Competitive motion synthesis based on hybrid control

By Zhang Liang, Jun Xiao*, Yueting Zhuang and Cheng Chen

We propose a simple and effective framework to deal with the problem of synthesizing interactive and competitive motions while reflecting the interactions. Two nontrivial issues are addressed in synthesizing two-character motions in competitive environment: how to reveal the embedded routines while keeping visual reality and how to build interactive models based on singly captured motions. To solve these issues, we employ a hierarchical framework: the finite state machine (FSM) controls the state transition in the higher layer, and the hybrid approach controls the action selection in the lower layer. The proposed approach contains two folds: first, a rule-based control scheme is proposed to simulate routine steps based on statistical analysis. Second, the interactive models are designed for simulating dense interactions between two players. The Relevance Vector Machine (RVM) algorithm is adopted to select attack styles and coupled with motion transition graph to determine combination blows. Here we apply the proposed framework of hybrid paradigm to boxing sport as an example. Copyright © 2009 John Wiley & Sons, Ltd.

Received: 25 March 2009; Accepted: 25 March 2009

KEY WORDS: character animation; motion capture and reuse; hybrid approach; RVM

Introduction

Interactive human motion is prevalent in movies and games. These competitive interactions are observed in various forms including antagonistic sports like boxing and kickboxing, kung fu, and taekwondo. Scenes of such forms which stress on compact and tit-for-tat interactions are conventional in computer games and character animations.

Compared with dancing motion, the generation of antagonistic sports makes remarkable differences. Such fighting sports are performed in an asynchronous manner. Not like coupled dancing in which one leads while another follows, the fighter tries to fool the opponent and seek the opportunity for a strike or prepare for a counterstrike. The key description of such asynchronous interactions should be “Against,” as the fighter dynamically schedules successive motion against the opponent both in time and space.

Comparing to creating and editing competitive motions manually, motion capture driven character takes advantage in efficiency. However, capturing the motion of multiple subjects together is a tough task in either optical or magnetic motion capture systems. The difficulty comes from track missing caused by self- or inter-occlusion and marker displacement derived from the interaction such as punch or rub. For example, our experiment shows that the markers attached to head and chest are liable to drift from original positions when the attack body hits or rubs their surface.

In this paper, we employ a hierarchical framework to solve the problem of simulating competitive motions based on singly captured motions. Our objectives are twofold: reflecting the complex interactions between two players while preserving the inherent routine manners. The outline of our approach is shown in Figure 1. It is composed of four steps: (1) Capture the motion of a single person using a motion capture system. (2) Create an FSM controller to determine the state transition in state layer, where the state set contains three states: balance, attack, and defend. (3) Generate a hybrid approach which combines empirical rules and interactive models to choose appropriate action in action.
layer. (4) Provide with scene and initial parameters, the process generates a variable-length animation of two-player competitive fighting automatically by stitching motion clips.

We demonstrate the effectiveness of our approach through the simulation of two-character fighting in boxing environment. The key contributions can be categorized as two issues:

1. A hybrid paradigm is adopted to control the motions of two players. The combination of empirical rules and probabilistic models exploits advantages of both approaches; exhibiting the embedded routine manners while keeping natural-looking.

2. We establish the interactive models based on supervisor training algorithms to capture and reflect the dense interactions. With tolerable manual interactions, the acceptance of simulated results is proved to reach an acceptable and durable level.

The paper is organized as follows. In the next section, we first review previous related work. Then our approach is elaborated in the subsequent sections, we first present the FSM adopted in the first layer, then we detail the empirical rules and interactive models separately. The results are shown in our experiment with detailed discussions. Finally, we conclude this paper with some future research topics in the last section.

**Related Work**

In the synthesis of two or multi-player character animation, the interactions between characters are extremely important as they determine whether the simulation seems nature-looking and reasonable. We review the related works as follows.

