Automatic 3D CBIR on Kinematical Human Motion

CHUN-HONG HUANG[1], CHING-SHENG WANG[2] and MENG-LIANG YU[1]
[1]Department of Computer Science and Information Engineering
Tamkang University,
Taiwan, R.O.C.

[2]Department of Computer and Information Science
Aletheia University,
Taiwan, R.O.C

Abstract: Automatic retrieval of human body motion is one of the most challenging issues of video and virtual reality research. In this paper, we propose a human motion retrieval system, which allows users to retrieve 3D kinematical motions. The system includes two major components for motion analysis and comparison. The first is a recognition unit of motion types which is based on Skeleton Discrimination Tree. The second is a unit for synchronization and motion similarity comparison. The comparative approach is based on mutative dynamic programming that considers the degree of the included angles of the vectors belonging to individual feature tracks. With Motion Capture Camera and animations represented by VRML, it is possible to automatically summarize, analyze and adjust the 3D motions of a real person. Users may provide a 3D VRML human motion object and find the similar human motions via our system.

Key-Words: Motion Retrieval, Motion Analysis, Skeleton Discrimination Tree, VRML, Human Skeleton

1 Introduction

Multimedia technologies introduce an interesting but challenging research issues. Particularly, Content-Based Information Retrieval (CBIR) on image is a difficult research topic. The purpose of CBIR is to retrieve from a set of images or an image database that may contain relevant information to a user's query. The proposed mechanism is used to overcome the difficulties of manual annotation of images, which could be indexed by their own visual content, such as color, shape, spatial relations, etc. Yet, few practical systems claim that the recognition rate is satisfiable. In the last decade, CBR on 3D objects or object motions became a popular topic. However, most motion retrieval systems are designed based on 2-D video information. Some research techniques for Content-Based 3D object retrieval have been reported [1][2]. Nevertheless, these techniques for Content-Based 3D object retrieval focused merely on static 3D objects, which cannot be represented automatically in 3D browser and, therefore, cannot meet users' need to control the operations to observe the objects from different viewpoints. In this paper, we aim at classification and retrieval on Human Kinematical Motions in a 3D space.

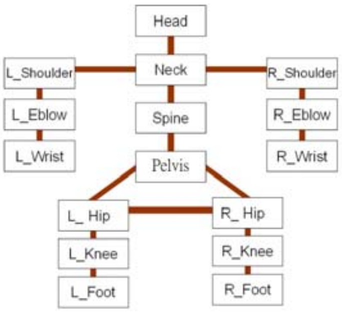

Fig.1. A Human Body Skeleton Model

The unit motion of kinematical domain as an example shows our contribution and the possible extension. According to a human Body Skeleton model (Figure 1), these sixteen joints can be considered as the feature points which represent different sets of body joint motions. In order to compare the animation tracks of different feature points, a recognition technique is required for the exiting manifold human motion types. Thus, an automatic retrieval system tells the user which set of kinematical actions is similar to those he/she is learning. We adopt a professional motion capture system, named VICON with 7 cameras to capture motion sequences representing the kinematical motions.

2 Related Work

The techniques for automatic retrieval of human body motion include tracking and behavior understanding. Object tracking in video streams is quite popular in the field of computer vision. As an extension of motion tracking, behavior understanding is the most important goal of research in human motion analysis and human motion recognition.

2.1 Tracking

The tracking technologies include model-based tracking, feature-based tracking and multi-camera tracking. As a rule, model-based tracking of human body can use stick figures [3], 2-D contour [4] or volumetric model [5], so that body segments can be simulated as lines, 2-D ribbons, and 3-D volumes accordingly. Feature-based tracking method uses sub-features such as distinguishable points or lines on the object to achieve tracking task. The benefit is that even in the presence of partial occlusion, some of the sub-features of the tracked objects remain to be visible. A good example to point out feature tracking can be found in Polana and Nelson [6]. The motions of arms and legs are considered to aggregate to the torso in their work. The moving person is bounded by a rectangular box, and the center of the box was selected as the feature point for tracking.

