On Automatic Actions Retrieval of Martial Arts

Timothy K. Shih[1], Ching-Sheng Wang[2], Yuan-Kai Chiu[1], Yi-Tsou Hsin[2], and Chun-Hong Huang[1]
[1]Department of Computer Science and Information Engineering
Tamkang University, Taiwan, R.O.C.
tshih@cs.tku.edu.tw
[2]Department of Computer and Information Science
Aletheia University, Taiwan, R.O.C.
cswang@email.au.edu.tw

ABSTRACT

Martial art actions can be represented via VRML animations or extracted by video tracking. We propose an action retrieval method, which allows users to retrieve similar actions of martial arts. The mechanism is based on a similarity function that compares animation tracks. A representation of human skeleton includes head, knee, elbow, wrist, etc further aggregates important features in martial art actions. Different weights are dynamically calculated according to motion sensitivity of feature points. As a result, the system can automatically retrieve similar martial art actions. The results are tested by professional kung fu master with a good satisfaction.

Key words: VRML, animation, automatic action retrieval, virtual reality, martial art

1. INTRODUCTION

Automatic retrieval of actions in 3D space is a challenge but useful technique. Examples of behavior understanding of video can be found in [1, 2]. Motion tracking and recognition of human interactions by a multi-layer finite state machine is presented in [1]. By using body pose vectors, human action recognition is presented in [2]. With recent Virtual Reality technologies, movies can be made my VR-based or Augmented Reality-based actors. Retrieval of actors in a 3D scene become a useful technology [3, 5], if animated actors in a scene database is to be reused [4]. In stead of using user pre-defined metadata [4], 3D animations should be retrieved based on the existing animation models. We look at one particular domain as an example showing our contribution and its possible extensions. Martial arts can be represented as VRML animations. A particular martial art, known as the Bar-Chi Spar, is our target application. Bar-Chi Spar was originated in Ho-Bai, China, in around 1368. The unique basic action of Bar-Chi Spar is clear, fast, powerful, and smooth. According to our Bar-Chi Spar guru, the use of head, shoulder, elbow, pud, tail, crotch, knee, and foot are the features which represent different sets of spar actions. Assuming that these features can be extracted from a VRML model, which teaches Bar-Chi Spar, it is possible to record the animation tracks of these important portions of a human body and to save it for comparison. Thus, an automatic retrieval system tells the user which set of spar actions is similar to the one he/she is learning.

In order to compare the animation tracks of different feature points, a normalization technique is required since different VRML actors may have different highs and weights. Also, animation tracks may have different lengths. We present the normalization of action tracks in section 2. A skeleton is required to represent human body, which is presented in section 3. A distance function which aggregates feature points in Bar-Chi Spar is presented in section 4. Our system is implemented using the 3D Studio Max and the Cortona VR Player. We compare the retrieval outcome with the reviews from three kung fu masters in section 5, before our conclusion section is presented.

2. NORMALIZATION OF ACTION TRACKS

Object animation tracks are not necessary represented with the same amount of tracking points. It is necessary to normalize animation tracks before we use these tracks in a similarity function. One way to normalize number of tracking points in two tracks is to add interpolation points. According to 3D geometric, three feature points form a circle in the three dimensional space. With a limited granularity, it is reasonable to use interpolated points on the circle as an approximation of an animation track. Assuming that we have three points P1=(x1,y1,z1), P2=(x2,y2,z2), and P3=(x3,y3,z3). Also, let point (X, Y, Z) represents the center of the circle. We can form the following three equations:

$\alpha_1=x_1-x_2$, $\beta_1=y_1-y_2$, $\lambda_1=z_1-z_2$, $\alpha_2=x_2-x_3$, $\beta_2=y_2-y_3$, $\lambda_2=z_2-z_3$

$$\alpha_1 * X + \beta_1 * Y + \lambda_1 * Z = C_1 \quad \ldots\ldots(1)$$
$$\alpha_2 * X + \beta_2 * Y + \lambda_2 * Z = C_2 \quad \ldots\ldots(2)$$
$$(\beta_1*\lambda_2 - \lambda_1*\beta_2)*X + (\lambda_1*\alpha_2 - \alpha_1*\lambda_2)*Y + (\alpha_1*\beta_2 - \beta_1*\alpha_2)*Z = C_3 \quad \ldots\ldots(3)$$

where C_1, C_2, and C_3 are constants. The first two equations represent two plans based on lines {P1, P2}, and {P2, P3}, respectively. Equation (3) represents a plan from the three points, P1, P2, and P3. By substituting

