

# Pose Selection for Animated Scenes and a Case Study of Bas-relief Generation

Meili Wang  
Northwest A&F University  
Xianyang 712100, China  
wml@nwsuaf.edu.cn

Shihui Guo  
Xiamen University  
Xiamen, China  
guoshihui@yeah.net

Minghong Liao  
Xiamen University  
Xiamen, China  
liao@xmu.edu.cn

Dongjian He  
Northwest A&F University  
Xianyang, China  
hdj168@nwsuaf.edu.cn

Jian Chang  
Bournemouth University  
United Kingdom  
jchang@bournemouth.ac.uk

Jian Zhang  
Bournemouth University  
United Kingdom  
jzhang@bournemouth.ac.uk

Zhiyi Zhang  
Northwest A&F University  
Xianyang, China  
zhangzhiyi@nwsuaf.edu.cn

## ABSTRACT

This paper aims to automate the process of generating a meaningful single still image from a temporal input of scene sequences. The success of our extraction relies on evaluating the optimal pose of characters selection, which should maximize the information conveyed. We define the information entropy of the still image candidates as the evaluation criteria.

To validate our method and to demonstrate its effectiveness, we generated a relief (as a unique form of art creation) to narrate given temporal action scenes. A user study was conducted to experimentally compare the computer-selected poses with those selected by human participants. The results show that the proposed method can assist the selection of informative pose of character effectively.

## CCS CONCEPTS

• **Software and its engineering** → *Software system models; Inheritance*; • **General and reference** → Performance;

## KEYWORDS

Action Snapshot, Information Entropy, Pose, Bas-relief

### ACM Reference format:

Meili Wang, Shihui Guo, Minghong Liao, Dongjian He, Jian Chang, Jian Zhang, and Zhiyi Zhang. 2017. Pose Selection for Animated

Scenes and a Case Study of Bas-relief Generation. In *Proceedings of CGI '17, Yokohama, Japan, June 27-30, 2017*, 6 pages. <https://doi.org/10.1145/3095140.3095171>

## 1 INTRODUCTION

A single still picture may indeed tell "a story proper," as claimed by Wolf [13] according to a narrow definition. Speidel [10] also argued that there exists the great potential of pictures to express temporal relationships. The creation of such static images is artistically challenging because it often involves multiple factors, such as behaviour, emotion, and storytelling considerations.

Body and actions convey important information, but action plays an even more essential role in creating a strong (aesthetic) impression [9, 10]. The goal of this research is to address this challenge and generate a still image for a virtual scene with animated character performances. Specifically, we need to select the optimal character pose to maximally convey the information contained in the scenario to the viewers. The method we proposed in this paper provides an informal summary of the whole series of animated scenes and serves as a reference for the users. The corresponding contributions are as follows:

- We select the optimal pose from an animation sequence by considering multiple factors that include local information (joint rotations) and global information (environmental contacts and inter-character interactions). The pose is selected to demonstrate significant changes in motion and relevant features, which contain the maximum information about the sequence of scenes.
- We validate the application of our approach in a case study for digital relief generation. We used the technology of three-dimensional(3D) printing to verify our approach and demonstrated our results in a variety of scenarios, ranging from dancing performance to sport activities. A user study is conducted to experimentally

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*CGI '17, June 27-30, 2017, Yokohama, Japan*

© 2017 Association for Computing Machinery.

ACM ISBN 978-1-4503-5228-4/17/06...\$15.00

<https://doi.org/10.1145/3095140.3095171>

evaluate computer-selected poses with a real human’s perception.

## 2 RELATED WORK

### 2.1 Pose Selection

In our example, the extraction and analysis of poses can guide the production of relief work and consequently help artists to generate a piece of fine art. The problem of selecting an optimal pose from an animation sequence is similar to key-frame extraction [6]. Key-frame extraction aims to extract and blend a series of frames to approximate the original motion. The number of keyframes is less than that of the original sequence, but is still greater than one. For example, researchers were able to select around 8% of the frames from motion capture sequences to create key-frame sequences [5]. In comparison, the generation of a relief must maximize the information conveyed in a single static posture [2].

