Vol.2009-CVIM-167 No.2
2009/6/9

Regular Paper

Keypose and Style Analysis Based on Low-dimensional Representation

Manoj Perera,[†1] Shunsuke Kudoh[†1]
and Katsushi Ikeuchi[†1]

Human motion analysis is a complex but extremely interesting and important research area in computer graphics and computer vision. This paper addresses three vital topics in human motion analysis related to keyposes: how to extract keyposes from dance motions, how to utilize them, and how to recognize the person and task that constitute a keypose. As our first topic, we propose a new method to extract keyposes from a given dance using the energy flow of the motion. Our experimental results and comparison with a previous keypose extraction approach show the high accuracy of keypose extraction with our new method. As our second topic, we propose a new method to reconstruct low-dimensional motion based on keyposes, and we illustrate the effect of keyposes in a given motion space on human perception. We utilize the keyposes extracted with our new method, formulate a model, and derive a low-dimensional motion based on our model. We also construct low-dimensional motion using uniform sampling poses, and we compare the results with those obtained from our method. As our third topic, we propose a novel approach to decompose motion into common and individual factors using the Multi Factor Tensor (MFT) model. By this method, we recognize person and task from the motion sequence.

1. Introduction

Human motion analysis is a complex but extremely interesting and important research area in computer graphics and computer vision. Researchers have proposed various methods of doing this. We emphasize keyposes related to human dance motions for our analysis purposes. Keyposes play an important role in understanding the structure of dance motions, recognizing the dance, and further analysis purposes such as style analysis. Therefore, there is increasing demand to investigate the role of keyposes in human motion analysis.

†1 The University of Tokyo

This paper addresses three vital topics in human motion analysis related to keyposes. Within our first topic, extraction, we describe an energy function to extract keyposes from human motions. In the second topic, utilization, we explain how to utilize the keyposes and present a method to create low-dimensional motion that has high impact on human perception. In the third topic, decomposition, we introduce a method to recognize the components of keyposes using Multi Factor Tensor (MFT) Analysis.

The remainder of this paper is organized as follows. Section 2 presents related work. Section 3 explains keypose extraction of dance sequences. Section 4 describes low-dimensional motion reconstruction based on keyposes and its impact on human perception. Section 5 describes the Multi Factor Tensor model to decompose human motion into common and individual factors. Finally, in Section 6 we present conclusions and describe future work.

2. Related Work

Recently motion capture data have been frequently used for character animation, reproducing human motions ([2],[14],[27],[21]) and vision applications ([7],[8]). In most cases, segmentation is observed as a basic technique to achieve goals and speed up the process. Not only for captured motion data, but also for motion data from video sequences, researchers have applied different techniques for segmentation or extracting keyposes/keyframes and have obtained various results.

Zelnik-Manor et al. introduced a method[44] based on a distance measure developed over a variety of temporal scales. Liu et al.[22] used a clustering-based adaptive keyframe extraction algorithm for improving 3D motion retrieval speed. Loy et al.[24] also applied clustering and selected the central frame of clusters as the keyframe. They successfully extracted keyframes in a sports event video sequence. Park et al.[25], Kovar et al.[13], extracted keyposes mainly for motion synthesis and retrieval. Similar work was conducted by Vermaak et al.[41], Fauvet et al.[6] for keyframe selection in video sequences focusing on background scene and camera motion. In our study we additionally consider music; our focus is on a different aspect and the above work does not comply with the objective of our work.

Assa et al. proposed a method to select keyposes based on embedding the

motion curve in low-dimensional space and applying a simple geometric algorithm to identify the important poses ([3]). They discussed different applications to which the proposed method can be applied. However, their results indicate that the presented method is more applicable to short motions such as tasks in our study, and in a dance motions there are several tasks connected together. They also iteratively select the keyposes locally until a satisfactory number of keyposes has been extracted related to a threshold, whereas in a dance motion the number of keyposes is fixed, although we don't have any prior knowledge of the number of keyposes. Yasuda et al. improved over the above method ([43]). For long motions that include several types of actions, to avoid loosing information for each action, they suggested segmenting the motion data ([4]). In[4], the researchers segmented motion data into high-level behaviors (e.g., walking, running), and also due to the time criterion they searched for the minimum duration of simple motions, which is not applicable to the scope of our study.

Many researchers have applied different approaches to regenerate, reproduce, or synthesize human motions. A Fourier-based method was utilized in[36] to generate human motion with behavioral characteristics. In[1], Amaya et al. presented a model to produce emotional motions based on signal processing. Brand et al. used *Hidden Markov Models* with entropy minimization in[5] to detect the style variations in sample data and applied the results to novel dance sequences. A two-mode PCA framework was described in[35] to linearly classify male and female walkers. Gao et al. introduced a three-mode expressive feature model in[8] to embed tunable weights on trajectories within the sub-space model to enable different style estimations in their experiments. In[19] a method is discussed for inferring a 3D body pose from silhouettes based on learned activity manifolds. Safonova et al. described a method in[28] that solved an optimization problem in low-dimensional space and synthesized physically realistic human motion. In[37] motion generation techniques were presented and the generated motions were parameterized according to speed or length. Vasilescu used tensor algebra in[38] to recognize the human motion signatures and applied it to motion synthesizing. In the second topic of our study we apply a different approach to obtain low-dimensional motions based on principal components incorporated with keyposes or uniform sampling poses.

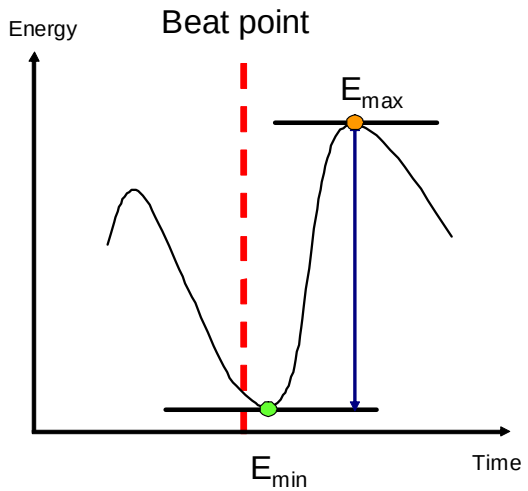


Fig. 1 Candidate Determination: The broken line shows the estimated music beat. The curved line describes the computed energy flow in the global coordinate system. The dark circles indicate the local minimum and local maximum energy points.

Researchers have put a considerable amount of effort into investigating style. Tenenbaum et al. introduced a bilinear model[31] for two-factor problems in separating style and content factors. They displayed two-mode examples in extrapolating fonts for unseen letters and translating face postures to new illuminations[32]. Given several sequences of walking silhouettes of different people, Su Lee et al. decomposed intrinsic body configuration through action (content) from the appearance (or shape) of the person performing the action (style)[18] and used the results for recognition. Style and content were separated on a nonlinear manifold and applied in interpolating the modes of gait styles and manner of smiling[20]. Gao et al.[7),8)] proposed a three-model principal component model and recognized the expressiveness in the style of human action when carrying different weights and when varying the walking pace of different people. Kannapan et al.[11),12)] also recognized the effort in human actions within a three-mode principal component analysis (PCA) framework.

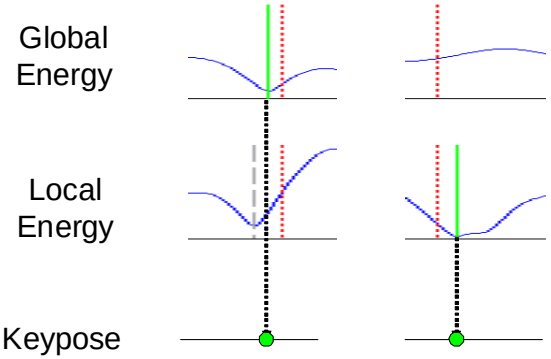


Fig. 2 Keypose selection: Here, the top and middle rows show two instances of global and local energy graphs. The dotted lines indicate the estimated music beats. The Gray-colored broken lines indicate the selected keypose candidates. The dotted arrows indicate the selected keyposes.

3. Keypose Extraction with Energy Analysis

3.1 Keypose Extraction Method

In 2004 Shiratori et al. used a method to extract keyposes ([29),30)]). The speed of end effectors and the musical beat were used in this method, but some issues need to be improved. We introduce a novel energy function for keypose extraction combined with music rhythm tracking as an addition to the above method for keypose extraction.

