial intelligence, its main content is how to study, using a variety of imaging system in place of the visual organ as the signal input means, by a computer to replace the brain to complete processing and interpretation of information [1-3]. The ultimate goal of computer vision, is to enable the computer to like that through visual observation and understanding of the world, among them, the video human motion analysis is one of the important topics in computer vision research field in recent years is also based on the front direction, much researchers attention [4].

Human motion analysis based on video is pattern recognition in recent years, one of the most active and intelligent man-machine interface and virtual reality in the field of study. It not only has important significance, but also in intelligent monitoring, sports analysis, has the broad application prospect animation generation [5]. Therefore, it has attracted researchers more and more interests, the main purpose of human motion analysis based on video, is containing the detection, video image sequence from a group of people in the identification, tracking of human behavior, and carries on the analysis and understanding of it, which, motion detection and tracking belong to the low-level vision problems, and behavior understanding and description belongs to the problem high level vision [6, 7].

In this paper, 3D human pose refinement based on video content, which is based on the known initial human 3D attitude information, according to the human body contour information in the video, the initial 3D of human body posture deformation, so as to obtain the accurate 3D structure of the human body information corresponding to the video in. In our practical applications, we combine matching method based on point set of contour, and comprehensive utilization of ZD skeleton information, to solve the human pose refinement based on video content. The experimental results show that, compared with the existing methods, our method can not only ensure the accuracy of processing results, but only through simple and efficient ZD operation to achieve deformation refinement.

II. OVERVIEW OF RESEARCH ON VIDEO TRACKING ALGORITHM

A. Research Progress of Human Pose Estimation

Human motion analysis refers to the use of some means, tracking, capture the motion of the human body, to obtain the body motion parameters and structure and posture reconstruction of human body, and its ultimate aim is to the understanding of human movement and it's application [8-10]. The primary task of human motion analysis and the basic problem is to obtain the body motion parameters and movement parameters, access to a variety of methods, including the use of special mechanical equipment or optical, electronic equipment, video human motion analysis of this study is to obtain the direct use of video information for human motion parameters! Video content analysis and understanding, often the main body contains motion and complex background, a moving subject dress casual, need to be calculated by the algorithm for human motion state.

The scholars at home and abroad human motion detection carried out a lot of research work, [11] conducted a detailed literature review of the problem. Briefly speaking, human motion detection at present there are two main ideas: one based on Detection Based on body and part of the test, each type of framework, different researchers according to the image features and

detection method using expression and different algorithms.

Human detection based on the whole human motion lies across the region as an objective, according to the image of the region is established to solve models such as [12] using the edge image, through the distance transformation matching image edge with the body contour shape Library of the sample, thereby positioning the human body possible positions on the image by wavelet transform to obtain images the characteristics, characteristics of the support vector machine (SVM) classifier is used for classification, pedestrian and non pedestrian image sample in the training stage model.

B. Survey and Research of Motion Tracking of Human Video Detection

In the early seventy's of the last century, psychologist Johan through mobile using two-dimensional illumination display device MLD to explore the visual analysis of human motion based on the system. As shown in Figure 1 joint left on the human model (black dots) have become a group of point set. Johansson found in the experiment show a two-dimensional image in MLD, observation can through the observation of MLD point sets in a variety of sports demonstration, recognition in less than 200 milliseconds, which include walking, running, climbing stairs and other simple periodic motion, including dancing, greetings, boxing and other complex motion, sometimes even can distinguish the presenter gender: but in demonstration of static or MLD image inversion conditions, the observer is unable to make the discrimination. This work on human motion recognition based on video image is of great enlightening significance.