Conventional methods selected the key frames by minimizing the errors between the original motion sequences and the reconstructed ones [5, 6]. However, the poses extracted in this way are most likely to be the most repeated poses, which, from the perspective of information theory, may not be optimal. In addition to this error-minimization framework, researchers proposed selecting staggered poses by encoding coordinated timing among movement features in the different body parts of a character [3]. This is based on the concept of *Extreme poses*, which perform perceptual events of significant motion changes [14].

### 2.2 Bas-relief Generation

Reliefs generated from 3D objects have been considered as a promising approach for creating bas-reliefs, allowing the reuse of existing 3D models. The challenge in this process is to visibly retain the fine details of an original 3D object while compressing its depths to produce an almost planar result[1].

The existing techniques are well designed and can produce visually pleasant and faithful reliefs while preserving appearance, accuracy, and details. These methods can preserve details and present a good visual effect, even with a high compression rate [8].

A large amount of attention has been paid to the effectiveness and efficiency of relief generation algorithms. Using techniques introduced in [7], digital reliefs can be generated in real time on a GPU or parallel system, so that a relief-style animation can be generated from a given 3D animation sequence. Researchers have introduced the gradient-based mesh deformation method, which can be used to generate plane surface bas-reliefs, curved surface bas-reliefs, and interactive shape editing of the bas-reliefs [15]. These techniques can generate relief animation sequences; however, for actual relief carving, it is not clear which pose to select to produce the physical relief. Current techniques do not address this question, even though it is an important consideration for artists during the process of relief creation.

## 3 METHODOLOGY

This section describes the algorithm for selecting the most informative pose from an animation sequence. Although there is no consensus about what determines a good pose, quality is intuitively related to how much information the poses give us about the whole performance. This paper proposed a novel method based on information theory to quantitatively evaluate the information contained in a pose.

In information theory, entropy  $H$  is the average amount of information ( $I$ ) contained in each message received. Here, a message stands for an event, sample, or character  $X$  drawn from the distribution of a data stream. This is mathematically formulated as the following expression [11]:

$$H(X) = E(I(X)) = E(-\log_b(P(X))) \quad (1)$$

$P(X)$  is the probability of a particular event  $X$ .

This work chooses  $b = 2$  (the unit of the entropy is a *bit*). When taken from a finite sample, the entropy can explicitly be written as:

$$H(X) = \sum_i P(X)I(X) = - \sum_i P(X)\log_b(P(X)) \quad (2)$$

Pose entropy  $H(X)$  considers the information contained in both the local and global features in the motion sequence. The local features refer to the joint rotation information and the global features refer to the event information. Such events include the contact between a foot and the ground plane and the interactions between characters. The pose (or a specific frame in an animation sequence) is selected as the optimal pose if its pose entropy  $H(X)$  has the maximum value of the all animation frames.

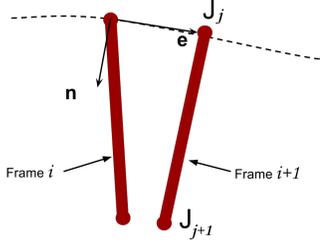
### 3.1 Local Features - Joint Information

A character pose is essentially a vector of the global transformation of the hip joint and the local rotations of other joints that represent the relative position of each joint to its parent joint. For the  $j^{th}$  joint, its motion trajectory is represented as  $\mathbf{m}_j \in R^{N_f \times 3}$  ( $N_f$  is the number of frames in this motion sequence). According to the concept of *extreme poses*, the information of a particular pose is emdeded in the changes of joint trajectories [3, 14]. Significant changes in the body poses create regions of higher curvature in the joint trajectory. A typical case is spinning in a ballet performance, where the local joints remain almost constant while the motion trajectory of each joint delivers its information in the world space.