Shapiro and Pighin implemented a hybrid framework by combining kinematic animation with physical simulation for the reason of developing each advantage. They build a layered framework of state controllers for handling the transitions between kinematic and physical situation. Zordan and Hodgins incorporated physical constrains into data-driven animation for synthesizing reactive character upper-body motions to various impacts by the opponent. They adopted their system in boxing and table tennis. Cooper et al. employed an active learning method to refine the motion controller upon user selected tasks during the data acquisition process. After training, only small amount of motion capture data are required for yielding such controllable controllers which aim for special tasks. The methods as we listed in this category did not deal with mutual interactions between multi-players, and they only concentrated on solving the problem of reaction to user controls or environment constrains.

Zordan et al. combined the motion capture data with physical simulation which responds to unexpected impacts using specialized search routine. Liu et al. adopted an optimal approach to solve the problem of external constrains and interactive constrains in multi-character motion synthesis. The authors here did not refer the field of persistent and mutual interactions between multi-players.

Hsu et al. presented a method to solve the synchronous manner interaction by generating the synthetic motion of dancers. Based on motion capture data, the motion of one dancer was used as a control signal for the other one to search for an example motion segment from databases. But unfortunately, this approach could not be
applied to our concerned competitive interactions as they are asynchronous both in time and space.

Kulpa et al.7 designed a motion representation which is independent from morphology and implemented CCD algorithm to deal with multi-character interactions with the representation in real-time. Lee and Lee9 presented a motion graph of current and target states for a motion set, and they adopted dynamical programming to pre-compute the transition action between the pairs of states in the graph. In synthesis, the controller searched the transition list to find the adaptive interaction for multi-characters in real time.

Kwon et al.9 incorporated the conception of primitive actions and modeled the transitions between the groups which contain similar actions based on probabilistic models. The final synthesis process was guided by trained transition model. But in their scheme, the training process of the probabilistic model was highly dependent on example motions.

Shum et al.10 presented a FSM mechanism called “Interaction Graph” which records the meaningful dense interactions from a bulk of samples. At run-time, the target motion of each character was searched by min–max or dynamical programming on the interaction graph. However, the authors concentrated more efforts on flat action transition without considering the routine rationality of transition.

State Transition

In our system, the controller of state transition is modeled by finite state machine (FSM). We separate an integrated boxing simulation into a series of basic semantic clips which are called “state cells”. The action of each player in a cell belongs to one of the states. When the player has finished the current state, the target state must be selected from the state set for the next cell, and the Markov chaining provides the approximate mechanism. Markov chaining favorites the state transition that simultaneously maximizes both the per-cell match to the observation and the temporal correlativity between successive cells.11 The optimized process can be formulated, for each player, as

\[ E = \sum_{i=0}^{n} \psi(\Phi_i, M_i) + \omega \sum_{i=1}^{n} P(\Phi_i|\Phi_{i-1}) \]  

where \( n \) is the number of state cells in the simulation result, \( \psi \) evaluates the matching error between the state features \( \Phi_i \) and the observation \( M_i \) corresponding to the state cell \( i \).

One potential choice for \( \psi \) is the energy function used to estimate the similarity for comparison. Based on global coordinate system, here we incorporate two features into the definition of \( \psi \): \( \Delta \text{dis}_i \) is the difference of distances of two players’ center of masses (COMs) at the initial and end of a cell, and \( \Delta \text{vel} \text{COM} \) is the average velocity of the player’s COM during the cell period. Based on such two features, the comparison provides greater sensitivity to the state correspondence. Thus, the chaining function uses \( \psi \) in following form:

\[ \psi(\Phi_i, M_i) = \psi(\Delta \text{dis}_i, \Delta \text{dis}_M) + \psi(\Delta \text{vel} \text{COM}, \Delta \text{vel} \text{COM}_M) \]  

where \( \Delta \text{dis}_M \) and \( \Delta \text{vel} \text{COM}_M \) corresponding to the observation \( M_i \) represent the prior average of \( \Delta \text{dis} \) and \( \Delta \text{vel} \text{COM} \) which are calculated from assistant method, and the function \( \psi \) is the Euclidean distance.