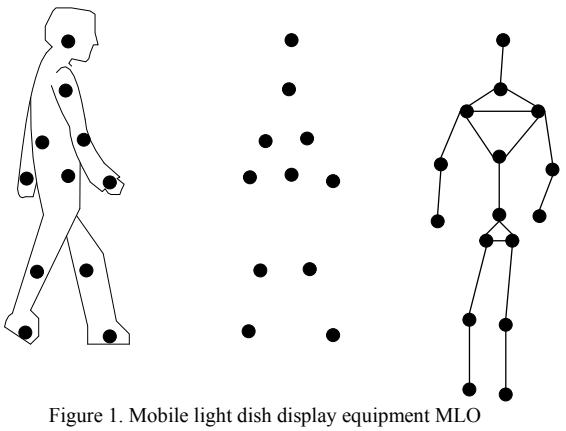

Figure 1. Mobile light dish display equipment MLO

Researchers for the Johansson experimental phenomena hold two different views: one view that the first reconstruction of 3D structure recognition, that people first set information recovery of joint points of the human body using the MLD display point, then according to the 3D structure recognition behavior; another point of view, advocated the recognition does not depend on the reconstruction, think people directly the motion information is used in MLD to identify, without reconstruction process. Different explanation to experimental phenomena corresponding to two schools of study behavior recognition: the former is devoted to the study of how to structure from motion estimation, and image feature representation and model makers focus on human motion is established, and pattern recognition methods on the model of the tomb.

C. The Deficiencies of Existing Algorithms

According to the number of cameras used, it can be divided into single objective method based on multi objective method. And based on the monocular method refers to the use of a camera to acquire images, only on video sequences acquired from a camera angle were analyzed, and the multi view refers to the use of multiple cameras, to process the video sequences acquired from multi angle. Because of the video data using single objective method, the monocular method than multi purpose application was more widely. However, a single method has a certain binding, i.e., obtaining the depth information in need of application fields, it is difficult to obtain satisfactory results, and multicast techniques through the video sequence multi angle analysis, better get depth information, but it often requires under laboratory conditions, compared with the single target method, added many restrictions.

III. THE 3D POSE ESTIMATION OF WEAK LABELING

A. Human Posture Reconstruction

The human body position with a 3D vector to express, and each joint orientation information with unit four-number (unituqatemoins) to describe the "human posture data movement in the database is represented by in position as the coordinate origin of coordinates.

Thus, 3D motion database, 3D human motion can be expressed as:

$$m(t) = (p(t), q^1(t), ...q^t(t), ...q^n(t)) \quad (1)$$

The filtering process is the first two-dimensional sliding template selection of specific structure, and then will be arranged according to the size of pixel values in the form template, the data sequence monotone. The output formula of two dimensional median for:

$$G(x, y) = \text{Mid } f(x-i, y-j), (i, j) \quad (2)$$

Gaussian smoothing filter is a widely used algorithm in image processing, mathematical expressions for Gaussian smoothing filter:

$$f'(x,y) = f(x,y) * g(x,y) \quad (3)$$

$$g(x,y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (4)$$

B. Algorithm Framework Signed on Sample

According to the background image model known to infer new data is the background or foreground object, is essentially a process of classification of two kinds of data, this paper uses hypothesis testing are classified. The statistical inference task is to test the model data is then accept or reject, accept the data as a background image, refuse is due to foreground objects. Some noise existed in single frame data, new video data first smoothing in the time domain, and then enter the model for statistical inference.

$\bar{\mu}_K^{(m)}(x,y)$ said the video sequence to mean $K \in \{H,S,V\}$ current frame adjacent M image as the center of the pixels in the P color components.