The discrete measure of the rotational curvature for the  $j^{th}$  joint is defined in [3] as

$$\kappa_{ji} = \frac{\mathbf{n}_{ji} \cdot \mathbf{e}_{ji}}{\|\mathbf{e}_{ji}\|} \quad (3)$$

where  $\mathbf{n}_{ji}$  is the unit normal vector of the  $j^{th}$  joint at frame  $i$ , and  $\mathbf{e}_{ji}$  is the edge vector of the  $j^{th}$  joint between the frames  $i$  and  $i + 1$ , as illustrated in Figure 1. This returns a curvature in the range of  $[0, 1]$  and preserves the extrema of the curvature of the original motion data [3].



**Figure 1: Sketch of the motion trajectory of joint  $J_j$ . The information contained in the joint rotations is measured by the change in the curvature of the joint trajectories.**



**Figure 2: Detection of global features. In the example of running, when the foot strikes and lifts from the ground, the postures are selected and highlighted in the colour of cyan.**

Different joints have different effects on the overall behaviour of the character, and thus different significance with respect to visual information delivery. Therefore, the curvature is further weighted by the influence of the limb length and motion magnitude. The longer the limb is and the faster the joint rotates, the greater a weight is applied to the curvature at this frame.

$$\omega_{ji} = |\mathbf{x}_j - \mathbf{x}_{j+1}| \times |\Delta \mathbf{m}_{ji}| \quad (4)$$

Here,  $\mathbf{x}_j$  is the position of the  $j^{\text{th}}$  joint in world space and  $|\Delta \mathbf{m}_{ji}|$  is the joint angle difference between two frames.

The information conveyed by joint rotation in a particular frame is calculated as:

$$H_i(X_L) = -p_i^L \log_2(p_i^L) \quad (5)$$

$$p_i^L = \sum_{j=0}^{N_j} \omega_{ji} / \kappa_{ji}$$

where  $N_j$  is the number of joints.

### 3.2 Global Features - Event Information

In addition to joint information, global features, such as interaction with the environment and other characters, are also considered when extracting the most informative pose. In most cases, especially for task-based or context-based animation, it is the interaction with the environment and other characters that conveys the most information about the motion performed by the character.

Here, all skeleton joints are iterated to check the interaction with the environment and other characters. Previous work only considers the foot contact event [3]. The inclusion of other joints allows other cases to be considered, for example ball handling in the movements of basketball and hand-shaking with another virtual character. Interaction events are detected by searching for joints whose world coordinates remain constant with respect to a specific object (for example, the ground plane), within a given tolerance and for a given minimum length of time. After finding such an interaction, it is propagated to the following frames until the relative position between the joint and environment exceeds the tolerance.

Event information is modelled as a binary signal (1 indicates an interaction event and 0 indicates no interaction). After iterating through all poses, the interaction probability for each joint is modelled as

$$p_j^G = \frac{N_i}{N_f} \quad (6)$$

where  $N_i$  is the number of frames containing interaction. By formulating the problem in this way, the joint where less interaction occurs contains more information, and thus contains more entropy. The information of global events in a particular frame is calculated as:

$$H_i(X_G) = - \sum_{j=1}^{N_j} p_j^G \log_2(p_j^G) \quad (7)$$

### 3.3 Weighted Pose Entropy

The local and global information are assumed to be independent and a common approach for considering these two factors simultaneously is to use the weighted sum formulation:

$$H(X) = \omega_L H(X_L) + \omega_G H(X_G) \quad (8)$$

However, this formulation introduces an additional problem: how to properly set the weight values. The magnitude of each component  $H(X_L)$  and  $H(X_G)$  differs from the other, thereby making it difficult for users to choose the appropriate weight values  $\omega_L$  and  $\omega_G$ . To solve this problem, both the local and global components are normalized by their respective maximum values  $H(X_G)_{max}$  and  $H(X_L)_{max}$  and minimum values  $H(X_G)_{min}$  and  $H(X_L)_{min}$ , as follows [4]:

$$H(X) = \omega_L H^*(X_L) + \omega_G H^*(X_G) \quad (9)$$

$$H^*(X_L) = \frac{H(X_L) - H(X_L)_{min}}{H(X_L)_{max} - H(X_L)_{min}}$$

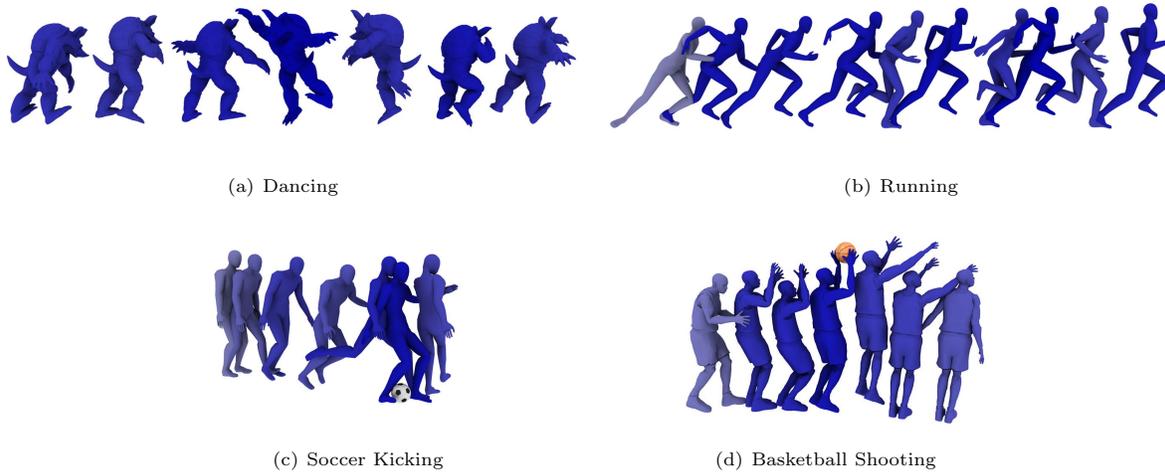
$$H^*(X_G) = \frac{H(X_G) - H(X_G)_{min}}{H(X_G)_{max} - H(X_G)_{min}}$$

After normalization, the values of each component will fall into a range of [0, 1]. In this case, the manipulation of the weights  $\omega_L$  and  $\omega_G$  directly relates the output of the pose selection to a preference for either local or global information.

Using Equation 9, a pose is selected as the most informative posture in an animation sequence if the entropy value of its frame is the maximum value of all frames in the animation sequence.

**Table 1: Statistics of the animation sequences used in this work. Time is in seconds**

Animation	Dancing	Running	Soccer Kicking	Basketball Shooting
No. Mesh Faces	75537	3191	3206	6061
No. Frames	1096	66	525	540
No. Joints	38	29	29	29
Time for pose selection	4.33	0.31	1.97	2.06

**Figure 3: Pose entropy for a variety of activities. Brighter colour indicates a pose with more information about the whole action.**

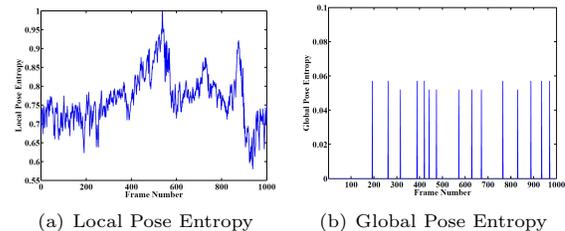
## 4 RESULTS

We used the Intel Xeon(R)(W3680) CPU (six cores clocked at 3.33GHz) to compute all our results. Table 1 presents a summary of the statistics for the demonstrated examples. As can be seen from the data, the time to find the optimal pose is largely determined by the number of frames and the number of performing characters.

### 4.1 Pose Selection

We apply our method for selecting the appropriate pose to a variety of human activities (see Figure 3 for results). The examples include ballet dancing, normal running, soccer kicking, shooting basketball hoops. Our method can successfully identify the critical events in an animation sequence with the assistance of the global feature when computing the pose entropy. This includes the moments of striking the soccer ball and letting go of the basketball.