We formulate our energy function for keypose extraction by detecting correspondence between each pose or frame with global and local coordinate systems. The global coordinate system is fixed on the floor, while the local coordinate system is attached to the waist of the human body. The local coordinate system, which is also the *body center coordinate system* has its origin at the waist position of the body. Here, the Z axis is the direction from waist to body, the Y axis is the frontal direction, and the X axis is perpendicular to these axes. Let the human body consist of $i = 1, \ldots, N$ marker positions, where m_i represents the weight

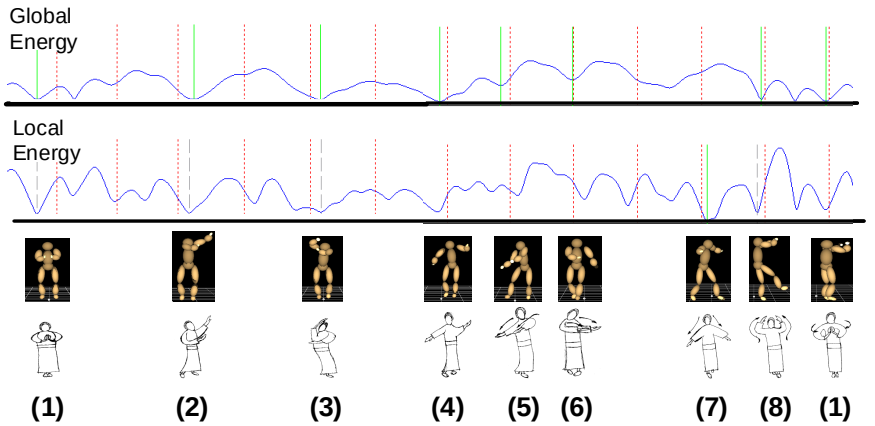


Fig. 3 *Aizu-bandaisan* dance keyposes 1: Top row and second row show the energy flow graph in global and local coordinate systems. The dark broken lines display the estimated musical beat. The gray-colored dashed lines represent the candidates, and the straight lines represent the extracted keyposes. (1), (2), (3) etc. represent the keypose number. The third row shows the extracted keyposes by our method in a viewer and the fourth row shows the keyposes drawn by dancing professionals.

for each marker position. Then $\boldsymbol{F}_t^G$, the global energy at time t is defined as

$$\boldsymbol{F}_t^G = ||\sum_{i=1}^{N} m_i \boldsymbol{V}_i(t)||, \tag{1}$$

where $\boldsymbol{V}_i(t)$ describes the velocity of the i th marker position at time t in the global coordinate system. The local energy $\boldsymbol{F}_t^L$ at time t is defined as

$$\boldsymbol{F}_t^L = ||\sum_{i=1}^{N} m_i \boldsymbol{v}_i(t)||, \tag{2}$$

where $\boldsymbol{v}_i(t)$ represents the velocity of the i th marker position at time t in the local coordinate system.

After computing the energy flow of the motion given, we determine the suitable candidates for keypose extraction. The keypose candidate determination process requires several thresholds as described below, and all the threshold values were obtained from example-based learning.

Candidate determination in global and local coordinate systems is conducted

separately, and later the determined candidate results are combined during keypose extraction. For each energy function, keypose candidates are calculated as follows.

Figure 1 displays the energy flow in the global coordinate system. The broken line represents the estimated music beat. First we find the local minimum around the music beat. We also find the local maximum around the music beat and find the energy difference between $\boldsymbol{E}_{max}$ and $\boldsymbol{E}_{min}$. If the local minimum is small enough, we select that as a candidate. If the energy difference is greater than the amount $\boldsymbol{T}_2$, we select the pose as a candidate. In this operation, we use two thresholds, $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$, which are obtained from example-based learning. In the same manner, the candidates for the keyposes are determined for the energy flow in the local coordinate system.

Figure 2 describes the keypose selection process from the determined keypose candidates in global and local coordinate systems. In accordance with a particular estimated rhythm beat, we search for the existence of a keypose candidate in the global coordinate system. If there is a global candidate, that global candidate is selected as a keypose. If there is no global candidate, but if there is a local candidate, the local candidate is selected as a keypose. This process is repeated until all the estimated rhythm beats are searched for the existence of keyposes corresponding to them. In a local energy flow graph, the effect of rotation is neutralized, and in practical situations, where the dancer is moving around a particular space such as a room, the effects of rotation and non-rotation are interconnected in a complex way. Our experimental results and the energy flow graphs indicate that the global energy candidates bear higher priority than local energy candidates.

3.2 Keypose Extraction Results

We conducted several experiments to evaluate the keypose extraction method. The extracted keyposes were compared with the dance masters' teachings. We extracted the keyposes of five Japanese folk dances, namely *Aizu-bandaisan*, *Jongarabushi*, *Donpan*, *Kokiriko-sasara* and *Kokiriko-theodori* dances. Figure 3 illustrates the keypose extraction results for the *Aizu-bandaisan* dance. According to dance masters' teachings the *Aizu-bandaisan* dance comprises eight keyposes. Our results show that the new method is capable of extracting all the keyposes.

Table 1 Summary Keypose Extraction Results: Summary of the keypose extraction results are displayed.

Dance	True	Our Method
Aizu-bandaisan	8	8
Jongara	12	12
Theodori	10	10
Sasara	10	10
Donpan	33	31

We note that using our previous approach for the *Aizu-bandaisan* dance, we were able to extract only four of the keyposes precisely, shown in Figure 3 as (1), (2), (3), and (4). In figure 3 the top two rows display the graphs of energy function in global and local coordinate systems respectively. The dotted lines and gray-colored broken lines in those rows display the estimated musical beat and the selected candidates. The dark straight lines represent the extracted keyposes. In Figure 3, (1) on the left side and (1) on the right side represent the same keypose, while they belong to two different but consecutive dancing cycles. In the third row, the pictures with a character show the keyposes seen through a viewer. The other pictures represent the keyposes drawn by the dancing professionals.

Table 1 summarizes the results of our keypose extraction experiments. The True column represents the number of total keyposes each dance has according to dance masters' textbooks. Our Method column indicates the number of keyposes extracted by the new method. Example-based learning works well, as shown in this table.

3.3 Effect of Threshold Values

This subsection explains what will happen to keypose extraction by varying thresholds. For this purpose we use Precision and Recall. Precision indicates the number of correct keyposes in the extracted poses divided by the total number

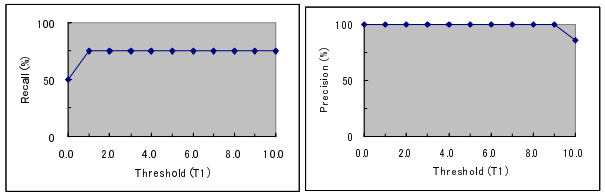


Fig. 4 Analysis of effect of $\boldsymbol{T}_1$ on keypose extraction: The graphs display the recall and precision results of threshold $\boldsymbol{T}_1$ on keypose extraction

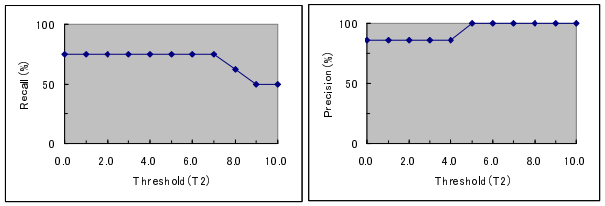


Fig. 5 Analysis of effect of $\boldsymbol{T}_2$ on keypose extraction: The graphs display the recall and precision results of threshold $\boldsymbol{T}_2$ on keypose extraction

of extracted poses. Recall indicates the number of correct keyposes in extracted poses divided by the total number of correct keyposes according to the textbook.

Note that in this analysis we consider only global energy for simplicity. Global energy and local energy have roughly the same tendency regarding keypose extraction with respect to variance in threshold values. Between global and local energy, global energy has the dominant influence on the keypose extraction result, so we focus on global energy only for this analysis.