Hypothesis testing on it for:

$$H_0 : \bar{\mu}_K^{(m)}(x,y) = \mu_K(x,y) \text{ and } H_1 : \bar{\mu}_K^{(m)}(x,y) \neq \mu_K(x,y) \quad (5)$$

The pixel belongs to the background region judgment for formula:

$$\begin{cases} \left|\bar{\mu}_H^{(m)}(x,y) - \mu_H(x,y)\right| < \delta_H \\ \left|\bar{\mu}_S^{(m)}(x,y) - \mu_S(x,y)\right| < \delta_S \\ \left|\bar{\mu}_V^{(m)}(x,y) - \mu_V(x,y)\right| < \delta_V \end{cases} \quad (6)$$

Based on the hypothesis testing, to mean the adjacent m frame in subsequent frames $\bar{\mu}_K^{(m)}(x,y)$ for inspection, so can inhibit the noise function, so that more accurate segmentation results. But m value can not be too large, otherwise it will over smoothing, changes caused by the loss of video object.

C. Dimension Reduced Processing Shape Context PCA

We use Hu moments and ShpaeCnotxet respectively (and its PCA dimension reduction) on the shape is described, and the analog video contours in contour libraries for shape matching. The reconstruction results of the statistical comparison as shown in Figure 2. As can be seen, the hsPaeconixet method has more advantages than the Hu method based on regional features, matching success steadily at the rate of 90%; even the analysis of the PCA will be a 60 dimensional vector information of the original fell into 2 dimensions by principal, the matching result is success rate compared with the Hu method, there is still a relatively higher matching rate of success.

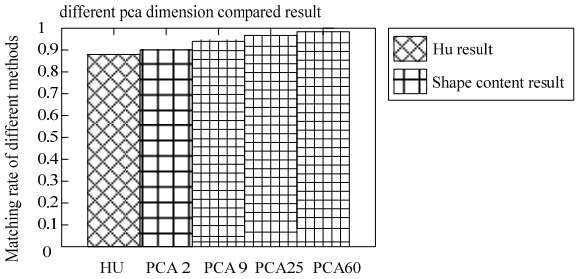

Figure 2.The reconstruction results of the statistical comparison.

D. The Clustering Tree and Clustering of Forest

After the C means clustering on motion data movement in the library; next, we for each class generated by constructing a clustering tree, order them, to describe the temporal motion data transfer between the relationship between categories, and all these clustering tree, is part of the cluster of forest. For a given class, steps to set the cluster tree as follows:

First of all, the tail frame of the class of the data segment, the original motion sequences found in the next data frame, if the category data segment the frame where the place and a different categories, the category of the frame where the segment as a child node categories. Figure 3 is a low level model of first-order Markov probability the transfer model, pre established probability between different data segment transfer.

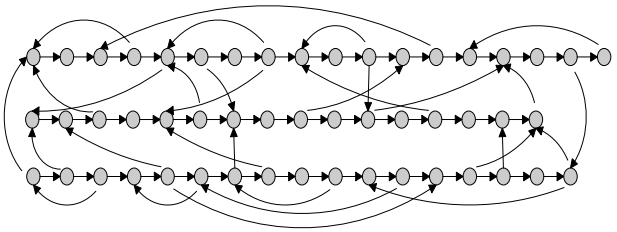

Figure 3. The low level model of first-order Markov probability the transfer model

Color information is compared with other feature information of target rotation, translation, deformation, occlusion is not sensitive, and compared with other feature information more convenient, is widely used in image processing, image segmentation, target detection and tracking in computer vision. In a video image, color distribution was shown in Figure 4.

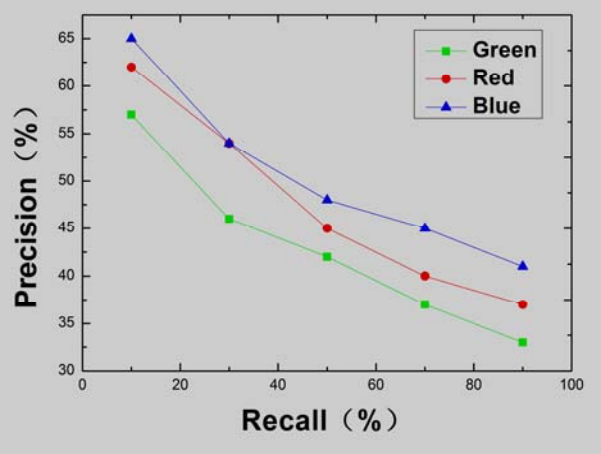

Figure 4. Color distribution in a video image.