Figure 4 presents both the local and global pose entropy for the ballet performance in Figure 5(a). Both terms are normalized into a range of  $[0, 1]$ . For the local pose entropy, we observe that the peaks occur at the moments when the character performs extreme poses. For the global pose entropy, the frames with non-zero values correspond exactly to the events of contacts between end-effectors (hands and feet) and the ground. In this example, the local and global entropy are

**Figure 4: Pose entropy of the armadillo ballet performance. (a) Local pose entropy. The three peaks correspond to the three selected poses in Figure 5(a). (b) Global pose entropy. The spikes indicate the frames where the interaction between end-effectors and environment happens.**

summed together with equal weights ( $\omega_{local}$  and  $\omega_{global}$  in Equation 9). By doing so, the local pose entropy outweighs the global component because of its larger value in this example.

### 4.2 Relief Generation

We here apply our selections of pose and viewpoint to a specific field: digital relief generation. This task is closely

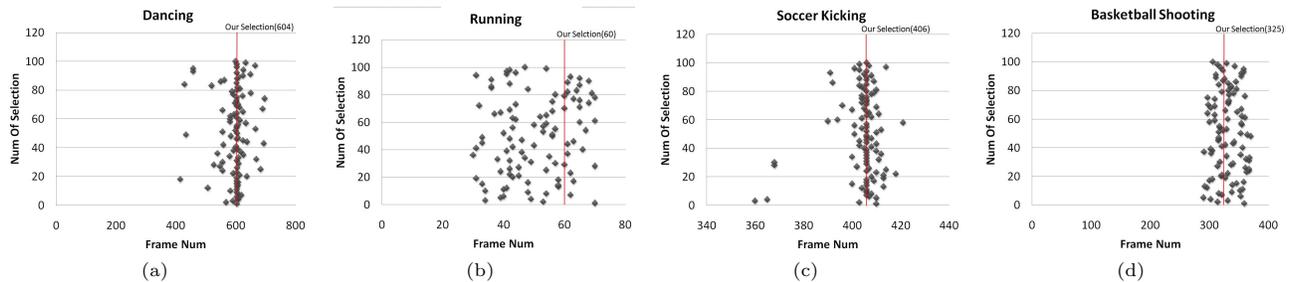


Figure 6: User study results for pose selection.



Figure 5: Relief generation of selected pose in Figure 3(a). (a) Digital relief. (b) 3D-printed relief.

related to our problem because that a piece of relief is an artistic work with embodied story-telling. The creation of relief is not only art-inspired but also technically challenging.

Once the optimal pose have been selected, we add saliency information into digital relief generation, as proposed in [12].

The generated bas-relief models and 3D printing results are shown in Figure 5.

### 4.3 User Study

We conducted a user study to validate and evaluate the outcomes of our proposed method. One hundred undergraduate students (50 male and 50 female) were hired as participants in this experiment. Before the experiment, the participants were informed of the content and procedure of the study and provided some background data. Participants were asked to watch six animation videos rendered at 24 frames per second and 1,920 x 1,080 pixels and select the most meaningful pose frames. The experimental results are shown in Figure 11.

**4.3.1 Discussion.** Figure 6 shows that our selected poses comply with the participant’s selection in most of the video sequences shown in Figure 3.

For Figure 5(a), Dancing, some samples are distributed near frame 430 because the ballet dancer has a slight leap in this period. In addition, the ballet dancer’s spinning starts at frame 580 and ends at about frame 620, so that most samples fall into this interval. However, the dancer performed the most stretched gesture at about frame 604, which agrees with our selection. For Figure 5(b), Running, this sequence simulates

running motion with high repeatability, so samples are evenly dispersed over the movement intervals of the runner.

In Figure 5(c), Soccer Kicking simulates someone before and after kicking a soccer ball. A few samples fall between frames 350 to 370 frame, because this is the interval in which the player sprinted before the kick. Most samples fall into the intervals before and after kicking the soccer ball, which falls between frames 400 to 420. However, the player started to touch the soccer ball around frame 406; thus, a sample is also gathered near frame 406. In Figure 5(d), Basketball Shooting, this group of actions simulates someone before and after shooting a ball into a hoop. An action showing a knee bend and jumping happens before and after shooting. In the animation, these main actions occur between frames 300 and 350, so most samples also centre on this interval.

## 5 CONCLUSION AND FUTURE WORK

This paper proposed a method to select an informative representative pose from an animation sequence. Several animation sequences were tested, and the results show that our method is applicable to scenarios involving one character and the performance of different types of activities.

