Action Snapshot: Pose and Viewpoint Selections

First Author · Second Author

Abstract Describing a scene via still images is important for a wide spectrum of fields, ranging from arts science. In this paper, we propose a concept, named *Action Snapshot*, to summarize the virtual scene with a single still image (snapshot). The input to our method is a scene which contains one or two characters, each performing an animation sequence. The goal of our method is to obtain the optimal pose of characters and viewpoint direction which is maximally meaningful in terms of information transmission. Our method is applicable to scenarios involving characters performing different types of activities. We use digital relief generation, one unique form of art creation, to validate our method. it can facilate the personalized artworks and souvenirs quickly prototyped for industry products. A user study is conducted to experimentally compare the outcome of computer-selected poses and viewpoints with participants' selection. The results show that the proposed method can assist in the selection of informative poses and perspectives from an animation-intensive scenario.

Keywords Scene Snapshot; Information Entropy; Pose Selection; Viewpoint Selection

1 Introduction

The need to display human activities in still images has long been one of the main forms of artistic creations. Examples include the cave painting, the oil painting, the relief sculptures and the thumbnail generation of animation and video sequences in recent years (Figure 1). The creation of such static images is artistically challenging since it often involves multiple factors,

Address(es) of author(s) should be given

(a)   (b)

Fig. 1: (a) Oil painting (*Oath of the Horatii*) by Jacques-Louis David. (b) Relief on the *Monument to the People's Heros* in Beijing.

such as behavior, emotion, and storytelling consideration. For example, in the case of relief sculpture, a sculptor needs to choose the best pose and perspective of a character (e.g. character a in Figure 1(b)). Such a problem is difficult to resolve because the selections of pose and viewpoint are often intertwined. Recent years saw a large numbers of productions of feature animation and visual effects films, which creates the increasing demand for optimal generation of thumbnail for these sequences.

The goal of this research is to address this challenge and generate a still image for a virtual scene with animated character performances. We name this image as *Action Snapshot*. Specifically, we need to select the optimal character pose and viewpoint to maximimally convey the information contained in the scenario to the viewers. It is worth pointing out that it is technically challenging, if not impossible, to provide a definite answer to such problem with artistic creation and subjective preferences. Therefore, a *snapshot* as proposed in this work is an informal summary of the whole scene and serves as a reference to users. We accomplish the goal of *scene snapshot* and make the following contributions accordingly:

– We select the optimal pose from an animation sequence by considering multiple factors, including 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 scene.

– We introduce the metric of Projected Motion Area (PMA) to quantitatively evaluate the information contained in a viewpoint direction and selects the viewpoint direction with the maximum value of PMA. Different from viewpoint selection for static geometry, the Projected Motion Area considers the whole animation sequences of multiple characters.

– We validate the application of our approach in a case study of digial relief generation in particular. We used the technology of 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 evaluate computer-selected poses and viewpoints with real human's perception.

The remainder of this paper is organized as follows. First, Section 2 reviews existing works related to pose and viewpoint selection. We then present details on how to select the optimal pose and viewpoint direction in Section 3. The results of our method are presented and evaluated by a user study analysis in Section 4. Section 5 concludes this paper with discussions of advantages and limitations of our approach.

2 Related Work

2.1 Pose Selection

Pose selection from an animation sequence plays an important role in terms of assisting artist to produce a piece of relief for carving. Therefore, extraction and analysis of the proper pose selection can guide the production of relief work and consequently help artists to generate a piece of fine artwork. The problem of selecting an optimal pose from an animation sequence is similar to key-frame extraction [9]. Key-frame extraction aims to extract and blend a series of frames to approximate the original motion. The number of key-frames is less than that of the original sequence, but still greater than one. For example, researchers are able to select around 8% of the frames from motion capture sequences to create key-frame sequences [8]. In comparison, the generation of a relief must maximize the information conveyed in a single static posture [3].

Conventional methods select the key frames by minimizing the errors between the original motion sequences and the reconstructed ones [9,8]. However, the poses extracted in this way are most likely to be the most-repeated poses, which, from the point of information theory, contain less information. In addition to this error-minimization framework, researchers have also proposed selecting the *staggered* poses by encoding coordinated timing among movement features in different body parts of a character [6].