Empirical Rules

Statistical Data

Our method exploits the distances of two players’ COMs and orientation as the observation features. To guarantee realistic, we extract these features from real boxing videos using computer vision techniques. The camera is calibrated by exploiting the fact that the actual size of the box court is fixed. The four edges of box court are detected by Hough Transform, and the corners of the court on the image plane can be determined. After this calibration, we are able to calculate the world coordinate of any point on the image. We then conduct background subtraction in the image to get the silhouette of two players, and calculate the centroid of each player in the world coordinate frame. In this way, we get the distance of two players’ centroids. The relational orientations between the players are manually designated by the user.

Rule-based Decision

Generally, in real boxing scene, player seeks for the attack opportunity by means of changing their positions...
at a deliberate pace and probing at times. On the other hand, for the passive defender, the instinctive response is greatly dependent on the reactive time which is related with distance. So the distance is one of the dominant features in establishing rule set. But there comes a new problem: how to build a quantitative model which reveals the embedded relationships between action selection and distance changing? We choose a divide-and-conquer strategy and build classification rules separately for each state.

**Attack Rules.** When the opponent has entered into the attack range, the player will choose a suitable attack style according to current situation. For example, excepting the handed habit and empirical factors, if the distance between two players is large, players are apt to choose straight punches or swing blows, with a reluctant selection of hooks. The reason is natural in the physical perspective: such long-range attack works by stretching the arms completely while the hook is suitable for short-range attack as the arms keep bending. According to physical features of each attack style, the directive rules are generated to guide action selection in attack state.

We group the major attack styles into three classes, far, middle, and near. Distance serves as determinant as it provides attack style compatible range. Two breakpoints \( \theta_{\text{Far-Middle}} \) and \( \theta_{\text{Middle-Near}} \) separate various attack styles into three candidate sets. It is not the end, as provided with the candidate styles, the player confronts a puzzled situation that which style should be selected and which one should be discarded? This will be further detailed in the interactive models.

**Balance Rules.** According to previous boxing experiment, 50–60% of the total time in boxing is taken up by players’ moving and probing. Moving provides the player a manner to observe the competitive situation for attack launching. Cooperated with moving, probing is used in a deceptive and tentative way to fool the opponent. Intuitively, we want to figure out the embedded rules of distance changing in state balance.

Based on the assistant method, we sample the distance of 55 time stamps between the players when they are in the state balance in a real boxing video, and create the corresponding histogram. As we find that the histogram exhibits the normality to some extent. We compare the distribution probability of samples with normal distribution probability to prove this point, and approximate the distance distribution in normal format:

\[
x \sim N(\mu, \sigma^2)
\]

where \( x \) represents the distance and the parameter pair \( < \mu, \sigma > \) is derived from experiment with values of \( <2.45, 0.20> \) here.

**Defend Rules.** The selection of defensive action depends on many factors, but as we demonstrated, there comes three significant ones: the blow speed, the distance, and the attack manner. Given the general accepted blow speed as 40 m/seconds, only two factors distance and attack manner are involved as parameters in establishing the defend rules.

Obviously, a long distance benefits the defender as it provides sufficient reaction time indirectly. Conversely, when the players get entangled with each other, the distance will not be sufficient for dodging or jumping. The attack manner also takes crucial effect due to their physical features elaborated above, and will bring distinct reactive times. According to the analysis, we create our physical-based rules as follows.

First, the reactive score \( \delta_i \) for the defender \( i \) can be model as

\[
\delta_i = e^{-\phi} \tag{4}
\]

This reactive score scales the emergency of response in a quantitative norm. The exponential format derived from previous medicinal experiments simulates the human reactive characteristics by incorporating a constant value \( \phi \) which denotes the blind zone of human nervous system.