A user study to validate the accuracy and effectiveness of the proposed method was conducted to experimentally compare the outcome of computer-selected poses with participants’ selections. The results showed that the proposed method outperformed other methods. To demonstrate the use of our proposed method, we used our meaningful results for bas-relief modelling.

Currently, our method is only tested for scenes with a single character. We want to extend our method to handle highly interactive scenes with a large number of characters, such as a carnival. This will be the focus of our future research.

## REFERENCES

- [1] Sami Arpa, Sabine Süsstrunk, and Roger D Hersch. 2015. High Reliefs from 3D Scenes. In *Computer Graphics Forum*, Vol. 34. Wiley Online Library, 253–263.
- [2] Volker Blanz, Michael J Tarr, Heinrich H Bülthoff, and Thomas Vetter. 1999. What object attributes determine canonical views? *Perception-London* 28, 5 (1999), 575–600.
- [3] Patrick Coleman, Jacobo Bibliowicz, Karan Singh, and Michael Gleicher. 2008. Staggered poses: a character motion representation for detail-preserving editing of pose and coordinated timing. In *Proceedings of the 2008 ACM SIGGRAPH/Eurographics*

- Symposium on Computer Animation*. Eurographics Association, 137–146.
- [4] Oleg Grodzevich and Oleksandr Romanko. 2006. Normalization and other topics in multi-objective optimization. In *Proceedings of the FieldsMITACS Industrial Problems Workshop*.
  - [5] Cihan Halit and Tolga Capin. 2011. Multiscale motion saliency for keyframe extraction from motion capture sequences. *Computer Animation and Virtual Worlds* 22, 1 (2011), 3–14.
  - [6] Ke-Sen Huang, Chun-Fa Chang, Yu-Yao Hsu, and Shi-Nine Yang. 2005. Key Probe: a technique for animation keyframe extraction. *The Visual Computer* 21, 8-10 (2005), 532–541.
  - [7] Zhongping Ji, Xianfang Sun, Shi Li, and Yigang Wang. 2014. Real-time Bas-Relief Generation from Depth-and-Normal Maps on GPU. In *Computer Graphics Forum*, Vol. 33. Wiley Online Library, 75–83.
  - [8] Jens Kerber, Meili Wang, Jian Chang, Jian J Zhang, Alexander Belyaev, and H-P Seidel. 2012. Computer Assisted Relief Generation A Survey. In *Computer Graphics Forum*, Vol. 31. Wiley Online Library, 2363–2377.
  - [9] Gotthold Lessing. 1976. Laocoon, or the Limits of Painting and Poetry. *Cosmopolitan Art Journal* 3, 2 (1976), 56–58.
  - [10] K. Speidel. 2013. Can a Single Still Picture Tell a Story? Definitions of Narrative and the Alleged Problem of Time with Single Still Pictures. *Diegesis* (2013).
  - [11] Pere-Pau Vázquez, Miquel Feixas, Mateu Sbert, and Wolfgang Heidrich. 2001. Viewpoint Selection using Viewpoint Entropy. In *VMV*, Vol. 1. 273–280.
  - [12] Meili Wang, Shihui Guo, Hongming Zhang, Dongjian He, Jian Chang, and Jian J Zhang. 2013. Saliency-based relief generation. *IETE Technical Review* 30, 6 (2013), 454–480.
  - [13] Wolf Werner. 2003. Narrative and narrativity: A narratological reconceptualization and its applicability to the visual arts. *Word & Image* 19, 3 (2003), 180–197.
  - [14] Richard Williams. 2009. *The Animator's Survival Kit: A Manual of Methods, Principles and Formulas for Classical, Computer, Games, Stop Motion and Internet Animators*. Macmillan.
  - [15] Yu-Wei Zhang, Yi-Qi Zhou, Xue-Lin Li, Hui Liu, and Li-Li Zhang. 2015. Bas-Relief Generation and Shape Editing through Gradient-Based Mesh Deformation. *Visualization and Computer Graphics, IEEE Transactions on* 21, 3 (2015), 328–338.