Figure 4 shows recall and precision when varying threshold $\boldsymbol{T}_1$ from 0 to 10. For a large range of values, precision and recall are constant. This shows that the selection of threshold value is not that sensitive for keypose extraction. Figure 5 shows recall and precision when varying threshold $\boldsymbol{T}_2$ from 0 to 10. The graphs show a higher percentage for a large range of values. We can see a higher percentage for a wide range of values, even for one energy graph. Even the worst case provides 50%. This shows that threshold is not very sensitive. The reason is we use the music beat, and we search keyposes around the music beat.

3.4 Weights in Energy Function

In our experiments we analyzed the effect of different weight settings on keypose extraction. We set the weights in our experiments as uniformly distributed. Another possibility of weight setting is based on mass distribution of the human body ([9),15)]).

The graphs in figure 6 show an example of energy computation according to different ways of weight assignments. At the top, the two graphs show global and local energy computation based on weights assigned according to mass distribution. At the bottom, they show weight computation based on uniformly distributed weights. Roughly, it appears that graphs based on both methods are the same. But if we analyze carefully, we can see that the energy flow based on weights according to mass distribution is more noisy and error-prone. Our observation on weight distribution provides little effect on energy function. But since uniform distribution of weights provides more weights to end effectors, it provides relatively clear characteristics.

4. Low-dimensional Motion Reconstruction

4.1 Motion Model

In this subsection, we describe the process of decomposing high-dimensional motion into low-dimensional representation. We incorporate a simple dimensionality reduction technique, such as Principal Component Analysis (PCA) for the purpose. For a particular motion space and for a particular set of poses, we create an eigen model relevant to the specific space with the use of the PCA technique. The eigen model, which consists of the eigen vectors and the average vector created by PCA analysis incorporated with the specific poses, is then used to generate the in-between motion in low-dimensional space, connecting the above poses. We define the eigen model incorporated with keyposes as *keypose space*. We demonstrate that keypose-based low-dimensional motion generation is significantly better than the low-dimensional motion generated based on eigen model incorporated with uniform time interval spaced poses. The details of the low-dimensional motion generation are explained below.

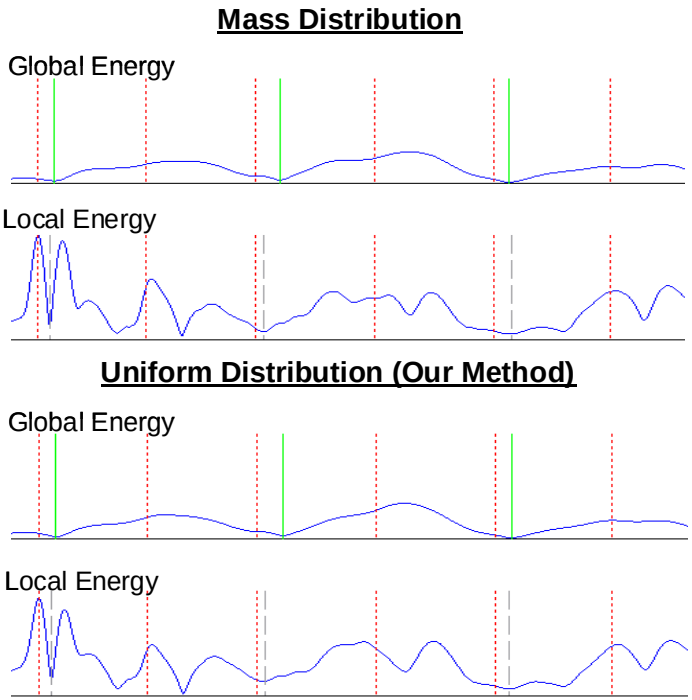


Fig. 6 Weight assignment in energy function: The two graphs on top display the energy flow when the weights are set according to mass distribution of the human body. The two graphs at the bottom show the energy flow when the weights are uniformly distributed.

4.1.1 Approach

Let a human pose, at any given time t be denoted as a line vector,

$$\varpi = [\vartheta_x^1, \vartheta_y^1, \vartheta_z^1, \ldots, \vartheta_x^i, \vartheta_y^i, \vartheta_z^i, \ldots, \vartheta_z^N], 1 \leq i \leq N \quad (3)$$

where ϑ_x^i represents normalized x position in the body center coordinate system and i represents the i(th) marker position in the human body. Similarly, ϑ_y^1 and ϑ_z^1 represent the normalized y and z positions in the body center coordinate system respectively. Each pose vector has a dimension of $(3 * N)$ where N is the number of marker positions in the body. We formulate our model by forming the covariance matrix $\boldsymbol{A}$, with the relevant keyposes for one dance within one cycle, for several people. We assume that there are $j = 1, \ldots, \mu$ number of people and κ number of keyposes in one dance cycle. Given K keyposes where $K = \mu * \kappa$, we compute the mean pose,

$$\varpi_0^{Kp} = \frac{1}{K} \sum_{k=1}^{K} \varpi_k^{Kp}. \quad (4)$$

In ϖ_0^{Kp} and ϖ_k^{Kp}, Kp indicates that the poses are correlated to keyposes. We compute the eigen vectors of the covariance matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{Q}^T$$

where $\boldsymbol{Q}$ is a $K * 3 * N$ mean subtracted pose matrix and $\boldsymbol{A}$ is a $K * K$ matrix. The kth line of $\boldsymbol{Q}$ can be described as

$$\boldsymbol{Q} = (\varpi_k^{Kp} - \varpi_0^{Kp}), 1 \leq k \leq K.$$

We compute the eigen vectors $\Upsilon_{1 \leq k \leq K}^{Kp}$ by Singular Value Decomposition (SVD). We build a linear mapping to obtain *eigen keyposes* as follows:

$$\Psi_{K'}^{Kp} = \sum_{k=1}^{K'} \Upsilon_k^{Kp\,T} (\varpi_k^{Kp} - \varpi_0^{Kp}) \quad (5)$$

where K', $1 \leq K' \leq K$. Consequently, we can represent any keypose in low-dimensional space as

$$\widehat{\varpi}_{K'}^{Kp} \approx \varpi_0^{Kp} + \sum_{k=1}^{K'} \Upsilon_k^{Kp} \Psi_{K'}^{Kp}. \quad (6)$$

Utilizing the above computed eigen prototype for the specific space, we generate the low-dimensional motions for various people included in our dataset. Let a pose that belongs to the original normalized motion of any person of the training set, which eventually lies in between any of that person's keyposes, be denoted as ${}_g^j\Gamma_k^{IN}$. In ${}_g^j\Gamma_k^{IN}$, Where k represents the dimension of the pose and IN represents that the pose lies between any of the keyposes of the person considered j represents that the pose belongs to the j'th person in the dataset, and g represents that the pose is the g'th pose in the cycle of the particular person. Let the dancing cycle of the relevant person contain ${}^jK$ poses in the cycle. Then, an eigen pose, which interconnects the eigen keyposes, is denoted as

$${}_g^j\Psi_{K'}^{IN} = \sum_{k=1}^{K'} \Upsilon_k^{Kp\,T} ({}_g^j\Gamma_k^{IN} - \varpi_0^{Kp}) \quad (7)$$

where $1 \leq g \leq {}^jK$. We define any pose that interconnects the keyposes in

low-dimensional representation as

$$ {}^{j}_{g}\widehat{\varpi}^{IN}_{K'} \approx \varpi^{Kp}_{0} + \sum_{k=1}^{K'} \Upsilon^{Kp}_{k}\, {}^{j}_{g}\Psi^{IN}_{K'}. \tag{8} $$

K' controls the percentage of the dataset in low-dimensional pose space that can be represented as

$$ \rho = \frac{\sum_{k=1}^{K'} \lambda_k}{\sum_{k=1}^{K} \lambda_k} \tag{9} $$

where λ_k represents the k'th most significant eigen value, which is obtained by SVD. The above low-dimensional motion from high-dimensional original motion is obtained by utilizing the keypose-based eigen prototype.

In the same manner we create the low-dimensional motion or the sequence of low-dimensional poses relevant to a particular person, based on an eigen model constructed with a certain number of poses that equals the same number of keyposes used in the keypose-based model. These poses are extracted by uniform sampling. From each dancing cycle of each person in the dataset, the same number of poses that equals the number of keyposes in one dancing cycle are extracted, making the time interval between each extracted pose within one cycle of each person to be the same.