Then, the spend time \( t \) of attack style is estimated in simple kinematics form

\[
t = \sqrt{\frac{2S}{a}} \tag{5}
\]

where \( S \) represents the distance the blow travels and \( a \) represents the average acceleration. The distance is defined as the norm of fist trajectory from the attack launching and the hitting. For example, the travel of straight punch is close to a straight line while the term of hook is close to a circular arc curve. All these travels can be calculated approximately in mathematical forms which will not be expanded here. Figure 2 shows the relationship between \( \delta \) and \( t \). Three defensive selections with corresponding reactive score ranges are labeled in the figure, and in our experiment, the blind constant is set as 0.30 seconds. The masked rectangle represents the blind zone for response.
Interactive Models

In this section, we describe how to simulate the various types of interactions between the players. More specifically, provided with coupled motion samples and user annotations, we explain how to model three conventional interactions by different probabilistic approaches.

Attack Selection

After the pruning unreasonable attack styles by the empirical rules, the following task is to determine the most probable style from the candidate set according to the current competitive situation, which is considered as high dimensional optimization process.

To achieve this goal, the regression method based on the conditional descriptors is designed and illustrated in this section. Here we represent the features describing the current situation as conditional descriptors by vector \( x \in \mathbb{R}^d \), and represent the output of attack style choice by vector \( t \in \mathbb{R}^m \). The relation between \( x \) and \( t \) is regressed using a set of basis functions

\[
t = \sum_{i=1}^{r} w_i \phi_i(x) + \varepsilon \equiv W\Phi(x) + \varepsilon
\]

where the set \( \{\phi_i(x) | i = 1, ..., r\} \) are the basis functions, \( w_i \) is \( \mathbb{R}^m \)-valued weighting vectors, and \( \varepsilon \) is a residual error vector. The weighting vectors are sorted into a weighting matrix \( W = (w_1, w_2, ..., w_r) \) with \( m \times r \) size, and the basis functions are sorted into a \( R^r \)-valued function \( \Phi(x) = (\phi_1(x), \phi_2(x), ..., \phi_r(x))^T \).

A trivial set of examples annotated by users \( \{(t_1, x_1), ..., (t_n, x_n)\} \) is needed for supervised learning, and Euclidean norm is used to measure the t-space prediction errors, so the optimized function here is of the form

\[
W := \arg \min_W \left\{ \sum_{k=1}^{n} ||W\Phi(x_k) - t_k||^2 + R(W) \right\}
\]

where the function \( R \) is a regularizer based on matrix \( W \). For clarity, we simplify Equation (7) by gathering the training data into a \( m \times n \) output matrix \( T = (t_1, t_2, ..., t_n) \) and a \( r \times n \) basis function matrix \( \Phi = (\Phi(x_1), \Phi(x_2), ..., \Phi(x_n))^T \):

\[
W := \arg \min_W \left\{ ||W\Phi - T||^2 + R(W) \right\}
\]

where \( ||.|| \) denotes the Euclidean norm.

Relevance Vector Machine (RVM) is a sparse Bayesian approach to classification and regression.\textsuperscript{12-13} It can generate robust results even on samples with low-dimension features. It introduces Gaussian priors on model weights, and each prior being governed by its own individual scale hyperparameters. The hyperpriors can be done analytically, and we encode a preference for smoother functions by defining an ARD Gaussian prior over the weights\textsuperscript{14}

\[
p(W|\alpha) = \prod_{i=1}^{r} N(w_i | 0, \alpha_i^{-1})
\]

where \( \alpha \) is a vector of \( r \) hyperparameters. The minimization function is then given by

\[
||W\Phi - T||^2 + \sum_{i=1}^{r} \left[ \frac{1}{\alpha_i} \exp\left( -\frac{w_i^2}{2\alpha_i^2} \right) \right]
\]