0-7803-8603-5/04/$20.00 ©2004 IEEE

$(x_1+x_2)/2$, $(y_1+y_2)/2$, and $(z_1+z_2)/2$ for X, Y, and Z, respectively, in equation (1), we have $C_1=[(x_1^2+y_1^2+z_1^2) - (x_2^2+y_2^2+z_2^2)]/2$. Similarly, we have $C_2=[(x_2^2+y_2^2+z_2^2) - (x_3^2+y_3^2+z_3^2)]/2$. Substituting x_1, y_1, z_1 for X, Y, and Z in (3) yields $C_3 = (\beta_1 * \lambda_2 - \lambda_1 * \beta_1) * x_1 + (\lambda_1 * \alpha_2 - \alpha_1 * \lambda_2) * y_1 + (\alpha_1 * \beta_2 - \beta_1 * \alpha_2) * z_1$. With the values of C_1, C_2, and C_3, we use Gauss elimination to solve the above three equations.

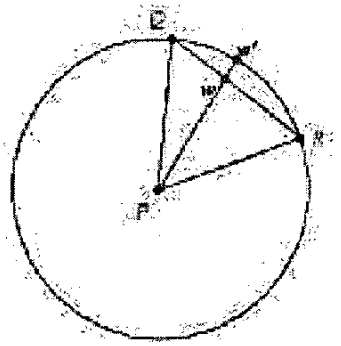

Figure 1: Adding an Interpolation Point *w'*.

To find an interpolated point on the circle, as illustrated in figure 1, we find the value of point *w* first. Since *w* is on a line constructed by points *Q* and *R*, we have $w=(1-\lambda)*Q + \lambda*R$, where $\lambda = [0..1]$. Assuming that *r* is the radius of the circle (i.e., $r = \|PQ\|$, or the distance between *P* and *Q*), we have $Pw' = Pw / \|Pw\| \cdot r$. Thus, the interpolated point *w'* is obtained. Note that, the interpolated points can be multiple. The number of interpolated points depends on the length between two tracking points, as well as the length of track. The normalization procedure takes two steps. The first step add interpolated points among tracking points such that two tracks to be compared result in the same number of points (interpolated points plus tracking points). The second step normalizes the lengths of tracks. The PCA mechanism applied to the bounding boxes of the two action tracks is used.

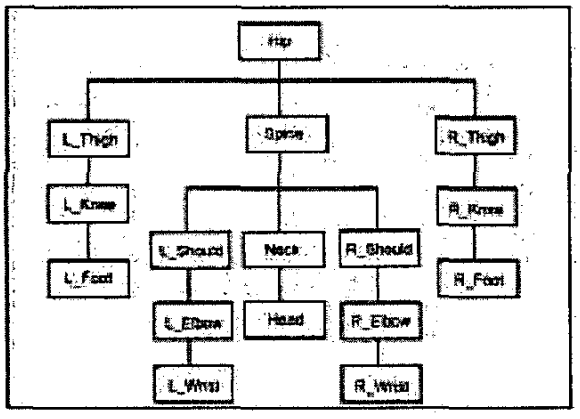

Figure 2: A Human Body Skeleton

3. REPRESENTATION OF SKELETON

A body skeleton for action tracking can be represented by 16 feature points, as illustrated in figure 2. These feature points are constructed as the first representation of VRML object, which stores a Bar-Chi Spar. In figure 3, each of these 16 feature points are represented. The second representation of a VRML object draws the avatar (shown in figure 5). In a practical situation, some feature points are less significant as compared to some points with a high momentum. For sack of computation efficiency, some of the feature points are omitted. In addition, momenta are computed based on portions of a human body. For instance, left thigh, left knee, and left foot are considered as an aggregation. In our model, we estimate the momenta of the following four portions:

Left Hand = left shoulder + left elbow + left wrist
Right Hand = right shoulder + right elbow + right wrist
Left Leg = left thigh + left knee + left foot
Right Leg = right thigh + right knee + right foot

According our Bar-Chi Spar master, the omitted features points (i.e., hip, spine, neck, and head) does not play important roles in spar actions.