2.2 Behavior Understanding

Behavior understanding involves action recognition and description. It could produce high-level description of actions and interactions. Action recognition involved in behavior understanding could be considered as a time-varying data matching problem. Dynamic programming has the advantages of conceptual simplicity and robust performance and has been widely used in the matching of human motion patterns. For instance, Yabe and Tanaka [7] proposed a technique based on Dynamic Programming to perform the similarity retrieval of human motion. Hidden Markov models (HMMs) is a kind of stochastic state machine which be employed in analyzing time-varying data with spatiotemporal variability. Both papers Brand et al. [8] and J. Yamato et al. [9] use HMMs to recognize human actions.

The organization of this paper is as follows. In section 3 presents the classification of motion types. A distance function which aggregates the feature points in human motion is proposed in section 4. In section 5 are Experimental Results and Comparison. Finally, a short conclusion is given in section 6.

3 Classification of Motion Types

For distinguishing motion types, we use Skeleton Discrimination Tree (Figure 2) based on the variation of momentum in 3D space. In the architecture of Skeleton Discrimination Tree, there are 24 groups that would be divided into five categories dynamically for each query procedure with different actions to decrease the computing loading. The significance of each feature point is very important according to our human motion templates. We disregard some feature points such as hip and spine. In addition, we treat hands and legs as aggregated units. This is due to the fact that shoulders, elbows, and wrists are naturally connected. Thus, we differentiate the degree of momenta among the left hand (LH), the right hand (RH), the left leg (LL), and the right leg (RL). The first approach of our similarity function, which compares the kinematical motions relies only on adding weights to the above four aggregated units. For example, suppose that a query spar has degree of momenta sorted as RH > LH > RL > LL. And, let RH_d, LH_d, RL_d, and LL_d be the differences of feature point vectors between the query action and a target action. We propose a skeleton discrimination tree, which is constructed for each query motions. For instance, if $LH_d > RH_d > LL_d > RL_d$, the tree is shown in Figure 2. According to the human skeleton, actions are discriminated into five categories. The category distance between category A and the rest are 1, 2, 3, and 4 for categories B, C, D, and E, respectively. For instance, category D includes a case $RH_d > LH_d > LL_d > RL_d$, which has distance 3 from category A. Note that there are 24 cases based on the momenta. Each case has its discrimination tree. The purpose of classifying motion types is to increase the retrieval rate and accuracy. A new motion type should be classified into its corresponding motion types. For example, the type of a query object is “Toss”, and then the query object will be compared with other objects in the same motion group which is stored in the database. It would reduce the time of computing and the overload with classification.

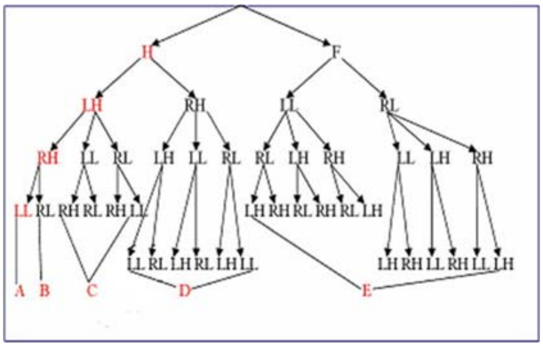
H
F
LH
RH
LL
RL
RH LL RL LH LL RL RL LH RH LL LH RH
LL RL RH RL RH LL LH RH RL RH RL LH
LL RL LH RL LH LL LH RH LL RH LL LH
A
B
C
D
E

Fig.2. Skeleton Discrimination Tree

4 Human Motion Similarity

After the classification to human movement types, we can compare the human body motion in detail to all joint parts. The work includes two parts of similarity measures. The first is the similarity between motions of each corresponding joint part of two human models. However, the similarity between two human bodies should be calculated not only the similarity between two persons' movement but the two persons' degree of synchronization among all body part motions. The second work of the two persons' degree of synchronization among all body part motions would also be described in this section.