We define an eigen pose, which interconnects the eigen uniform sampling poses, as

$$ {}^{j}_{g}\widetilde{\Psi}^{IN}_{K'} = \sum_{k=1}^{K'} {\Upsilon^{Un}_{k}}^{T} ({}^{j}_{g}\Gamma^{IN}_{k} - \varpi^{Un}_{0}) \tag{10} $$

where Υ^{Un}_{k} represents the eigen vectors obtained by applying SVD to uniform sampling poses, and ϖ^{Un}_{0} denotes the mean pose computed from uniform sampling poses. The low-dimensional motion generated by uniform sampling pose-based eigen prototype can be denoted as

$$ {}^{j}_{g}\widetilde{\varpi}^{IN}_{K'} \approx \varpi^{Un}_{0} + \sum_{k=1}^{K'} \Upsilon^{Un}_{k}\, {}^{j}_{g}\widetilde{\Psi}^{IN}_{K'}. \tag{11} $$

4.2 Results

We used motion capture data of the *Aizu-bandaisan* dance of eight people to conduct our experiments. Our dataset contains the motion data of several dancing cycles of five female and three male dancers. We randomly selected one dancing cycle from each person, extracted the keyposes as described in 3, and

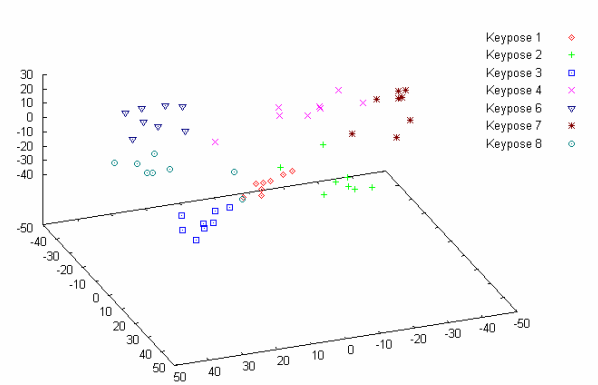


Fig. 7 Eigen keyposes: The distribution of keyposes used in the keypose-based method plotted onto eigen space is displayed. These eigen keyposes are constructed using the first three most significant eigen values incorporated with the keypose model. The keyposes are shown in different colors and shapes.

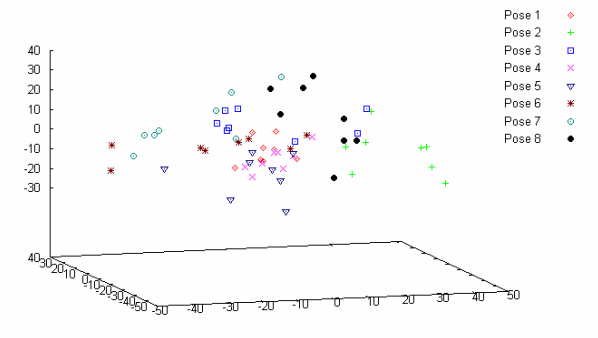


Fig. 8 Uniform sampling poses: The distribution of uniform sampled poses used in the uniform sampling-based method plotted onto eigen space is displayed. Like the eigen keyposes, the eigen uniform sampled poses are also constructed using the first three most significant eigen values incorporated with the uniform sampled pose method. Uniform sampling poses are displayed in different colors and shapes.

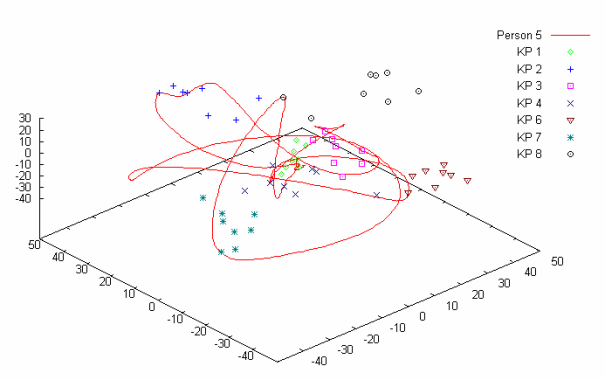


Fig. 9 Eigen motion generated based on keyposes: This shows the eigen motion obtained with the keypose eigen prototype for person five in our dataset. The red curve displays the eigen motion for one dance cycle of person five. KP indicates the keyposes, and the different shapes denote the eigen keyposes of the model.

constructed the keypose-based eigen model to evaluate our model. Similarly the uniform sampling pose-based eigen model was constructed by extracting poses, using the same randomly selected dancing cycles.

Figure 7 and figure 8 illustrate the intermediate stage of the low-dimensional motion generation process using the keypose-based eigen model and uniform sampling pose-based eigen model. Figure 7 displays the eigen keyposes, constructed with the first three most significant eigen values incorporated with the keypose-based model, where $K' = 3$, according to Equation 7. Figure 8 shows the uniform sampled eigen poses, constructed in a similar manner to eigen keyposes. These two figures illustrate that the eigen keyposes are clustered nicely, where uniform sampled eigen poses are not effectively clustered.

Figure 9 illustrates another intermediate stage of the low-dimensional motion representation process with a keypose-based framework. It describes the eigen motion of one dance cycle of person five, which belongs to the keypose space. The eigen motion ${}^{j}_{g}\Psi^{IN}_{K'}$, is obtained according to Equation 7 where $1 \leq g \leq {}^{j}K$.

Figure 10 displays the low-dimensional motions created by the keypose-based framework. The characters display an instance of one-, two-, three-, and four-dimensional motions of person five beginning from the left side. Our results

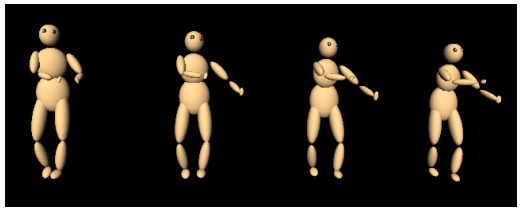
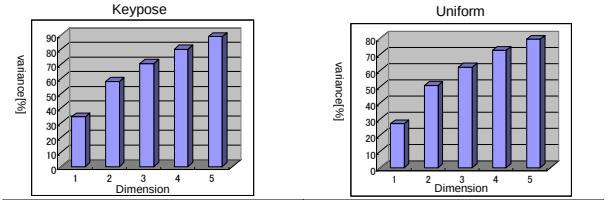


Fig. 10 Low-dimensional motion representation of person five: An extraction of the low-dimensional motion representation of person five is displayed. The characters show one-dimensional, two-dimensional, three-dimensional, and four-dimensional motion starting from the left side.



Dimension	1	2	3	4	5
Total Variance	34	58.2	70.3	80.1	88.8

Dimension	1	2	3	4	5
Total Variance	27.1	50.9	62.2	72.4	79.3

Fig. 11 Variance distribution of approach 1: The graphs display the distribution of total variance covered by the first few eigen postures in both keypose-based and uniform sampling pose-based methods. The left side graph displays the variance distribution of eigen key postures, and the right side graph shows the variance distribution of eigen uniform sampling poses. Dimension indicates the number of principal components associated with the variance.

also demonstrate that three-dimensional motion is quite impressive and efficient for further motion analysis processes and other applications. Therefore we used three-dimensional motions for low-dimensional motion comparison and analysis purposes in all our experiments.

4.3 Results

We evaluated the results of low-dimensional motion representation by keypose-based eigen model and uniform sampling pose-based eigen model. Figure 11 illustrates the total variance covered by the first few eigen vectors or principal components. The graphs show that the keypose-based principal components cover a greater percentage of total variance than the uniform sampling

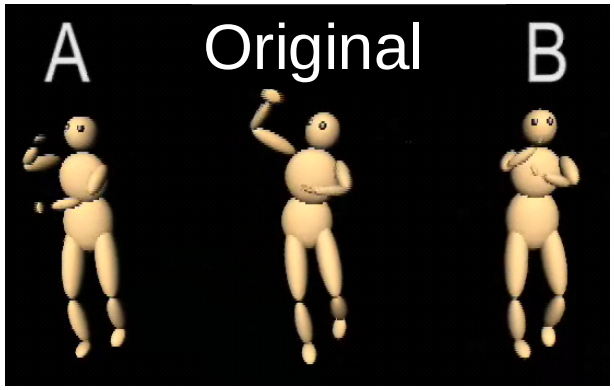


Fig. 12 A still capture of the user study: A still capture of the video that we used for the user study is displayed. In this sequence, the character marked as A shows the low-dimensional motion created by keyposes, and the character marked as B shows the low-dimensional motion created by uniform sampling poses. The middle character displays the original motion of the same person.