This regularization penalty form of \( R(W) \) in latter part of Equation (8) pushes unnecessary weighting parameters to zero and thus produces a sparse model. In our experiments we use kernel bases in the form

\[
\Phi(x) = (K(x, x_1), ..., K(x, x_n))^T
\]

These kernel bases based on the kernel function \( K(x, x_j) \) instantiated at each training sample \((t_i, x_i)\), which offers the RVM relevant examples effectively. Our experiments use the Gaussian kernel \( K(x, x_j) = \exp(-\mu ||x_i-x_j||^2) \) with coefficient \( \mu \) estimated from the scatter matrix of our training data, and is proved to produce acceptable results.
**Combination Blow**

Besides the single attack style specified above, it is always noticed that the players carry on a series of successive attack styles rapidly in practice. This is called combination blow which is a favorite attack manner for enhancing the damage intensity. This attack manner should be regarded as combination of conventional routine training and private brilliant creativity which influence on blow strategy as how and when.

In order to reflect the routine, we incorporate the prior probability into directed transition graph construction. A combination blow example motion stream can be regarded as a sequence of attack styles stitched along the time axis. We represent it as a transition graph denoted by \( e = (A, G, E) \) where \( A \), \( G \), and \( E \) are an attack style sequence, groups which styles belonged to, and a set of edges, respectively.

As Figure 3 shows, \( e_i \) is a combination blow example, the nodes represent the attack styles, and the weighting edge represents the transition. The weight \( w(a_i,a_j) \) records the transition number from the style \( a_i \) to \( a_j \) appeared in all combination blow samples in database. Then the probability \( P(T_{ij}) \) of transition from style \( a_i \) to \( a_j \) can be modeled as

\[
P(T_{ij}) = \frac{w(a_i,a_j)}{\sum_{j=1}^{n} w(a_i,a_j)}
\]

where \( n \) is the number of all attack styles originated from \( a_i \) in database.

In the simulating stage, once an attack style \( a_i \) is successfully performed, estimation is made to judge whether successive style \( a_{i+1} \) is needed, and if so, which kind of style is mostly desired. This estimation is calculated by a scoring function

\[
S_{ij} = \text{Suc}(a_i) \cdot P(T_{ij}) \cdot e^T + S_{i-1,j} \quad \text{with} \quad S_{0,1} = 1
\]

Where \( \text{Suc}(a_i) \) is a Boolean function which is true if and only if \( a_i \) works effectively without being dodged, and the power \( \tau \) evaluates the player’s experimental level of handling combination blows. The attack style which retains the maximum score takes priority over the others and be the successive attack style. In practice, a threshold \( \theta_{\text{end}} \) is set to confine the rounds of combination blows in a reasonable scale.

**Response Model**

Compared with the previous work,\(^4\) we employ an improved hybrid approach to create response model. In order to build a transition-to motion which serves as both a reference for the response and a sponsor for successive animation, the dynamics is implemented to generate an informative character trajectory, which provides the next searching process with comparative basis. The comparison between the simulated sequences and the ones in database is conducted after the sequences are aligned both in time and root.

We represent each motion frame \( f_i \) in classical vector form \( (P_{0i}, \theta_{0i}, \ldots, P_{n_i-1}, \theta_{n_i-1})^T \), where the \( P_i \) and \( \theta_i \) are the position and orientation of joint \( I \), respectively. Within the comparison of frame windows \( S_1 = \sum_{i=1}^{c} f_{1i} \) and \( S_2 = \sum_{i=1}^{c} f_{2i} \) with size \( c \),\(^15\) the distance between them is defined in a reformative form as

\[
D(S_1, S_2) = \sum_{i=1}^{c} w_i \left( \sum_{j=1}^{\lambda} w_{Rj} \left( \sum_{l=1}^{\gamma} w_{Pj} |P_j(f_{1i}) - P_j(f_{2i})| + w_{Rj} |\theta_j(f_{1i}) - \theta_j(f_{2i})| \right) \right)
\]