4.1 Motion Similarity

The implementation of our approach requires a VRML parser, which decomposes objects of kinematical motion into different nodes. We put emphasis on the information regarding coordinates and sizes. Color information and appearance node (i.e., material) are not used. Before comparing the detail motion, we should extract the trajectories of feature joints from VRML kinematical model. The coordinates of motion are consecutive points which can be subdivided into series of vectors. The serial data can be used for the comparison between two human motions based on Dynamic Programming, which solves problems by combining the solutions of sub-problems and is applied to optimization problems typically. It is also a pattern matching approach widely used in the field of bio-information, security, and image processing etc. Assuming we have two human motion models *Q* and *T*. *Q* is the query model, and *T* is one of the template models which stored in the database. Both of these models have 16 joints.

$$Q=(q_1,q_2,q_3\ldots,q_h\ldots,q_{16})\ ,\ T=(t_1,t_2,t_3,\ldots,t_h,\ldots,t_{16})$$

where *h* is one of a human model's 16 joints, and t_h is the motion of joint *h* in *T* model, q_h is the motion of joint *h* in *Q* model. Assuming motion t_h has *m* points, and motion q_h has *n* points. Then, the assumptions can be expressed as follows:

$$t_h=(t_{h1},t_{h2},t_{h3},\ldots,t_{hi}\ldots,t_{hm}),$$

and

$$q_h=(q_{h1},q_{h2},q_{h3},\ldots,q_{hj}\ldots,q_{hn})$$

The vectors of q_h and t_h are $\overrightarrow{V_{q_h}}$ and $\overrightarrow{V_{t_h}}$

$$\overrightarrow{V_{qh}} = (\overrightarrow{v_{q_{h1}}}, \overrightarrow{v_{q_{h2}}}, \ldots \overrightarrow{v_{q_{hi}}}, \ldots \overrightarrow{v_{q_{h(n-1)}}})$$

Let *n*-1 to be *a*, we obtain $\overrightarrow{V_{qh}} = (\overrightarrow{v_{q_{h1}}}, \overrightarrow{v_{q_{h2}}}, \ldots \overrightarrow{v_{q_{hi}}}, \ldots \overrightarrow{v_{q_{ha}}})$, and

$$\overrightarrow{V_{t_h}} = (\overrightarrow{v_{t_{h1}}}, \overrightarrow{v_{t_{h2}}}, \ldots \overrightarrow{v_{t_{hj}}}, \ldots \overrightarrow{v_{t_{h(m-1)}}})$$

Let *m*-1 to be *b*, we have $\overrightarrow{V_{t_h}} = (\overrightarrow{v_{t_{h1}}}, \overrightarrow{v_{t_{h2}}}, \ldots \overrightarrow{v_{t_{hj}}}, \ldots \overrightarrow{v_{t_{hb}}})$.

The similarity $Mot(\overrightarrow{v_{q_{hi}}}, \overrightarrow{v_{t_{hj}}})$ between the vectors $\overrightarrow{v_{q_{hi}}}$ of q_h and $\overrightarrow{v_{t_{hj}}}$ of t_h can be calculated by Dynamic Programming with the calculation steps and optimal structures as equation 1. θ is the included angle of vectors between q_{hi} and t_{hj}.

$$1.\ Mot(\overrightarrow{v_{q_{hi}}}, \overrightarrow{v_{t_{hj}}}) = \begin{cases} 0 & if\ \ i=0\ or\ \ j=0, \\ Mot(\overrightarrow{v_{q_{h(i-1)}}}, \overrightarrow{v_{t_{h(j-1)}}})+3 \quad if\ \ 0^o<\theta\le 2^o, & \\ Max(Mot(\overrightarrow{v_{q_{h(i-1)}}}, \overrightarrow{v_{t_{hj}}}), Mot(\overrightarrow{v_{q_{hi}}}, \overrightarrow{v_{t_{h(j-1)}}})) \ + \begin{cases} (0) & if\ \ 2^o<\theta\le 20^o, \\ (-1) & if\ \ 20^o<\theta\le 45^o, \\ (-2) & if\ \ 45^o<\theta\le 90^o, \\ (-3) & if\ \ 90^o<\theta\le 180^o, \end{cases} & \end{cases} \qquad \textbf{-----(1)}$$

$$(1\le i\le a) \quad (1\le j\le b)$$

$$2\ MS(q_h,t_h) = Max\,\{Mot(\overrightarrow{v_{q_{hi}}}, \overrightarrow{v_{t_{hj}}})\,|\,1\le\ j\le b\}$$