2.2 Viewpoint Selection

A second problem addressed in this paper is to select an optimal projection direction given a selected pose. This is similar to the problem of viewpoint selection for a 3D mesh. To address the viewpoint selection problem, researchers have proposed various measures to quantitatively assess the goodness of a view. A collection of these measures (or *view descriptors*), including surface visibility, viewpoint entropy, silhouette stability, mesh saliency and symmetry, can be found in [21].

A related problem is how to determine an optimal camera path for an animated character motion. Some researchers have proposed to choose a camera path by maximizing the space swept out by the character skeleton [16]. This is useful for determining the viewpoint in a dynamic scenario. However, the generation of relief requires selecting the projection direction for a specific pose, rather than a whole sequence. Our method first selects the optimal pose and then determines the projection direction for the selected pose. By doing so, the problem of determining the projection direction shares its similarity with the conventional problem of viewpoint selection for a static mesh.

2.3 Bas-relief Generation

Digital fabrication is a way of preserving cultural heritage and one of its applications is relief generation [1, 17]. Reliefs generated from 3D objects have been considered as a promising approach to create bas-reliefs, allowing the reuse of existing 3D models. The challenge in this process is to visibly retain fine details of an original 3D object while compressing its depths to produce an almost planar result.

Song et al. [22] were the first to describe a method to generate bas-reliefs from 3D shapes automatically instead of projecting a 3D shape to the viewing plane. Subsequently, researchers developed different feature-preserving methods to generate bas-relief models with rich details [12,27,13,23,2,4,5]. The existing techniques

are well designed and can produce visually pleasant and faithful reliefs while preserving the appearance, accuracy and details. These methods can preserve details and present a good visual effect, even with a high compression rate [15]. Researchers also presented a novel method for bas-relief generation with additional intuitive style control [10].

A large amount of attention has been paid to the effectiveness and efficiency of relief generation algorithms. Using techniques introduced in [14,30,31,11], 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 introduced gradient-based mesh deformation method which could be used to generate plane surface bas-reliefs, curved surface bas-reliefs and shape editing of the bas-reliefs interactively [29]. 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 one of the important considerations for artists during the process of relief creation.

Recently, Schüller [20] proposed a unified framework to create bas-reliefs with the target shapes, viewpoints and space restrictions. Their approach is similar to our research; however, our work starts with an animation sequence, and proposes a novel framework that uniquely includes pose and viewpoint selections in our relief creation. This aspect distinguishes our work from its predecessors.

3 Methodology

Our goal is to generate a still image from a scene sequence of character animation. This involves two subtasks: selection of the *optimal* pose and the *optimal* viewpoint direction. The following sections explain our approaches to address these two tasks.

3.1 Pose Entropy and Selection

This section describes the algorithm to select the most informative pose from an animation sequence. Although there is no consensus about what determines a good pose, the quality is intuitively related to how much information they give us about the whole performance. This paper proposed a novel method, based on the *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 a distribution of data stream. This is mathematically formulated as the following expression [24]:

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

where b is the base of the logarithm and can be set to 2, e, or 10, depending on the unit of entropy. In addition, $I(X)$ denotes the information contained in a variable 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:

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

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 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 among the whole animation frames.

3.1.1 Local Feature - Joint Information

A character pose is essentially a vector of the global transformation of the hip joint and the local rotation of other joints that represent the relative position of each joint to its parent joint. Selection of the most informative pose is of practical application in the animation industry. *Extreme poses*, as animators call them, are ones where the characters perform perceptual events of significant motion changes [6,28]. These selected poses convey the information characterizing geometric aspects of the movement and are used as key-frames in an animation sequence. Similar to the work in [6], the information embedded in local joint rotations is evaluated based on the changes of motion trajectory for each joint.

For the jth joint, its motion trajectory is represented as $\mathbf{m}_j \in R^{N_{frame} \times 3}$ (N_{frame} is the number of frames in this motion sequence). Significant changes in the body poses create regions of higher curvature in the joint trajectory. A typical case is the 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 jth joint is defined in [6]

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

$\mathbf{n}_{ji}$ is the unit normal vector of the jth joint at the frames i, and $\mathbf{e}_{ji}$ is the edge vector of the jth joint between the frames i and $i+1$, as illustrated in Figure 2. This returns a curvature in a range of $[0, 1]$ and preserves the extrema of the curvature of the original motion data [6].

Different joints have different effects on the overall behavior 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 weight is applied to the curvature at this frame.

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

here $\mathbf{x}_j$ is the position of jth joint in world space, $|\Delta\mathbf{m}_{ji}|$ is the joint angle difference between two frames.