where window weight \( w_i \) takes the highest value for the initial frames and decreases for the subsequent frames. The weights \( w_{Pj} \) and \( w_{Rj} \) scale the linear and angular distances for each sensitive region of human body such as limbs. According to the observation in experiment, the distance between two windows exhibits the sensitivity to easy-stretched and easy-shift body regions as they provide maximum diversities between two frames. Here \( \lambda \) and \( \gamma \) are the number of sensitive regions and inclusive joints, and \( w_{Rj} \) is the region weight tuned by training observation and, has higher value for the upper limbs in boxing simulation.

**Result and Discussion**

**Experiment Platform**

We perform our experiments on a PC (Intel Pentium Dual-Core 2.4 GHz; 3.5 Gbytes of RAM; Nvidia Quadro
FX 4500 graphics card). Our character model has 42 DOFs, consisting of six DOFs for the pelvis, six DOFs for each limb, three for the neck, three for the head, and six for the spine. We capture 10 types of boxing motions and a stream of combination blows motion. Each of the former contains actions performed by left-handed and right-handed players and is 15 minutes long with 90 Hz sampling rate, and the latter contains variable length sequences with the total time of 10 minutes long at 90 Hz. The segments derived from latter combination stream are obtained using the segmentation technique, as the example stream is segmented at frames where the contact force exhibits local minima. Each segment motion contains some basic semantic and is employed in constructing transition graph in interactive models. The format of output animation sequence is the combination of motion sections, and these sections obtained from captured data represent action element with minimal granularity in boxing sport.

**Experiment Results**

Figure 4 shows three representative actions of the player with gold color in state balance, defend, and attack, respectively. In Figure 5, the simulated results of attack generation are exhibited particularly, and three successive key-frame sequences of different attack styles are listed in the figure. Figure 5(a) shows the hook style performed by the player with cyan color upon the component in short range entangled. Figure 5(b) shows the long-range attack style performed by the player with gold color, which is swing blow with the character of curved trajectory. Figure 5(c) shows the straight punch which is suitable for both long and middle ranges performed by the player with cyan color.

**Quantitative Evaluation**

In the phase of quantitative evaluation, two cross-validation tests are conducted for justifying our design decision in statistical estimation. There are two stages in our experiment. In the first stage, the simulated boxing animations are compared with the real ones in statistical performance. As illustrated in previous works, statistical results can be used as one valuable factor in evaluating the simulated reality. In the second stage, as the complement to the former tactic, we incorporate human subjective opinions into our reality judgment by the assistant of manual annotation. User can annotate the unrealistic parts of the simulated animation in time axis, and the visual reality percentage is evaluated by

\[
Rp_i = \frac{F_{\text{all}} - F_{\text{annotation}}}{F_{\text{all}}} \quad (15)
\]

where \(Rp_i\) is the reality percentage judged by user \(i\), and \(F_{\text{all}}\) is the number of the simulated frames while \(F_{\text{annotation}}\) is the one of annotation frames. We rely on such two-stage judgment based on statistical and visual criterion, which is proved to be reliable and reasonable.

To make our experiment more convincing, hierarchical manual interactions are applied both in training and testing by considering users variant comprehensions of boxing sport. Three levels are employed here for separating up grading users as the junior set which contains the users with little knowledge about boxing, the moderate set which contains the users with proper acquaintance, and the senior set which contains the users with professional comprehension. A hundred people act as volunteers in our experiment with 50 men and 50 women. Among these people, about 20 per cent of them fall into senior group, 30 per cent in moderate, and the rest in junior. Meanwhile, about 14 per cent of them are left-handed persons. They are partitioned into five

![Figure 4. State simulation result. (a)Balance. (b)Defend. (c)Attack.](image-url)
equal subsets following the above proportions for fivefold cross validation.