Table 2 Results summary of the user study: Summary of keypose-based and uniform sampling pose-based low-dimensional motion representation comparison with original motion.

	Ratio	Percentage
Sequence 1	32/34	94%
Sequence 2	29/34	85%
Sequence 3	31/34	91%
Total	92/102	90%

pose method. They also indicate that when we generate low-dimensional poses from high-dimensional pose space, the low-dimensional poses generated from the keypose-based eigen prototype contain a larger amount of total information.

Further, to investigate the amount of impact that the low-dimensional motions generated with these two methods has on human perception, we conducted a user study. Several low-dimensional motion cycles of different people were shown to a group of adults selected randomly, who participated in an open campus event. Participants in the study were asked to compare both low-dimensional motions, and select the motion that matched best to the original dancing motion of the same person.

Figure 12 displays a still capture of the video that we used in our user study. We randomly selected three dancing sequences from our dataset for the comparison. Thirty-four people spared the time for us and answered our questionnaire. Each person selected the best match for three sequences. The user study results are summarized in Table 2.

The ratio in Table 2 represents the number of people who selected the keypose-based low-dimensional motion representation as the best match with the original motion compared to uniform sampling pose-based low-dimensional motion, out of the total number of participants who participated in the user study, for each dancing sequence. The percentage denotes the above ratio converted into a factor of percentage. The results demonstrate that a significantly higher percentage of participants selected the keypose-based low-dimensional motion representation as the best match over uniform sampling pose-based method. It also indicates that the keypose-based method can maintain the essential subtle factors of human motion in low-dimensional motion representation over the other method, as distinguished by human perception.

In addition to the above analyses, we also analyzed the actual low-dimensional motions (three dimensional) created according to Approach 2. Our results demonstrate that for some people the uniform sampling pose-based method produces unnatural human postures where the keypose-based method does not. Figure 13 illustrates an example of such an occurrence. In figure 13 the left side character, which represents the keypose-based low-dimensional motion, displays a natural posture, while the right side character, which represents low-dimensional motion based on uniform sampling poses, displays an unnatural posture. It shows that the right hand has moved inside the human body in the right side character. The middle character shows the original motion of the person.

Figure 14 describes another unnatural posture of uniform sampling pose-based low-dimensional representation. Similarly the left side character represents the

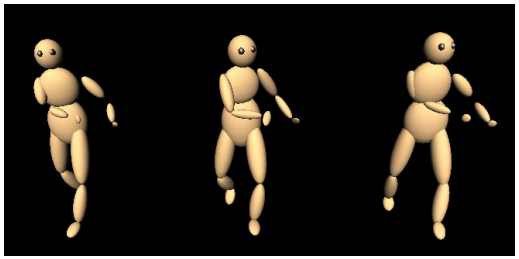


Fig. 13 An example of unnatural posture: The left side character shows the low-dimensional motion representation based on keyposes. The middle character displays the original motion. The right side character shows the low-dimensional motion obtained by uniform sampling poses.

keypose-based low-dimensional motion, the middle character original motion, and the right side character the uniform sampling pose-based low-dimensional motion. In figure 14 (a) the right side character shows an unnatural head and leg posture. Figure 14 (b), a side view of the same instance, clearly shows that the leg of the uniform sampling pose-based character has bent to the other side unnaturally. In the same instance, the keypose-based character shows a natural posture.

5. Recognition using Multifactor Tensor Analysis

5.1 Task Model

Vasilescu used tensor algebra[38] to recognize human motion signatures and applied it to motion synthesizing. Later Vasilescu et al. addressed the problem of multidimensional data analysis in image face recognition and tensor textures using multilinear algebra[39),40)]. Recently, several research projects have utilized multilinear models in solving problems[10),23),42)]. Our model presents a multi factor tensor (MFT)-based framework for decomposing motion data and recognizing tasks and motion styles or human identities. In this model, we follow a novel approach that considers higher order tensor factorization[16),17)].

As described above, dance motion consists of a set of motion sequences segmented by keyposes that occur after every musical rhythm interlude. In dance sequences, similar motion segments are often repeated in each dance cycle. Provided with a group of similar motion segments, we decompose them into a com-

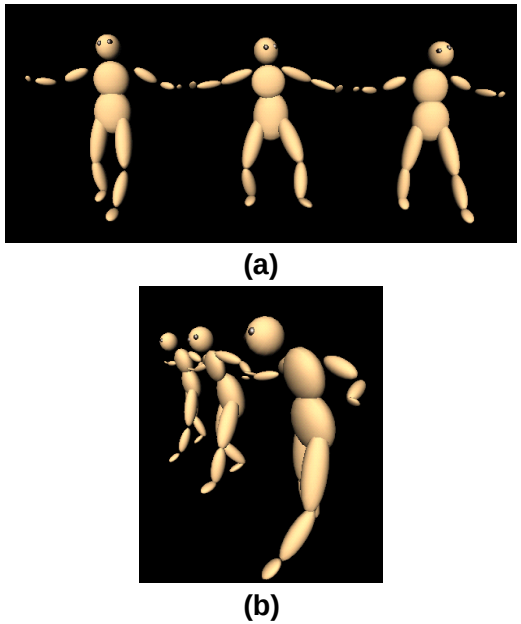


Fig. 14 Another example of unnatural posture: In (a) the left side character shows the low-dimensional motion representation based on keyposes. The middle character displays the original motion. The right side character shows the low-dimensional motion obtained by uniform sampling poses. (b) shows a side view of the same instance.

mon *person invariant* motion factor, which does not vary by person, and a *person dependent* factor, which does vary by person. We define *task* as the former common factor and *style* as the latter dependent factor. The framework of decomposing human motion into task and style is called a *task model*. In practice, we use MFT analysis to decompose human motion, and we apply the result in recognizing tasks and human motion styles.

Before formulating our MFT model, we normalize the human motion data as described in[26)]. The appearance of each person is made equal to a predefined physical model during the normalization.

5.2 Multi Factor Tensor (MFT) Analysis

We acquire motion-capture data of dance sequences of μ people for κ cycles

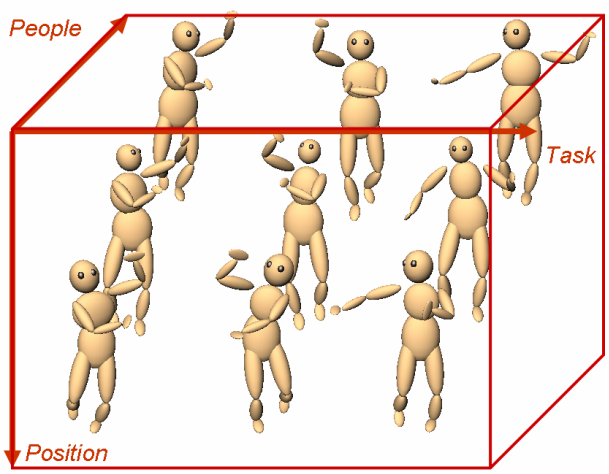


Fig. 15 Three Mode MFT: Formulation of third order MFT model where *position*, *people*, and *task* represent the first mode, the second mode, and the third mode respectively. The data tensor is formulated so that slicing the tensor parallel to the *task* axis produces slices that contain the data representing different tasks performed by the same person. Similarly, slicing the tensor parallel to the *people* axis produces slices that contain the data representing the same task performed by different people.

where each cycle has a duration of η frames. After the segmentation process using music analysis, we normalize the segments and arrange them appropriately into our data tensor, according to the different approaches described below. The arrangement of motion segments into the data tensor is shown in Figure 15. In the first approach, we select one task category from the motion-capture data for each person for learning in our model. In the second approach, we leave out some task categories from some people in formulating the model. We decompose the MFT using SVD[16] by flattening it in task-mode or people-mode, and factorize the motion data into *task* factor or *style* factor respectively. The task is defined as a person-invariant factor and the style as a person-dependent factor.