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

$$H_i(X_{local}) = -p_i^{local}log_2(p_i^{local}) \tag{5}$$

$$p_i^{local} = \sum_{j=0}^{N_{joint}} \omega_{ji}\kappa_{ji}$$
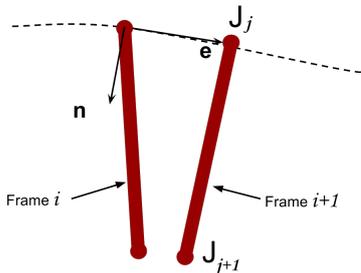
Fig. 2: Sketch of motion trajectory of joint J_j. The information contained in the joint rotations in measured by the change in the curvature of joint trajectories.

3.1.2 Global Feature - 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.

All skeleton joints are iterated to check the interaction with the environment and other characters. Previous work only considers the foot contact event [6]. The

Fig. 3: Detection of global features. The temporal events when environmental interactions happen are detected. Postures with such interactions are assumed to transfer greater information than those without such interactions. In the example of running, when the foot strikes and lifts from the ground, the postures are selected and highlighted in the color of cyan.

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, 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 modeled 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 modeled as

$$p_j^{global} = \frac{N_{interaction}}{N_{frame}} \tag{6}$$

where $N_{interaction}$ is the number of the frames containing interaction and N_{frame} is the total number of frames. By formulating the problem in this way, the joint where less interaction occurs contains more information, and thus contains more entropy:

$$H(X_{global}) = -\sum_{j=1}^{N_{joint}} p_j^{global}log_2(p_j^{global}) \tag{7}$$

3.1.3 Weighted Pose Entropy

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

formulation:

$$H(X) = \omega_{local}H(X_{local}) + \omega_{global}H(X_{global}) \tag{8}$$

However, this formulation introduces an additional problem: how to properly set the weight values. The magnitude of each component $H(X_{local}), H(X_{global})$ differs from the other, thereby making it difficult for users to choose the appropriate weight values $\omega_{local}, \omega_{global}$. To solve this problem, both the local and global components are normalized by their respect maximum $(H(X_{global})_{max}, H(X_{local})_{max})$ values and minimum $(H(X_{global})_{min}, H(X_{local})_{min})$ values [7]:

$$H(X) = \omega_{local}H^*(X_{local}) + \omega_{global}H^*(X_{global}) \tag{9}$$

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

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

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

Based on Equation 9, a pose is selected as the most informative posture in an animation sequence if the entropy value of this frame is the maximum value in the animation sequence.

3.2 Projected Motion Area and Viewpoint Selection

Finding the optimal viewpoint is as important as selecting meaningful poses for characters. To measure the goodness of a viewpoint, existing researchers have proposed a few *view descriptors*, such as projection area, viewpoint entropy, surface visibility and mesh saliency [21]. However, previous works generally select the viewpoint for a static geometry. In contrast, our task needs to consider the temporal information contained in an animation sequence.

In this work, we propose the concept of Projected Motion Area (PMA) to determine the optimal viewpoint of the action. *Motion area* is the area swept by a body link $L_{j/j+1}$ within a fixed time interval (normally between the ith and $(i+1)^{th}$ frames) [16]. Figure 4 sketches the motion area and its projection. The body link is connected by two joints where the jth joint is the parent of the $(j+1)^{th}$ joint. The trajectory of each joint in the world space are given as a curve with dimension $N \times 3$. A projection plane in the orthogonal projection is only dependent on the viewing direction. The projection matrix $\mathbf{P}(\nu)$ derives from the viewing direction ν can be expressed as $\mathbf{P}(\nu) = [\nu_1 \ \nu_2]^T$, where
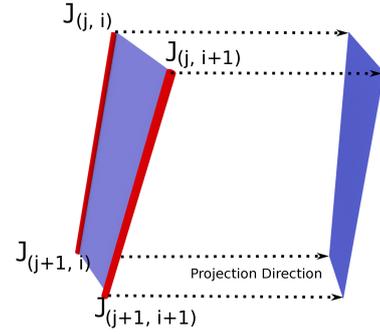
Fig. 4. Sketch of motion area (on the left) and its projection area (on the right) for one skeleton link (J_j, J_{j+1}) in two consecutive frames $(i, i+1)$. Weighted Projection Area is a weighted sum of the projection areas of all skeleton links in multiple frames.