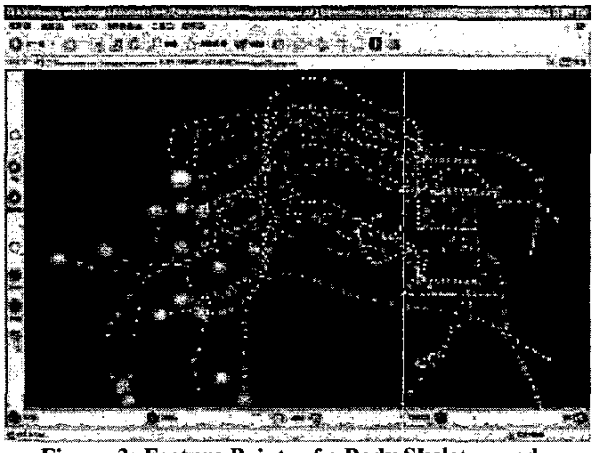

Figure 3: Feature Points of a Body Skeleton and Animation Tracks (with Interpolation Points)

4. SIMILARITY OF SPARS

The implementation of our system requires a VRML parser, which decompose objects of a spar avatar into different sections. We only consider coordinates and sizes. Color information and appearance (i.e., material) are not used. Coordinate information include rotation and translation. It is necessary to convert rotation information into relative coordinates, with size information normalized in a post process. According to the skeleton, rotations include multiple origins. For instance, the rotation of left knee is based on the rotation of left tight. Assuming that

p_1 represents the feature point of left thigh and p_2 represents the left knee, and $p_1=(x_1,y_1,z_1)$, $p2=(x_2,y_2,z_2)$. We can use:

$$R=\begin{vmatrix} a^2(1-\cos\theta)+\cos\theta & ab(1-\cos\theta)-c\sin\theta & ac(1-\cos\theta)+b\sin\theta \\ ab(1-\cos\theta)+c\sin\theta & b^2(1-\cos\theta)+\cos\theta & bc(1-\cos\theta)-a\sin\theta \\ ac(1-\cos\theta)-b\sin\theta & bc(1-\cos\theta)+a\sin\theta & c^2(1-\cos\theta)+\cos\theta \end{vmatrix}$$

where (a, b, c, θ) represents the rotation vector and angle. This yields $p_2'=R \bullet p_2 + p_1=(x_2'+x_1, y_2'+y_1, z_2'+z_1)$. And p_2' is the final coordinate. After rotation information is converted, spar avatars contain only relative coordinates. These coordinates are consecutive points which can be subdivided into series of vectors. Assuming that each of the two vector series of two avatars contains a pair of points, (p1,p2) and (p1',p2'), and $p_1=(x_1,y_1,z_1)$, $p_2=(x_2,y_2,z_2)$, $p_1'=(x_1',y_1',z_1')$, $p_2'=(x_2',y_2',z_2')$. Let

$V = (Vx = x_2-x_1, Vy = y_2-y_1, Vz = z_2-z_1)$
$V' = (Vx'=x_2'-x_1', Vy'=y_2'-y_1', Vz'=z_2'-z_1')$

The difference of the two vectors is $d = |Vx-Vx'| + |Vy-Vy'| + |Vz-Vz'|$. When we compare the differences between two spar tracks, the differences of vectors are accumulated. To compare two Bar-Chi Spar actions, not only the tracks of feature points are considered. The significance of each feature point is very important, according to our kung fu guru. As we mentioned before, we disregard some feature points such as hip and spine. And, we treat hands and legs as aggregated units. This is due to the fact that, shoulder, elbow, and wrist are naturally connected. Thus, we differentiate the degree of momenta among left hand (LH), right hand (RH), left leg (LL), and right leg (RL). The first approach of our similarity function which compares two Bar-Chi Spar actions only relies on adding weights to the above 4 aggregated units. 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 spar and a target spar. We further enhance the accumulated similarity measure as:

$$d = n_1 * RH_d + n_2 * LH_d + n_3 * RL_d + n_4 * LL_d$$

The weights should be arranged in a non-increasing order (i.e., $n_1 >= n_2 >= n_3 >= n_4$). Note that, a default setting in our system is $n_1 = 4$, $n_2 = 3$, $n_3 = 2$, and $n_4 = 1$. However, the weights can be adjusted by a kung fu master. Another approach to enhance the significance of feature points relies on the momenta as well. However, the relation of aggregated units in a body skeleton is also considered. We propose a *skeleton discrimination tree*, which is constructed for each query action. For instance, if $LH_d > RH_d > LL_d > RL_d$, the tree is show in figure 4. 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 include a case $RH_d > LH_d > LL_d > RL_d$, which has distance 3 from category A. In addition, a special condition (shown as "Special") is used with different weights in the tree. Note that, there are 24 cases based on the momenta. Each case has its discrimination tree. The result is used in weight calculation (i.e., the use of n_1, n_2, n_3, and n_4). If two spars are in the same category, the weights are the same. Otherwise, the weights are decided by the distance (i.e., *dis* in table 1). The weights are added according to the order of body portions in the query action. Portions of the same kind in both the query and the target actions use the same weights for enhancement. The effect will enlarge the differences of potions with larger momenta, which are important according to our kung fu master.