We consider a third order tensor:

$$\mathcal{P} = \mathcal{Q} \times_1 \boldsymbol{S}_{position} \times_2 \boldsymbol{S}_{people} \times_3 \boldsymbol{S}_{task}, \tag{12}$$

where $\mathcal{P}$ is the tensor obtained by arranging the samples associated with all the three dimensional factors and $\mathcal{Q}$ is the core tensor that controls the interaction between the three different mode factors: position, people, and task. The mode matrix $\boldsymbol{S}_{position}$ spans the parameter space of positions, $\boldsymbol{S}_{people}$ spans the parameter space of people, and $\boldsymbol{S}_{task}$ spans the parameter space of tasks. $\boldsymbol{S}_{position}$, $\boldsymbol{S}_{people}$, and $\boldsymbol{S}_{task}$ represent column-based orthonormal matrices. In general, if we think n-mode tensor, we denote $\tilde{\mathcal{Q}}_{\vartheta,i}$ and $\tilde{\mathcal{Q}}_{k,i}$

$$\tilde{\mathcal{Q}}_{\vartheta,i} = \mathcal{Q}_{\vartheta,i} \times_1 \boldsymbol{S}_1 \cdots \times_{q-1} \boldsymbol{S}_{q-1} \times_{q+1} \boldsymbol{S}_{q+1} \cdots \times_n \boldsymbol{S}_n$$
$$\tilde{\mathcal{Q}}_{k,i} = \mathcal{Q}_{k,i} \times_1 \boldsymbol{S}_1 \cdots \times_{q-1} \boldsymbol{S}_{q-1} \times_{q+1} \boldsymbol{S}_{q+1} \cdots \times_n \boldsymbol{S}_n,$$

where $\mathcal{Q}_{\vartheta,i}$ is the tensor belonging to the cluster being probed and $\mathcal{Q}_{k,i}$ is the tensor belonging to each cluster of training samples, and where $k = 1, \ldots, C$ and $i = 1, \ldots, N$ are the cluster and intra-cluster indices respectively. Here, intra-cluster means the sub-elements of each cluster, and each cluster comprises $i = 1, \ldots, N$ intra-cluster elements. $\tilde{\mathcal{Q}}_{\vartheta,i}$ and $\tilde{\mathcal{Q}}_{k,i}$ represent core tensors when the whole data tensor with n-mode factors is factorized in q-mode. The $\tilde{\mathcal{Q}}_{\vartheta,i}$ and $\tilde{\mathcal{Q}}_{k,i}$ core tensors have a complex relationship with $\boldsymbol{S}_{position}$, $\boldsymbol{S}_{people}$, and $\boldsymbol{S}_{task}$ mode matrices. A cluster, which represents a set of column vectors, characterizes a different person or a different task when collected according to the mode of factorization. ϑ means that the particular core tensor belongs to the cluster or class that has to be recognized. When our model is factorized in people-mode or task-mode, C indicates the number of persons or tasks contained in the training samples respectively. Using this factorization to consider the mode-q space, we can decompose the mode-q component as:

$$\hat{S}_q = \sum_i \sum_{k^*} ||\tilde{Q}_{\theta,i} \times_q S_q - \tilde{Q}_{k^*,i} \times_q S_q||^2, \tag{13}$$

where

$$k^* = \arg\min_k ||\tilde{Q}_{\theta,i} \times_q S_q - \tilde{Q}_{k,i} \times_q S_q||^2.$$

Here, $\hat{\boldsymbol{S}}_q$ represents a matrix that contains the minimum distances from the probed segment's intra-cluster elements to the respective intra-cluster elements of the training samples for mode-q factorization of the tensor. This matrix exists in the tensor subdomain space, with mappings to the corresponding cluster and intra-cluster indices. We compute the distances from each intra-cluster element of the probed segment to the corresponding intra-cluster element of the training samples in the tensor subdomain space. Thus we find out the minimum distance and the related training sample in the minimization argument.

5.3 Task and People Recognition

5.3.1 Recognizing Known Components

In this approach, we assume that all the motion segments to be probed are known to the data tensor. Given a motion segment to be probed, we decompose it using our framework and recognize its category or its identity. To recognize the task category, we flatten the data tensor in task-mode, and to detect the identity of the motion segment, we flatten the data tensor in people-mode. In our MFT model, we assume that the samples are represented as vectors $\boldsymbol{g}_i \in \mathbb{R}^\tau$, where $i = 1, \ldots, N$ is the intra-cluster index and τ is the dimension of $\boldsymbol{g}_i$. Our algorithm maximizes the following objective function in the recognition stage:

$$\begin{aligned} E^{\Psi}(\boldsymbol{g}_i, k, \hat{\boldsymbol{S}}_q|_{q=1}^n, W) &= \max_k \sum_{\boldsymbol{g}_i} W(\hat{\boldsymbol{S}}_q) \\ &= \max_k \sum_{g_i} \begin{bmatrix} w_1 & & \\ & w_2 & \\ & & \ddots \\ & & & w_N \end{bmatrix} (\hat{\boldsymbol{S}}_q). \end{aligned} \tag{14}$$

Here, k denotes the cluster index of the training sample, and w_i denotes a weight for the q-mode factorization of the model. k is dependent on the mode of factorization and it represents the person identity index or the task identity index of the training samples, depending on the mode of factorization in person-mode or task-mode respectively. The parentheses surrounding $(\hat{\boldsymbol{S}}_q)$ indicate the multiplication process. W is the weight matrix obtained by applying weight w_i to the diagonal elements of the identity matrix. A reasonable choice of values is selected for the weight matrix by observing the motion sequence beforehand. Weights are assigned in increasing order by considering the number of links to the body center.

For each training cluster in our objective function we compute the sum of the minimum distances, found out earlier in $\hat{\boldsymbol{S}}_q$ from each intra-cluster element of the probed segment to the corresponding intra-cluster element of the training samples. A higher total value of the sum of the minimum distances indicates that there are many intra-cluster elements, which are quite similar to the probed segment's intra-cluster elements that belong to the particular cluster. On the other hand, if there are fewer links from the body center to a position in the body, they have less influence on the difference of the motion. The level one positions, such as waist positions, which are directly connected to the body center with one link, have quite similar kinds of motion among different types of motions. Positions connected with two links have more dissimilarity compared to the level one positions. Regarding this factor and our assumption that there is no elasticity in the human body, we have assigned the weights in the weight matrix in order to produce a larger value when there is more influence from the particular intra-cluster element. Therefore, the cluster or the training sample that produces the largest objective function value of our model is the most similar category to the probed category. We can solve for the probed sequence using the above objective function.

5.3.2 Recognizing Alien Components

Unlike the previous approach, this approach has no assumptions. The motion segments to be probed can belong to any task or person category that is known or unknown to the data tensor. As in the first approach, we assume that the samples are represented as vectors $\boldsymbol{g}_i \in \mathbb{R}^\tau$, where $i = 1, \ldots, N$ is the intra-cluster index and τ is the dimension of $\boldsymbol{g}_i$. We use our MFT model to determine a functional value vector $\mathcal{F}_k^\star$ in the tensor subdomain that characterizes the variational difference, which is a set of the *relative style/task variational values* in the tensor subdomain, for each mode-factor of every element in the training set. We compute the relative style/task variational values of each category in the training set, in tensor subdomain space, by formulating the MFT model with a similar category for every category in the training set from a different cycle, and by flattening the tensor in the appropriate mode. Considering the q-mode factorization of the MFT model, we compute ${\mathcal{F}_k}^\star$ in the tensor subdomain as follows:

$$\begin{aligned} {\mathcal{F}_k}^\star = E^{\varphi}(\boldsymbol{g}_i, k, \hat{\boldsymbol{S}}_q|_{q=1}^n, W) &= \max_k \sum_{\boldsymbol{g}_i} W(\hat{\boldsymbol{S}}_q) + \varepsilon_k \\ &= \max_k \sum_{g_i} \begin{bmatrix} w_1 & & \\ & w_2 & \\ & & \ddots \\ & & & w_N \end{bmatrix} (\hat{\boldsymbol{S}}_q) + \varepsilon_k, \end{aligned} \tag{15}$$

where ε_k represents the error value in the tensor subdomain. ε_k value is deter-

mined experimentally. Given the motion sequence to be probed, we segment it with the musical analysis method formulated, as in the previous case, into our framework and compute $\mathcal{F}_{Pr}{}^{\star}$ as follows:

$$\begin{aligned}\mathcal{F}_{Pr}{}^{\star} &= E^{\varphi}(\boldsymbol{g}_i, Pr, \hat{\boldsymbol{S}}_q|_{q=1}^{n}, W) \\ &= \max_{Pr} \sum_{\boldsymbol{g}_i} W(\hat{\boldsymbol{S}}_q) \\ &= \max_{Pr} \sum_{g_i} \begin{bmatrix} w_1 & & \\ & w_2 & \\ & & \ddots \\ & & & w_N \end{bmatrix} (\hat{\boldsymbol{S}}_q). \end{aligned} \tag{16}$$

where $\mathcal{F}_{Pr}{}^{\star}$ is the functional value in the tensor subdomain for the motion segment to be probed and Pr indicates that the functional value of the model is computed while formulating the probed motion segment into our framework. We sequentially compare the distances from this value to the trained value vector within a threshold ε_{Pr}, that is specified and identify the category of the probed motion segment. If nothing is detected within the range of our specifications, we assume that the probed motion segment belongs to a new category that is not contained in our database.