IV. RESULTS AND SIMULATION

A. District of the City of Inspection Results

From the table 1, the calculation time is mainly spent on contour extraction and dimensionality reduction, in the reconstruction phase, of the 2000 frame data are the total time spent in reconstruction of about 810 seconds. The average time of single frame processing for 0.0405 seconds, even considering the whole pre-processing stage time, the average time of the single frame processing still for about 1.59 seconds, with a chapter on the method of Hu moments and based on AMlS (mean time to process a single frame of 13.5327 seconds), the calculation efficiency is greatly improved.

TABLE 1. ATTITUDE RECONSTRUCTION PROCESSING TIME STATISTICAL

	(2D contour library generation/contour extraction)& point sampling	Shape Context &PCA	Embedding	LSH
Pre-Processing Stage	~510 (sec)	~1050 (sec)	~150 (sec)	~580 (sec) (create table)
Reconstruction stage	~170 (sec)	~350 (sec)	~60 (sec)	~230 (sec) (query)

In order to further verify the physical reality of the reconstruction results, we generate synthetic video data sequences by Posers software, and motion reconstruction for the synthesis of video data. In order to better simulate the real video sequences, we outline of synthetic data adding noise processing closed operation (after the first expansion, corrosion), and motion reconstruction of adding noise to the processed data. For example, we are on a segment length of 439 frames (segment number is 29) plus noise synthetic motion sequences are reconstructed, as shown in Figure 5, and the reconstruction results with the real motion data on the quantitative comparison.

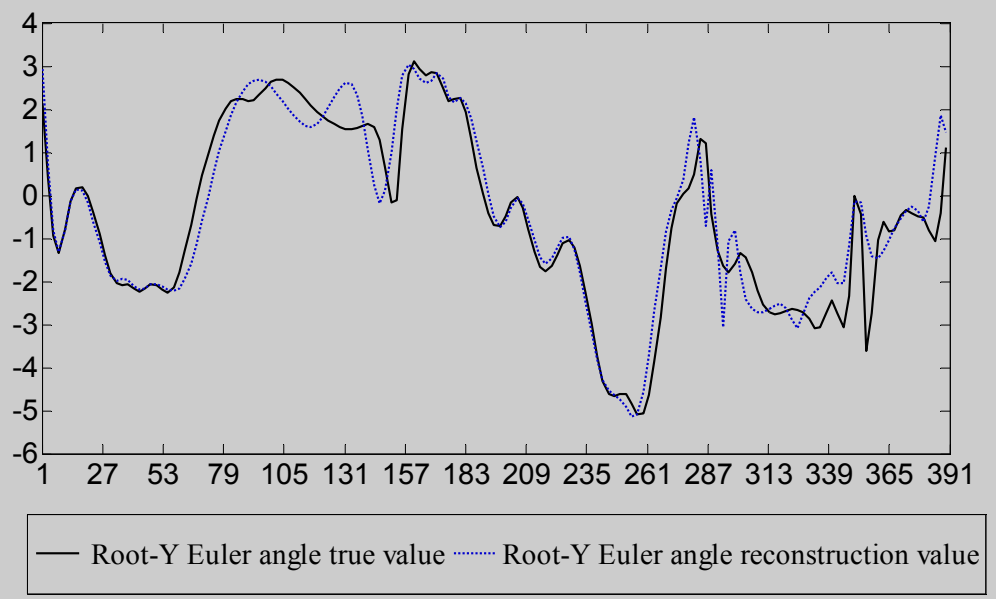

Figure 5. The plus noise synthetic motion sequences

B. The Body Contour Extraction of Static Background

Motion segmentation, foreground contour extraction from the human body motion video is the essential step in the motion analysis technique based on contour. This paper selects the trampoline as a case study for trampoline; moving camera motion information not contained in video, so in the treatment in the process, we can use the static background of human motion segmentation based on technology is relatively easy to deal with.