ν_1, ν_2 are basis vectors of the projection plane and are both perpendicular to ν.

For a complete animation sequence, we sum up the projected motion area of all joints and frames:

$$P^{ma} = \sum_{i=1}^{N_{frame}} \sum_{j=1}^{N_{joints}} \frac{MA_{proj}}{MA_{orig}} \tag{10}$$

where MA_{proj}, MA_{orig} are the areas of the original *motion area* (the polygon on the left of Figure 4) and its *projection area* (the polygon on the right of Figure 4).

The optimal viewpoint V^* is found by computing P^{ma} for all sample points on the viewpoint sphere and selecting the one with the maximum P^{ma}:

$$V^* = \max_{\nu} \sum_{k=1}^{N_{char}} P^{ma}, \forall \nu \in \Omega \tag{11}$$

where N_{char} is the number of characters in the animation performance, and Ω is the set of all sample points on the viewpoint sphere.

4 Results

We used the Intel Xeon(R)(W3680) CPU (six cores clocked at @ 3.33GHz) to compute all our results. Table1 presents a summary of the statistics on the models and computing time. As can be seen from the data, the time cost to find the optimal pose is largely determined by the number of frames. After the pose is determined, the time cost to find the optimal viewpoint does not present significant differences. It is worth noting conventional methods of viewpoint entropy [24] computes the projection area for each triangular face, which means that the time cost for finding the optimal viewpoint grows linearly with the number of triangular faces. However in

Animation	Dancing	Basketball Dunking	Basketball Shooting	Soccer Kicking	Running	Two-person Arguing	Two-person Boxing
No. Mesh Faces	75537	3191	6061	3206	3191	6412	6412
No. Frames	1096	35	540	525	66	1992	150
No. Joints	38	29	29	29	29	58	58
Time for pose selection	4.33	0.25	2.06	1.97	0.31	10.62	2.37
Time for viewpoint selection	2.16	0.32	1.01	1.06	0.28	4.69	0.57

Table 1: Statistics of the animation sequences used in this work. The unit of time cost is second.

(a) Dancing

(b) Running

(c) Soccer Kicking

(d) Basketball Shooting

(e) Two-person Arguing

(f) Two-person Boxing
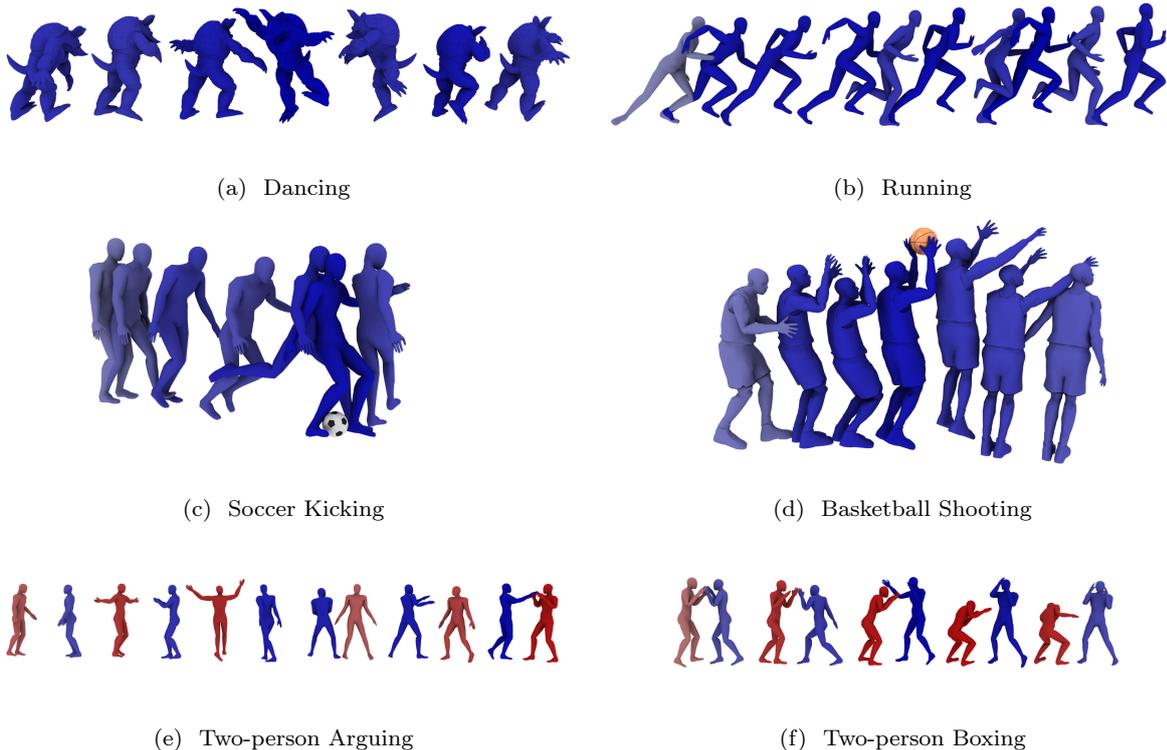
Fig. 5: Pose entropy in a variety of activities. A pose with higher colour brightness means a pose with more information about the whole action.