Table 1. Correspondence relation between q'_h and t'_h

Serial vectors of $'_h$	**A**	**B**	S	F	H	**H**	C	B	**C**	**B**
Timer of q'_h	**T1**	**T2**	T2.4	T3	T4	**T5**	T6	T6.5	**T7**	**T8**
	⇳	⇳	⇳	⇳	⇳	⇳	⇳	⇳	⇳	⇳
Timer of t'_h	**T1**	**T2**	T3	T4	**T5**	T6	**T7**	**T8**	T9	T10
Serial vectors of t'_h	**A**	**B**	S	F	**H**	H	**C**	**B**	C	B

In our proposed DP matching approach (equation 1), the decision of the score is accorded with the range of included angle. If $0^0 \leq \theta \leq 2^0$, the score increases 3, and if $2^0 < \theta \leq 20^0$, the score is invariable. There is the punishment with the score, if $20^0 < \theta^{\,0}$, because the included angle of two vectors between two relevant motion sequence is extended. So when $90^0 < \theta \leq 180^0$, the score deducts 3. We experimented with some combination sets of angle and the value of the score, and obtained the combination with better efficiency as shown in above formula. The similarity between two corresponding motions of Q and T is shown as follow:

$$Similarity_{(M_h)}(Q,T) = MS(q_h, t_h)$$

4.2 Synchronization Similarity

We also need to consider the degree of synchronization between different human motions. For example, we assume that a person raises his right hand first and his left hand next. If the similarity retrieval is accomplished only by the comparison between human body parts' motions, it is possible to find a motion that a person raises both of his hands in the same time. To solve this problem, we use the corresponding relation between two human motions obtained by Dynamic Programming, and calculate the synchronization similarity. To apply the DP matching approach (equation 2), we can find the longest common sequence of q_h and t_h.

Table 1 represents the correspondence relation (the longest command sequence) between q'_h and t'_h. Assume that the longest common sequence is associated by A, B, S, F, H, H, C, B, C and B. Orderly, A, B, S, F, H, H, C, B, C and B represent different vectors in a motion sequence. We can observe the time period of q'_h that is from T1 to T8, and t'_h goes to T10, although they have the same motion sequence. Therefore, we calculate the synchronization similarity to solve this problem. The algorithm is based on Dynamic programming, as shown in equation 2. q'_{hi} and t'_{hj} are the elements of the motion sequence of q'_h and t'_h respectively. T_{qi} and T_{tj} are the timers of q'_h and t'_h respectively. The synchronization similarity between motion q'_h and t'_h is given as the follow:

$$Similarity_{(T_h)}(Q,T) = TS(q'_h, t'_h)$$

$$1.\ n \leftarrow lenght[LCS] \qquad \#\, LCS \text{ is the common sequence of } q_h \text{ and } t_h$$

$$2.\ Timer(q'_{hi}, t'_{hj}) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0, \\ Timer[q'_{h(i-1)}, t'_{h(j-1)}] + 1 & \text{if } q'_{hi} = t'_{hj} \text{ and } T_{qi} = T_{tj}, \text{and } i > 0, j > 0 \\ Max(Timer[q'_{h(i-1)}, t'_{hj}], Timer[q'_{hi}, t'_{h(j-1)}]) & \text{if } q'_{hi} \neq t'_{hj} \text{ or } T_{qi} \neq T_{tj}, \\ & \text{and } i > 0, j > 0 \end{cases} \qquad \textbf{-----(2)}$$

$$(1 \leq i \leq n) \quad (1 \leq j \leq n)$$

$$3.\ TS(q'_h, t'_h) = Max\,\{Timer(q'_{hi}, t'_{hj} \mid 0 \leq j \leq n\}$$

5 Experimental Results and Comparison

In this paper, we use Skeleton Discrimination Tree to classify the motions into five categories dynamically. By this way, unallied human motions can be filtered. It will raise the recall value and improve the precision. Currently, there are 213 human motions stored in our database. We experimented with a desktop PC of Pentium-4 2.4 GHz. If an input query of human motion is not able to be classified into the corresponding motion type, it will cost about 2 minutes for the object retrieval. Moreover, if a query of human motion that had been classified into its belonging motion type, the average time of retrieval is about 69 seconds. The overall recall rate has risen from 82% to 89%, and the precision is about 86%. We compare the retrieval outcome with the reviews from three professors of physical education. Three professors give the comparison with the help of our graduate students. We also asked fifty undergraduates to evaluate the accuracy. The results are efficient and reasonable.