5.4 Experiments

We applied the MFT model described in the previous section to task recognition and person identification problems following two approaches to demonstrate the efficiency of our algorithm. The experiments were conducted on dance motion sequences from the *Aizu-bandaisan*, acquired by the motion-capture system. The data were captured at the rate of 120 frames/second and the noise reduction was done using a Gaussian filter.

For the experiments, we segmented the motion-capture data as described above. We normalized the segmented motion data using the vectorization method. The normalized segments were used to formulate the MFT model. The motion sequence to be probed could be from any cycle of the dance performance. We conducted the experiments as follows.

5.4.1 Experiment 1

We followed the "Recognizing Known Components" method described in 5.3.1 in performing this experiment. In this approach, we assumed that the motion segment to be probed did not belong to any alien category such as an unknown task or an unknown person compared to the data categories that were already contained in the data tensor.

Table 3 Recognizing Known Components: Summary of the results in recognizing known categories.

Category	No. Of Motion Segments	Correctly Recognized	Accuracy
Task Recognition	48	47	97.91%
Person Recognition	48	42	87.5%

We selected the tasks from one cycle of motion data and from every person to formulate the MFT model in order to fulfill the assumption. As described above, we factorized the data tensor according to different mode spaces, such as *task*-mode space and *people*-mode space, and applied the MFT model to recognize the required probed sequences. To recognize to which task category the required motion segment belongs, we flattened the data tensor in *task*-mode. To examine the identity of the motion segment we flattened the tensor in *people*-mode. Provided that a motion segment belonged to any person at a different time instance, we recognized the task category and the identity of the person performing the task.

For the experiment, we used 48 motion segments as probe sequences and six tasks that were performed by eight persons during a different cycle and that were not used as training data. The average computation time for each recognition process took about 70 seconds. The recognition results using our model are displayed in Table 3.

5.4.2 Experiment 2

This experiment was done according to the "Recognizing Alien Components" method as explained in 5.3.2. In contrast to the "Recognizing Known Components" method, we have made no assumptions that the task category and the person category of the motion segment be familiar or contained in the data tensor beforehand. We recognized any alien motion segment category, which is any kind

Table 4 Recognizing Alien Components 1: Summary of the results in recognizing unknown tasks performed by unknown people.

Category	No. Of Motion Segments	Correctly Recognized	Accuracy
New Task Recognition	24	22	91.66%
New Person Recognition	24	20	83.33%

Table 5 Recognizing Alien Components 2: Summary of the results in recognizing known tasks performed by unknown and known people.

Category	No. Of Motion Segments	Correctly Recognized	Accuracy
Task Recognition of Alien Person	12	10	83.33%
Task Recognition of Known Person	12	11	91.66%

of person or task for which we have no information in our model. For this, we left out several people's data and selected four different tasks from four different people in formulating the MFT model. Our main attempt was to determine whether the motion segment to be probed was new to the database by following elimination with no match in the tensor subdomain variation value.

As in the previous experiment, we factorized the data tensor in different mode spaces, such as *task*-mode space and *people*-mode space, and applied the tensor model in recognizing whether the motion segment belonged to any new category. To recognize a new task of a known person, we flattened the data tensor in *task*-mode, and to recognize a new person performing a known task, we flattened the data tensor in *people*-mode. The thresholds ε_k and ε_{Pr} were set to 0.05 times ${\mathcal{F}_k}^\star$ and ${\mathcal{F}_{Pr}}^\star$ respectively. The average computation time for each recognition process took about 40 seconds. The results of 24 motion segments from different cycles where the categories were new to the data tensor are displayed in Table 4.

In addition to the above experiments, we conducted experiments to recognize categories known and unknown to the training samples, and to specify the category if known. We examined the recognition ability of the model by experimenting with known and unknown tasks performed by people known and unknown to the data tensor. Interestingly, unlike the previous approach where only one closest category was selected as the result in the "Recognizing Alien Components" method, we sometimes got several categories that lay within the specified functional variation value ranges in the tensor subdomain. The multiple categories recognized in this experiment are largely similar to each other, such as the same task performed by two different persons with quite similar motions. The reason for multiple results is that we examined whether the probed motion segment's relative functional variation value lay within the trained functional variation value ranges with thresholds ε_k and ε_{Pr}. We found that it was difficult to set the ideal thresholds to recognize only one category as they are determined experimentally. If the thresholds are not set appropriately, we might miss recognizing some categories due to noise or some other reasons. There is a trade-off between the processing time and the recognition accuracy while setting the ε_k and ε_{Pr} thresholds.

For fine tuning, we need to set the thresholds ε_k and ε_{Pr} appropriately, since the tasks that we use for the experiments are not separable completely. For task recognition, we set ε_k and ε_{Pr} thresholds as 0.12 times ${\mathcal{F}_k}^\star$ and ${\mathcal{F}_{Pr}}^\star$ respectively. We experimented with 12 motion segments that belonged to known tasks performed by persons unknown to the data tensor, and 10 were recognized correctly. We also experimented with 12 motion segments that belonged to known tasks performed by people known to the data tensor, and 11 were recognized correctly. The summary of the above results is displayed in Table 5. In each case, first we checked whether the probed segment belonged to a new category or not. If it were not new, then we determined in which functional variation value range in the tensor subdomain it lay. Where we had few selections within the specified ranges, the one with the least style variation in the tensor subdomain was selected.

6. Conclusion and Future Work

In this paper we introduced novel frameworks to demonstrate the importance of keyposes in human dance motions in addressing three vital topics: how to extract the keyposes, how to utilize the keyposes, and how to use the keyposes for recognition. To extract the keyposes we proposed a new energy function combined with a music analysis approach. We compared our keypose extraction results with dancing professionals' teachings, and the results demonstrated high accuracy. Further, to utilize the keyposes, we generated low-dimensional motions corresponding to keypose space and uniform sampling pose-based eigen space. We evaluated our framework with synthetic and human perception experiments, which demonstrated the impact on human perception of keypose based method in low-dimensional motion space over a uniform sampling pose method. The third topic that we presented illustrated how to recognize keyposes using MFT analysis.

Currently, we are interested in investigating fusion of keyposes and physiology ([33),34)]), and motion summary creation. We are also interested in exploring the reasons behind human perception in low-dimensional motion space and the percentage of motion data required to be distinguished by the human eye. Finally, we are interested in further investigating more important characteristics of keyposes and the impact keyposes have on a given motion space.

Acknowledgments

This work is supported in part by the Japan Science and Technology Corporation (JST) under the CREST project, and in part by the Grant-in-Aid for Scientific Research on Priority Areas of the Ministry of Education, Culture, Sports, Science and Technology.