our method, the time cost to find the optimal viewpoint is dependent on the number of skeleton joints, instead of the number of mesh faces.

4.1 Pose Selection

We apply our method of selecting the appropriate pose to a variety of human activities (see Figure 5 for results). The examples include ballet dancing, normal running, soccer kicking, basketball shooting, arguing and boxing between two persons. 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 and releasing the basketball. In the example of two-person arguing, the critical event is identified when one person initializes body contacts with another person.

Figure 6 presents both the local and global pose entropy for the ballet performance in Figure 5a. 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 is performing 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 summed together with equal weights (ω_{local} and ω_{global} in Equation 9). By doing so, the local pose

entropy outweights the global component due to their larger value in this example of ballet performance.

(a) Local pose entropy   (b) Global pose entropy
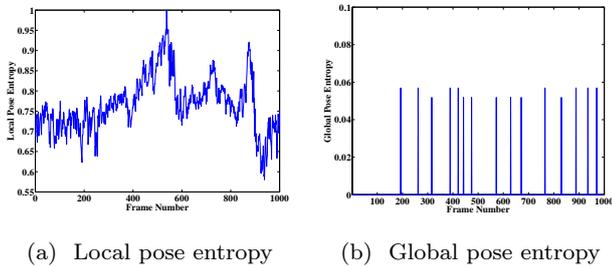
Fig. 6: Pose entropy of the ballet performance of armadillo. (a) Local pose entropy. The three peaks correspond to the three selected poses in Figure 5(a). (b) Global pose entropy. The spikes indicate the frames where the interaction between end-effectors and environment happens.

(a) Screenshot of animation sequence
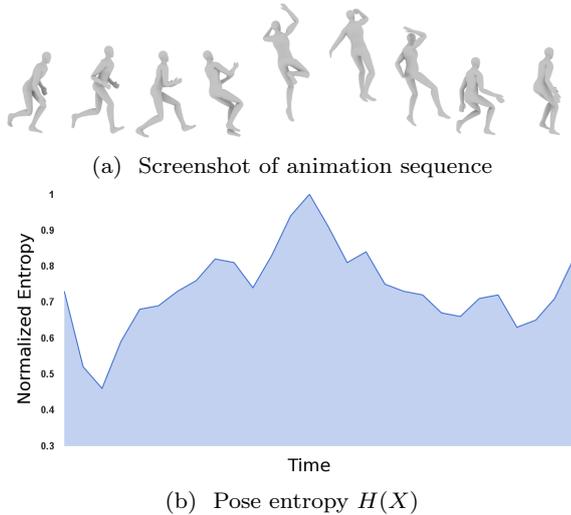
(b) Pose entropy $H(X)$

Fig. 7: Pose entropy of an animation sequence of performing a basketball dunk.

Figure 7 presents an example animation sequence of a basketball dunk, showing both the screenshots and pose entropy. The result shows that the extreme poses, when the character's links (both arms and legs) are fully stretched, have the highest pose entropies and thus are selected as the most informative poses.

4.2 Viewpoint Selection

After the optimal pose is determined, the next step is to find the appropriate viewpoint which conveys the information efficiently. The results in this section verify the effectiveness of the proposed method to determine the optimal viewpoint.

Three different methods were tested and compared with their selected projection direction (Figure 5(a-f)). The methods include our proposed method, [16] and [24]. The results show that the standard method to compute viewpoint entropy with surface visibility does not achieve a satisfactory result (Figure 9). In comparison, the results of our method (Figure 8) and the standard method of *motion area* (Figure 10) are similar. This is because the specific projection direction for this pose coincides with the global projection direction.