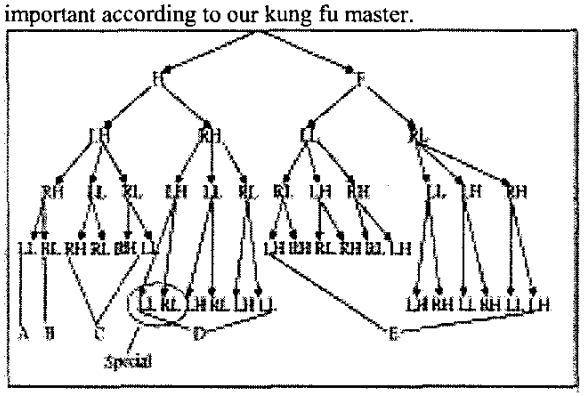

Figure 4: Skeleton Discrimination Tree

Table 1: Weights for Similarity Calculation

Groups *dis* to A	A (*dis* =0)	B (*dis* =1)	C (*dis* =2)	D (*dis* =3)	E (*dis* =4)
n_1	1	2	3	4	5
n_2	1	1	2	3	4
n_3	1	1	1	1	1
n_4	1	1	1	1	1

5. EXPERIMENTAL RESULTS AND ANALYSIS

We constructed 16 Bar-Chi Spar actions. Each spar is used as a query to compare the similarity with other spars. However, we divide the rest 15 spars into three groups for an easier comparison by human. A query spar is compares to 5 spars in a group, with results range from 1 (more similar) to 5 (less similar). Three Bar-Chi Spar gurus give the comparison with the help of our graduate students. The average similarities from the three experts are shown in table 2. The results of similarity distance from computer are shown in table 3 (see an example in figure 5). If the momenta of the 4 body portions have the same weight, the results are less realistic. If we use the default weights (i.e., $n_1 = 4$, $n_2 = 3$, $n_3 = 2$, and $n_4 = 1$), the results are acceptable. Finally, the adjustable weights in table 1 yield a best solution (with actual distance between spars shown in table 3).

Table 2: Similarity Averages from 3 Kung Fu Gurus

Ave.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		1.7	3.7	4	1.3	4.3	5	4	1.7	2.67	1.67	1.33	1.67	4.33	4	3.67
2	1.7		3.3	4.3	1.3	4.3	4.7	3.7	1.3	3.67	1.67	1.67	1.67	4	4.33	3.33
3	4	3.7		2.7	2.3	2.3	1.3	1.7	3	5	4	3.67	3.33	3	4	1
4	3.3	4.7	3		2.3	1.7	3	2	3.3	4	2.67	2	3.33	3.67	2.67	3.33
5	2	1	4.7	3.7		3.7	4.7	3.7	1	3.67	2	2	1	4.67	4.33	3
6	3.7	3	2.3	2.7	3.3		2.3	2.7	3	4.33	2.67	2	3.67	2	3.67	3.67
7	4.3	3	1	4	2.7	3		1	3	3.67	4.33	3.33	4.33	1.67	4.33	1.33
8	4.3	2.7	1	4	3	3	1		2.7	4.33	4	3	3.67	3.67	3.67	1
9	2.7	1.3	4.7	4.3	2	4.7	3.3	2.3		3.67	1	2.33	1	4.67	4.33	2.67
10	1.3	2.7	4.3	4.3	2.3	5	3.3	2.3	2.3		2	1	3	4	3.33	3.67
11	3.3	2	3	4.7	2	4	4	2.7	1.3	3		2	1	4.33	4	3.67
12	1.7	2.7	4	5	1.7	4	3.7	2.3	2.7	2.33	2.33		2.67	3.67	3.33	3
13	3	1.7	4.3	4.7	1.3	4.7	3.7	2.7	1	3	2	1.33		4	4	3.67
14	2.3	4.3	2	2.3	4	1.3	1.7	3	4	5	4	2	3		4	2
15	3	3.3	3.7	2.3	2.7	4.3	4	3	2	1	1.33	3	4	3.33		3.33
16	4.3	2.7	1	4	3	3.3	1.3	1.7	4	4.67	3.33	2.67	2.33	3	3.67	