References

1) Amaya, K., Bruderlin, A. and Calvert, T.: Emotion from motion, *Proceedings of Graphics Interface*, pp.222–229 (1996).
2) Arikan, O. and Forsyth, D. A.: Interactive Motion Generation from Examples, *Proceedings of ACM SIGGRAPH*, pp.483–490 (2002).
3) Assa, J., Caspi, Y. and Cohen-Or, D.: Action Synopsis: Pose Selection and Illustration, *Proceedings of ACM SIGGRAPH*, pp.667–676 (2005).
4) Barbic, J., Safonova, A., Pan, J.Y., Faloutsos, C., Hodgins, J.K. and Pollard, N.S.: Segmenting Motion Capture Data into Distinct Behaviors, *Proceedings of Graphics Interface*, pp.185–194 (2004).
5) Brand, M. and Hertzmann, A.: Style Machines, *Proceedings of the 27'th Annual conference on Computer Graphics and Interactive Techniques*, pp.183–192 (2000).
6) Fauvet, B., Bouthemy, P., Gros, P. and Spindler, F.: A Geometrical Keyframe Selection Method Exploiting Dominant Motion Estimation in Video, *Proceedings of International Conference on Image and Video Retrieval*, pp.419–427 (2004).
7) Gao, H. and Davis, J.: An Expressive Three-Mode Principal Components Model of Human Action Style, *Image and Vision Computing*, Vol.21, No.11, pp.1001–1016 (2003).
8) Gao, H. and Davis, J.: Recognizing Human Action Efforts: An Adaptive Three-Mode PCA Framework, *Proceedings of International Conference on Computer Vision*, pp.1463–1469 (2003).
9) Jenson, R. K.: Changes in segment inertia proportion between 4 and 20 years, *Journal of Biomechanics*, Vol.22, No.6/7, pp.529–536 (1989).
10) Jia, K. and Gong, S.: Multi-Resolution Patch Tensor for Facial Expression Hallucination, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (2006).
11) Kannapan, V. and Davis, J.: Expressive Features for Movement Exaggeration, *SIGGRAPH Conference Abstracts and Applications*, p.182 (2002).
12) Kannapan, V. and Davis, J.: Recognizing Human Action Efforts: An Adaptive Three-Mode PCA Framework, *IEEE Workshop on Motion and Video Computing*, pp.139–144 (2002).
13) Kovar, L. and Gleicher, M.: Automated Extraction and Parameterization of Motions in Large Data Sets, *Proceedings of ACM SIGGRAPH* (2004).
14) Kovar, L., Gleicher, M. and Pighin, F.: Motion Graphs, *Proceedings of ACM SIGGRAPH*, pp.473–482 (2002).
15) Kudoh, S.: Balance Maintenance for Human-like Models with Whole Body Motion, *Ph.D. Dissertation 2004*.
16) Lathauwer, L. D., Moor, B. D. and Vandewalle, J.: A multilinear singular value decomposition, *SIAM J. Matrix Anal. Appl.*, Vol.21, pp.1253–1278 (2000).
17) Lathauwer, L.D. and Vandewalle, J.: Dimensionality reduction in higher-order signal processing and rank-$(R_1, R_2, \ldots, R_N)$ reduction in multilinear algebra, *Linear Algebra and its Applications*, Vol.391, pp.31–55 (2004).
18) Lee, C.-S. and Elgammal, A.: Gait Style and Gait Content: Bilinear Model for Gait Recognition Using Gait-Resampling, *Proceedings of the 6th International Conference on Automatic Face and Gesture Recognition*, pp.147–152 (2004).
19) Lee, C.-S. and Elgammal, A.: Inferring 3D Body Pose from Silhouettes using Activity Manifold Learning, *Proceedings of IEEE Conference on Computer Vision and*

Pattern Recognition (2004).
20) Lee, C.-S. and Elgammal, A.: Separating Style and Content on a Nonlinear Manifold, *Proceedings of The IEEE International Conference on Computer Vision and Pattern Recognition* (2004).
21) Lee, J., Chai, J., Reitisma, P. S.A. and Pollard, N.S.: Interactive Control of Avatars Animated with Human Motion Data, *Proceedings of ACM SIGGRAPH*, pp.491–500 (2002).
22) Liu, F., Zhuang, Y., Wu, F. and Pan, Y.: 3D Motion Retrieval with Motion Index Tree, *Computer Vision and Image Understanding*, Vol.92, No.2-3, pp.265–284 (2003).
23) Liu, W., Lin, D. and Tang, X.: Hallucinating faces: TensorPatch super-resolution and coupled residue compensation, *Proceedings of Computer Vision and Pattern Recognition*, pp.478–484 (2005).
24) Loy, G., Sullivan, J. and Carlsson, S.: Pose-Based Clustering in Action Sequences, *Proceedings of IEEE International Workshop on Higher-Level Knowledge in 3D Modeling and Motion Analysis* (2003).
25) Park, M. and Shin, S.Y.: Example-Based Motion Cloning, *Computer Animation and Virtual Worlds*, Vol.15, No.3-4, pp.245–257 (2005).
26) Perera, M., Shiratori, T., Kudoh, S., Nakazawa, A. and Ikeuchi, K.: Task Recognition and Person Identification in Cyclic Dance Sequences with Multi Factor Tensor Analysis, *IEICE Transactions on Information and Systems*, No. 5, pp. 1531–1542 (2008).
27) Pullen, K. and Bregler, C.: Motion Capture Assisted Animation, *Proceedings of ACM SIGGRAPH*, pp.501–508 (2002).
28) Safonova, A., Hodgins, J.K. and Pollard, N.S.: Synthesizing Physically Realistic Human Motion in Low-Dimensional, Behavior-Specific Spaces, *Proceedings of ACM SIGGRAPH* (2004).
29) Shiratori, T., Nakazawa, A. and Ikeuchi, K.: Detecting Dance Motion Structure through Music Analysis, *Proceedings of IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, pp.857–862 (2004).
30) Shiratori, T., Nakazawa, A. and Ikeuchi, K.: The Structure Analysis of Dance Motions using Motion Capture and Musical Information, *IEICE Transactions on Information and Systems D-II (Japanese)*, No.8, pp.1583–1590 (2005).
31) Tenenbaum, J. B. and Freeman, W. T.: Learning Bilinear Models for Two Factor Problems in Vision, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp.554–560 (1997).
32) Tenenbaum, J.B. and Freeman, W.T.: Separating style and content with bilinear models, *Neural Computation*, Vol.12, pp.1247–1283 (2000).
33) Thoroughman, K.A. and Shadmehr, R.: Electromyographic correlates of learning an internal model of reaching movements, *Journal of Neuroscience*, Vol. 19, pp. 8573–8588 (1999).
34) Thoroughman, K.A. and Shadmehr, R.: Learning of action through adaptive combination of motor primitives, *Nature*, Vol.407, pp.742–747 (2000).
35) Troje, N.: Decomposing Biological Motion: A Framework for Analysis and Synthesis of Human Gait Patterns, *Journal of Vision*, Vol.2, No.4, pp.371–387 (2002).
36) Unuma, M., Anjyo, K. and Takeuchi, R.: Fourier Principles for Emotion-based Human Figure Animation, *Proceedings of ACM SIGGRAPH*, pp.91–96 (1995).
37) Urtasun, R., Glardon, P., Boulic, R., Thalmann, D. and Fua, P.: Style-Based Motion Synthesis, *Computer Graphics Forum*, Vol.23, No.4, pp.799–812 (2004).
38) Vasilescu, M.: Human Motion Signatures: Analysis, Synthesis, Recognition, *Proceedings of International Conference on Pattern Recognition*, pp.456–460 (2001).
39) Vasilescu, M. and Terzopoulos, D.: TensorTextures: Multilinear Image-Based Rendering, *Proceedings of ACM SIGGRAPH*, pp.336–342 (2004).
40) Vasilescu, M. and Terzopoulos, D.: Multilinear Independent Components Analysis, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (2005).
41) Vermaak, J., Pirez, P., Gangnet, M. and Blake, A.: Rapid Summarisation and Browsing of Video Sequences, *Proceedings of British Machine Vision Conference* (2002).
42) Wang, H., Wu, Q., Shi, L., Yu, Y. and Ahuja, N.: Out-of-Core Tensor Approximation of Multi-Dimensional Matrices of Visual Data, *Proceedings of ACM SIGGRAPH* (2005).
43) Yasuda, H., Kaihara, R., Saito, S. and Nakajima, M.: Motion Belts: Visualization of Human Motion Data on a Timeline, *IEICE Transactions on Information and Systems*, No.4, pp.1159–1167 (2008).
44) Zelnik-Manor, L. and Irani, M.: Event-Based Video Analysis, *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp.123–130 (2001).