Figure 3 shows the results of our Human Motion Retrieval System. The action models are the side baseball pitching with left hand. The query action is in the upper left, and the similarities are after with query image. All the similarities are side baseball pitching, but with different time to perform the action. In our system interface, the query and similar objects are represented by the image form in order to decrease the loading of computer. User may click the magnifier-button to see the objects in different viewpoint by navigating the 3D Browser.

In the field on 3D human motion analysis and retrieval issues are new challenges; therefore, it is difficult to find highly related work. A few related technique can be found in [10][11]. The discussion in [11] experimented on two types of usability tests for ten subjects using a desktop PC of Pentium-4 2.4 GHz. The first test evaluates the efficiency in retrieval by measuring the time and the number of trials spent in detecting a requested motion segment. Experimental data set consists of ordinary and task motions including various moving and manipulating behaviors. The map was learned from 51 data files, from which many motion segments can be clipped. The second usability test employed 17 motion files which are the same motion styles. In [10] and [11], the authors mention the precision rations, but did not describe the recall rations. We compare our approach with [10] and [11] in table 2. Their precision of Motion Map is better than our proposed method. However, we might have the largest amount of experimental data set and definitely the shortest time for computation.

In the field on 3D human motion analysis and retrieval issues are new challenges; therefore, it is difficult to find highly related work. A few related technique can be found in [10][11]. The discussion in [11] experimented on two types of usability tests for ten subjects using a desktop PC of Pentium-4 2.4 GHz. The first test evaluates the efficiency in retrieval by measuring the time and the number of trials spent in detecting a requested motion segment. Experimental data set consists of ordinary and task motions including various moving and manipulating behaviors. The map was learned from 51 data files, from which many motion segments can be clipped. The second usability test employed 17 motion files which are the