**Performance on Statistical Comparison.** Table 1 shows the comparison between the data achieved from real boxing video and our simulated boxing animation. The columns correspond to three states and sub-classified actions corresponding to these states. The data in all but the last two rows record the percentage of time taken by corresponding action when compared to total time. The symbols ‘#’ and ‘/’ are used for identifying simulated animation and real video, respectively. Rows #1, #2, and #3 are three simulated samples with the total time of 15, 25, and 35 minutes long at 24 fps, and rows /1, /2, and /3 are three real video samples with same times and fps. Based on the average data of the two types, the row “A-ER” corresponds to action error rate while row “S-ER” corresponds to state error rate, the only difference between them is the granularity.

**Performance on Manual Annotation.** Figure 6(a) reveals the acceptances of the simulated animations annotated by different groups as the section number increased. The interactive models are trained by users randomly selected from three groups. These acceptances are measured according to reality percentage of simulated animation sequence. The red curve records the average reality percentage Rp judged by the users belonging to senior group, which spaces out with the moderate and junior ones above. It is easy to find that the moderate one with green color is close to the junior one with blue color. All these ones keep stable in spite of fluctuating in local ranges as the section numbers increased.

Figure 6(b) evaluates the influences of training samples with increased sizes on user acceptances. For this purpose, we want to figure out how the professional knowledge will affect our learning mechanism with measurement of reality percentage. In this experiment, given that the simulated results are judged by senior
level group, we apply the samples which are trained by three groups to interactive models. The junior curve exhibits unsteadiness as sample size enlarged, and in contrast, the other two with proper knowledge guidance grow gently and enter a stable zone after the size reach to a sufficient degree.

Figure 6 shows the comparison of reality percentages between some potential approaches labeled in the figure and RVM algorithm applied for attack selection as the training samples increased. The random generated approach is used as benchmark. With the pre-conditions of restricted section number of 40, moderate level training samples, and senior group judgment, the acceptances of RVM and SVM both surpass that of Markov chaining dramatically. The RVM and SVM curves are close to each other and share stable periods while the RVM takes advantage moderately in the figure.

**Table 1. Comparison between real boxing data and simulated animation data**

<table>
<thead>
<tr>
<th></th>
<th>Attack</th>
<th>Defend</th>
<th>Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single(%)</td>
<td>Combination(%)</td>
<td>Jump(%)</td>
</tr>
<tr>
<td>#1</td>
<td>9.57</td>
<td>11.24</td>
<td>1.99</td>
</tr>
<tr>
<td>#2</td>
<td>10.23</td>
<td>13.26</td>
<td>0.89</td>
</tr>
<tr>
<td>#3</td>
<td>10.15</td>
<td>12.87</td>
<td>1.23</td>
</tr>
<tr>
<td>#Avg.</td>
<td>9.98</td>
<td>12.46</td>
<td>1.37</td>
</tr>
<tr>
<td>'1</td>
<td>11.15</td>
<td>13.26</td>
<td>0.78</td>
</tr>
<tr>
<td>'2</td>
<td>13.26</td>
<td>13.12</td>
<td>1.85</td>
</tr>
<tr>
<td>'3</td>
<td>10.75</td>
<td>13.21</td>
<td>0.86</td>
</tr>
<tr>
<td>'Avg.</td>
<td>11.72</td>
<td>13.20</td>
<td>1.16</td>
</tr>
<tr>
<td>A-ER</td>
<td>14.85</td>
<td>5.60</td>
<td>18.10</td>
</tr>
<tr>
<td>S-ER</td>
<td>9.95</td>
<td>6.13</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. (a) The acceptances of different level users. (b) The influences of different training samples. (c) The comparison between RVM and other algorithms.