The solution formula of video frame CDM as follows:

$$CDM(x,y,i)=\begin{cases} d, \ if \ \ d \geq T \\ 0, \ if \ \ d < T \end{cases}, d = \left| I_L(x,y,i+1) - I_L(x,y,i) \right| \tag{7}$$

Throughout the coordinate position of all video frames (x, y), thereby forming a complete background, the procedure can be used the following formula description:

$$M(x,y) = (ST(x,y) + EN(x,y))/2 \tag{8}$$

$$B(x,y) = I(x,y,M(x,y)) \tag{9}$$

C. Experimental Scene

The proposed 3D human motion generation method based on video using motion database based on 3D in the field, technical route and extract information from the ZD 3D movement contour information. Due to the dependence on the domain specific knowledge of 3D motion database, this method can only be used for motion reconstruction of specific motion types; but at the same time, it is also a general the system framework of the type: the motion database determines the applicable scope of movement types.

The main contents of this paper include: 3D profile reconstruction based on contour matching; 3D pose refinement based on video content; combined with the human motion reconstruction motion library spatio-temporal model. The whole process of 3D video based motion generation of video was shown in Figure 6.

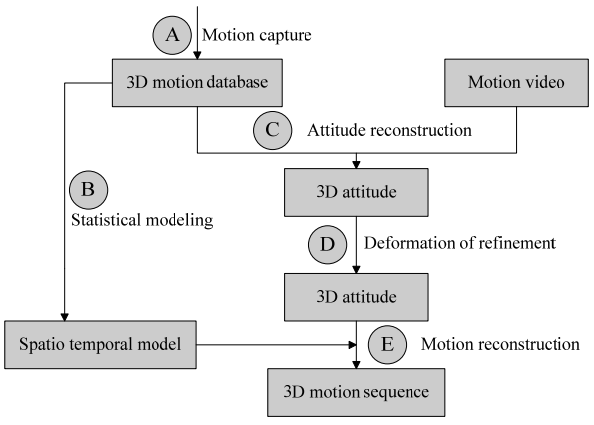

Figure 6. The whole process of 3D video based motion generation of video.

D. Experimental Results and Analysis

The main line of this paper is a monocular 3D human pose estimation for human-computer interaction, including its special conditions: low network camera video capture system is usually a picture of poor quality, interactive system requires that the core algorithm also has the accuracy, robustness and real-time, system initialization to automation. In this paper the main clue of the research under the guidance of the system separately, estimation of monocular 3D human pose three research core based on constraints algorithm is designed.

The motion characteristics of mining, the main steps for design reference system framework, the main process was shown in Figure 7.