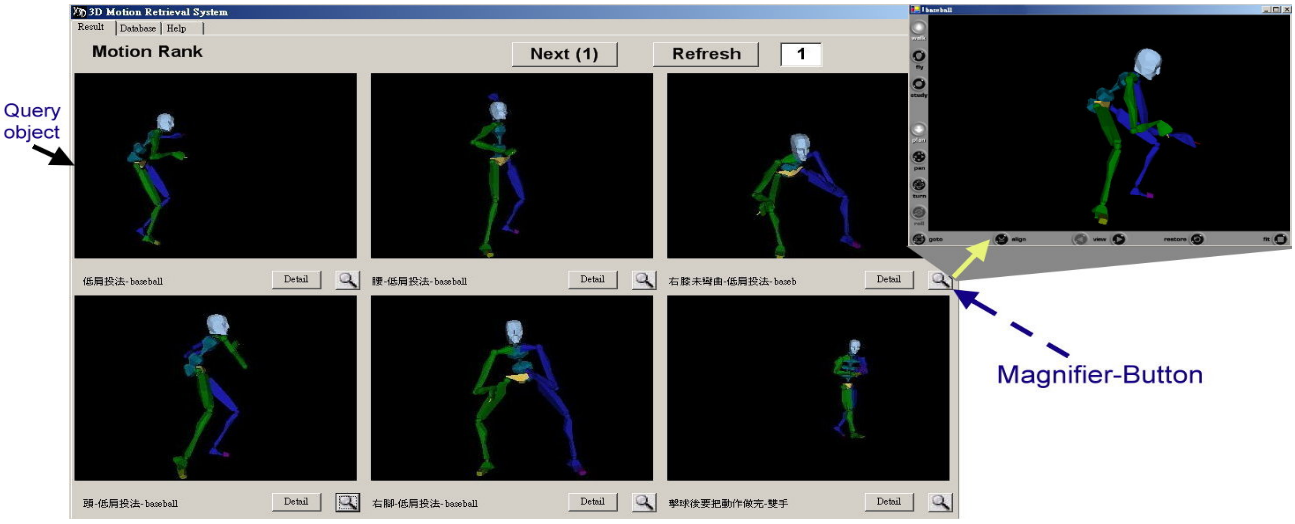

Fig.3. The results of Human Motion Retrieval System

same motion styles. In [10] and [11], the authors mention the precision rations, but did not describe the recall rations. We compare our approach with [10] and [11] in table 2. Their precision of Motion Map is better than our proposed method. However, we might have the largest amount of experimental data set and definitely the shortest time for computation.

Table 2. Comparison with Motion Map and Multidimensional Indexing with Voting

	The Proposed Method	Motion Map [11]		Multidimensional Indexing with Voting[10]
		First type	Second type	
Time	**1.09 (min.)**	**17 (min.)**	**6 (min.)**	--
Precision	**86%**	**86.25%**	**98%**	**84.6%**
Recall	**89%**	--	--	--

6 Conclusion and future works

With respect to the literatures in the Virtual Reality application, human motion retrieval is an interesting but challenging issue. In this paper, we proposed the classification function, which is needed before comparing with the trajectories of each human body joints. Besides, we used a series of vectors as the feature information to compute the similarity which is based on mutative Dynamic Programming. We calculated the similarity between different persons' motions and the degree of synchronization among the discriminative parts of body motions. By our proposed 3D human motion retrieval system, animation designers could find the animation objects and reuse them efficiently to facilitate the 3D games and the movie industry. It can level down the cost of production and time. Sport players can improve their skills by using our system as well. In the near future, we will extend the system to solve the complexity kinematical motions, and investigate the training scheme on estimation of skill levels in sports. This will lead to drill the well-trained player in sports.

References:

[1] Dejan V.and D. Saupe, 3D Model Retrieval, *Proceedings of Spring Conference on Computer Graphics 2000*, Comenius University Press, Bratislava, Slovakia, May 2000, pp. 89-93.

[2] Motofumi T. and Suzuki, A similarity retrieval of 3d Polygonal Models Using Rotation Invariant Shape Descriptors, *IEEE International Conference on Systems, Man, and CyBernetics(SMC2000)*, 2000, pp2952-2956.

[3] I.A. Karaulova, P.M. Hall, A.D. Marshall, A hierarchical model of dynamics for tracking people with a single video camera, *British Machine Vision Conference*, 2000, pp.352–361.

[4] I.-C. Chang, C.-L. Huang, Ribbon-based motion analysis of human body motions, *Proceedings of the International Conference on Pattern Recognition*, Vienna, 1996, pp.436–440.

[5] J.K. Aggarwal, Q. Cai, W. Liao, B. Sabata, Non-Rigid motion analysis: articulated & elastic motion, *Comput. Vision Image Understanding* 70 (2), 1998, pp. 142–156.

[6] R. Polana, R. Nelson, Low level recognition of human motion, *Proceedings of the IEEE CS Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, TX, 1994, pp.77–82.

[7] Takeshi Yabe, and Katsumi Tanaka, Similarity Retrieval of Human Motion as multi-stream time Series Data, *Proceeding of International Symposium on Database Applications in Non-Traditional Environments (DANTE'99)*, 1999, pp. 279-286.

[8] M. Brand, N. Oliver, A. Pentland, Coupled Hidden Markov models for complex action recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 994–999.

[9] J. Yamato, J. Ohya, K. Ishii, Recognizing human action in time-sequential images using Hidden Markov model, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1992, pp. 379–385.

[10]Ben-Arie, J., Wang, Z., Pandit, P. and Rajaram, S., Human Activity Recognition Using Multidimensional Indexing, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 24 No. 8, pp. 1091-1104, August 2002.

[11]Yasuhiko Sakamoto, Shigeru Kuriyama and Toyohisa Kaneko, Motion Map: Image-based Retrieval and Segmentation of Motion Data, *Eurographics/ACM SIGGRAPH Symposium on Computer Animation*, 2004, pp.259-266.