**Discussion**

The results in Table 1 indicate that our simulated boxing animations are close to the real ones in statistical layer. Considering two abnormal data of jump and dodge in row A-ER, both rates seem unexpected higher than others. Because it is hard to simulate human nervous mechanism in mathematical model precisely which makes great importance to both dodge and jump, and some factors like fatigue and damage degree in real scene will add the complexity additionally. All these
reasons make our physical-based rules appear unstable in simulating such two actions. But considering the low share of total simulation as they account for, these flaws do not impact on our simulated result seriously. Moreover, if we check the row S-ER which extends the restriction to state granularity, all the error rates are confined to lower than 10 per cent which is tolerable definitely in practice.

In Figure 6(a), as the number of output sections increases, the reality percentages of three groups remain stable in their own separated range, and it indicates that the unreality caused by our approach couples with sections increase in linear relationship. The acceptance of senior curve seems lower than the other two with moderate gap, as the professional knowledge may help to figure out some flaws omitted by junior and even moderate groups. In this paper, we are inclined to the professional users’ opinions as they are more accurate and convincible. As illustrated in Figure 6(b), the samples provided by junior users who are lack of professional guidance exhibit unstable characteristic, and the others guided by proper professional experiences achieve similar results. We also find an interesting aspect: our approach behaves insensitive to training samples guided by moderate or higher level users. Meanwhile, it maximizes the ‘margin’ between moderate and junior ones extremely. Considering the lack of professional guidance, this finding provides us with an eclectic way in practice. Generally, training samples guided by moderate group are used for convenience with no sharp sacrifice of simulated reality.

In the last experiment, RVM and SVM generate close results and reach to the stable state synchronously. But the testing time of SVM is about seven times longer than that of RVM in our experiment, and the trained SVM uses 47 relevance vectors and error rate is 10.15%, while the corresponding ones of RVM are 6 and with a improved rate of 9.06% which owes to higher sparsity of RVM. Compared to SVM, RVM brings dramatic reduction in complexity with moderate performance improvement.

Conclusions

In this paper, we present a novel layered and hybrid approach to synthesize two-player competitive motions in automatic manner. The synthesis of such motions is broken into twofold issues as state control and action selection. As the key issue detailed here, a hybrid approach which integrates empirical rules and interactive models is adopted to guide the appropriate action selection. The experiment results show that our approach achieves stable and well performance on simulating dense and asynchronous interactions. In the future, we would like to develop the proposed approach for constructing full limb interactive models. With such generalized models, we could simulate more competitive and free sports such as wushu and karate, each of which contains a variety of asynchronous dynamics motions.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant No.60525108, the National Key Technology R&D Program of China under Grant No.2007BAH11B03, Program for Changjiang Scholars and Innovative Research Team in University (IRT0652), the National Science Foundation for Post-doctoral Scientists of China under Grant No. 20080431327.

References


Authors’ biographies:

Zhang Liang got his bachelor’s degree from the Department of Computer Science at Zhejiang University (2007) majoring in software engineering. Now he is a Ph.D. student of the College of Computer Science at ZJU, working in Microsoft Visual Perception Lab of ZJU. His research fields include computer animation and computer vision.

Jun Xiao got his bachelor’s degree from the Department of Computer Science at Zhejiang University (2002) majoring in computer science and engineering. Now he is an Assistant Professor of the College of Computer Science at ZJU, working in Microsoft Visual Perception Lab of ZJU. His research fields include computer animation, digital entertainment, and multimedia retrieval.

Yueting Zhuang received his Ph.D. degree from the Department of Computer Science at Zhejiang University (1998). From 1997 to 1998, he worked as a visitor at the Department of Computer Science and Beckman Institute of UIUC. He now is the professor and dean of the College of Computer Science at ZJU. His current research interests are computer animation, multimedia analysis, digital entertainment, and digital library technology.

Cheng Chen got his bachelor’s degree from the Department of Computer Science at Zhejiang University (2004) majoring in computer science and engineering. Now he is a Ph.D. student of the College of Computer Science at ZJU, working in Microsoft Visual Perception Lab of ZJU. His research fields include 3D human motion acquisition/ animation and computer vision.