Table 3: Similarity Distances from Computer

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	5216	11303	5532	5252	12107	10429	10286	5940	4880	4533	4945	4931	18716	6293	10450
2	5548	0	8161	4368	1505	13642	7519	7327	4078	3858	4284	4104	1548	17468	5432	7540
3	12171	10472	0	10855	10452	17007	5952	6674	10320	10323	11478	10818	10254	20708	11326	5617
4	6115	4363	8687	0	4368	13166	8176	8073	6009	5282	4677	4970	3905	18925	5502	8176
5	5568	1457	8026	4364	0	13495	7385	7193	4054	3460	4101	3927	1138	17008	5106	7406
6	12083	13541	15574	13097	13429	0	15126	15171	13332	13636	13196	13667	13057	22340	13285	15134
7	11731	9702	6067	10330	9682	16680	0	6370	9398	9454	10840	10092	9484	20724	11045	3862
8	11519	9249	6721	9962	9228	16264	6302	0	9030	9378	10349	10033	9030	19969	10811	5570
9	6341	4108	8463	5782	4096	12373	8325	8027	0	4673	5324	4962	3481	16040	5712	8404
10	4595	3659	8258	5086	3368	12625	8014	7968	4646	0	4559	2504	2960	16799	4089	8111
11	4562	4310	9554	4545	4175	12178	9509	9155	5283	4551	0	4099	3566	17613	5300	9542
12	4861	3445	10124	4408	3403	13254	8935	8886	4946	2667	4032	0	3185	17728	4077	9075
13	4985	1555	7292	3762	1158	13168	6651	6459	3251	2925	3296	3064	0	17299	4328	6672
14	18345	17333	20322	18695	16978	22377	20221	19884	16640	17362	18008	17742	17249	0	19159	20291
15	6154	5330	10865	5011	5251	13112	10247	10143	6592	5121	6070	4055	5054	19486	0	10327
16	11778	9835	5774	10382	9815	16764	3912	5708	9607	9652	10919	10285	9617	21053	11202	0

6. CONCLUSIONS

In the Virtual Reality literature, it is hard to find similar research for comparing human actions, especially in martial arts. The most difficult challenge of our research is on the design of a realistic similarity function between actions. With the dynamic adjustment of weights between different momenta of human skeleton, our system performs reasonably. Yet, the results of similarity comparison may have drawbacks. For instance, if actions in a set of spars only have small differences, the comparison results may show failures.

A few issues are listed in our future work. Firstly, since the movements in Bar-Chi Spar are relevant to the center of a human body, it is easier to normalize the scale of momenta. But, in other application domains (e.g., ice skating, jump, skiing, etc), redundant track portions should be eliminated. However, to decompose a series of actions and to extract key features are difficult since different sports might have different focuses. Another future direction of our work is to incorporate multiple spar players. Spatial and temporal relations are used to estimate the interactions. Finally, with object extractions in video technology, it is possible to analyze spar actions by real persons.

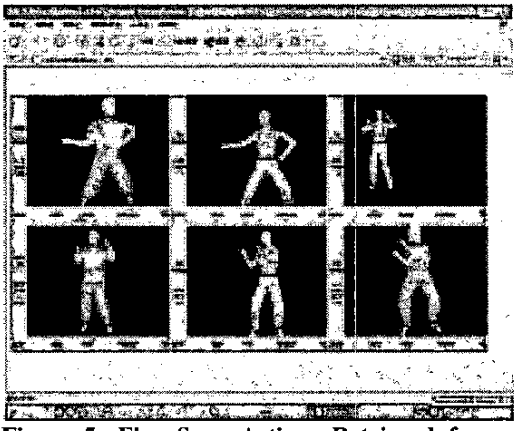

Figure 5: Five Spar Actions Retrieved from a Query (the Upper-Left VRML Viewer)

REFERENCES

[1] Sangho Park, Jihun Park, and Jake K. Aggarwal, "Video Retrieval of Human Interactions Using Model-Based Motion Tracking and Multi-layer Finite State Automata", CIVR 2003, LNCS 2728, pp. 394–403, 2003.

[2] 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.

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

[4] Akanksha, Huang Z., Prabhakaran B. and Ruiz, Jr. C. R. "Reusing Motions and Models in Animations", Proceedings of EGMM 2001, pp. 11-22, 2001.

[5] C. Zhang and T. Chen "Indexing and retrieval of 3D models aided by active learning" Proceedings of the ninth ACM international conference on Multimedia, Ottawa, Canada, pp. 615 – 616, 2001.