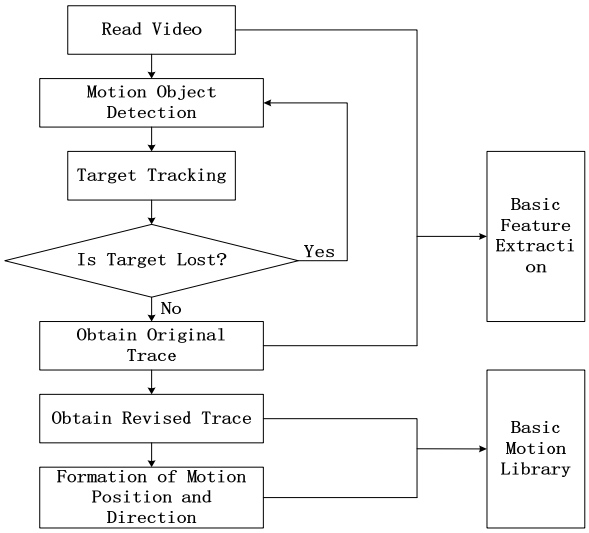

Figure 7. The main motion process characteristics of mining

V. CONCLUSION

3D human pose refinement based on video content, which is based on the known initial human 3D attitude information, according to the human body contour information in the video, the initial 3D of human body posture deformation, so as to obtain the accurate 3D structure of the human body information corresponding to the video in. In our practical applications, we combine matching method based on point set of contour, and comprehensive utilization of ZD skeleton information, to solve the human pose refinement based on video content. The experimental results show that, compared with the existing methods, our method can not only ensure the accuracy of processing results, but only through simple and efficient ZD operation to achieve deformation refinement. In addition, the deformation of our technique is based on the generic processing framework, can be used for pose refinement deal with objects of various types of known initial 3D information, embodies the good generality.

REFERENCES

[1] D. Abbott, S. Cunningham, G. Daniels, B. Doyle, J. Dumlop, D. Economo, T. Farmer, D. Farrant, C. Foley, B. Fox, M. Hedley, J. Herrmann, C. Jacka, "Development and Evaluation of Sensor Concepts for Ageless Aerospace Vehicles – Threats and Measurands," NASA, vo. 2, no. 11, pp. 68-77, 2002.

[2] Li JianZhong, Li JinBao, Shi ShengFei, "Concepts, Issues and Advance of Sensor Networks and Data Management of Sensor Networks," Journal of Software, vol. 14, no. 10, pp. 1717-1727, 2003.

[3] Tilak S, Abu-Ghazaleh NB, Heinzelman W, "A Taxonomy of Wireless Micro-sensor Network models," Mobile Computing and Commuications Review, vo. 1, no. 2, pp. 1-8, 2002.

[4] David Culler, Deborah Estrin, Mani Srivastava, "Overview of Sensor Networks", IEEE Computer Society Officers, vol. 11, no. 10, pp. 15-25, 2005.

[5] T. Matsuo, Y. Kaneko, M. Matano, "Introduction of Intelligent Vehicle Detection Sensors", IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems, vol. 5, no. 12, pp. 709-713, 1999.

[6] J. Deng, R. Han, S. Mishra, "Inrusion tolerance and Anti-traffic Analysis Strategies in Wireless Sensor Networks," IEEE 2004 International Conference on Dependable Systems and Networks, vo. 10, no. 11, pp. 477-490, 2004.

[7] Q.R. Li, L.Y. Wei, S.F. Ma, "The Model Analysis of Vehicles Situation and Distribution in Intersections Abased on Markov Process," IEEE International Conference on Intelligent Transportation Systems. vol. 2, no. 10, pp.1076-1080, 2003.

[8] Qinghai Wang, Haihui Ma, Yuhuan Chi, Yin Li, Linchong Dong, " Chinese word knowledge improvement based on HNC ", Journal of Chinese Information Processing, vol. 26, no. 2, pp. 35-39, 2012.

[9] Yao Meng, Tiejun Zhao, Sheng Li, Jianming Yao, " Based on the evaluation of the English syntactic structure disambiguation and self evaluation rule correction", Journal of Computer Research and Development, vol. 39, no. 7, pp. 802-808, 2002.

[10] Starck J L, Candès E J, Donoho D L. "The curvelet transform for image denoising".IEEE Trans. on Image Proc,vol. 6, no. 11, pp. 131-141, 2002.

[11] Li Jian, Pan Qing, Yang Tian. "Color Based Grayscale-fused Image Enhancement Algorithm for Video Surveillance", In Proceeding of the third international conference on image and graphics, pp. 47-50, 2004.

[12] D. Abbott, S. Cunningham, "Development and Evaluation of Sensor Concepts for Ageless Aerospace Vehicles --- Threats and Measurands," NASA/CR-2002-211772, vol. 11, no. 2, pp.25, 2002.