4.3 Relief Generation

We here apply our selections of pose and viewpoint to a specific field: digital relief generation. This task is closely related to our problem, since that a piece of relief is an artistic work with embodied story-telling. The creation of relief is not only art-inspiring but also technical challenging.

Once optimal pose and viewpoint have been selected, we add saliency information into digital relief generation as proposed in [25]. Saliency defines the level of significance of an observed local feature or object and links the distribution of human attention to visual encounters. In addition to a bilateral filter that integrates spatial and intensity information, we adopt a cross-bilateral filter that provides an extensible approach to generating bas-relief by taking different information into consideration, such as height, perspective direction, color, lighting and texture. The visual perception information is defined as the angle between the surface normal and view direction. The visual information is added as a third Gaussian kernel in our filter. An input models may be transformed directly into a piece of relief by a linear compression, but such an operation may cause local features to be lost and the resulting relief look dull because of the lack of incorporated depth information. Instead, the height information is processed with a non-linear compression as implemented in [25].

The technology of additive manufacturing, or commonly known as 3D Printing, is used to prototype the relief models in Figure 11 . The printer model is Projet 3510 SD. The 3D printer mesh volume is 298x185x203mm, the resolution is 345x375x790DPI, the precision is 0.025-0.05mm, and the material is VisiJet M3 X. The size of the printed bas-reliefs is 100mm×100mm×4mm. [1]

---

[1] the parameters is available at :http://www.magicfirm.com/professional-3d-printer/ProJet-3510SD
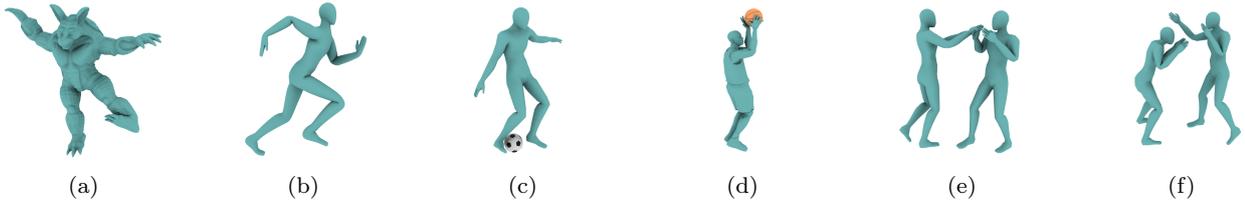
Fig. 8: Result of viewpoint selection for the selected poses in Figure 5 with our method.
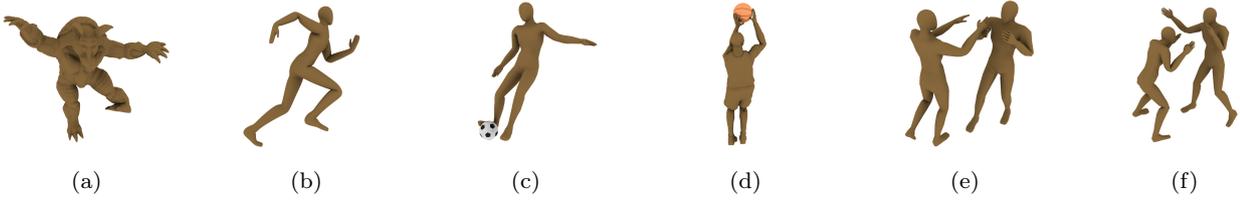
Fig. 9: Result of viewpoint selection for the selected poses in Figure 5 with the method in [21].
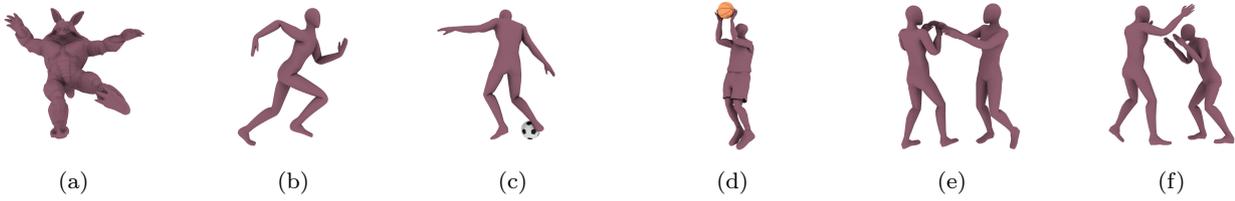
Fig. 10: Result of viewpoint selection for the selected poses in Figure 5 with the method in [26].
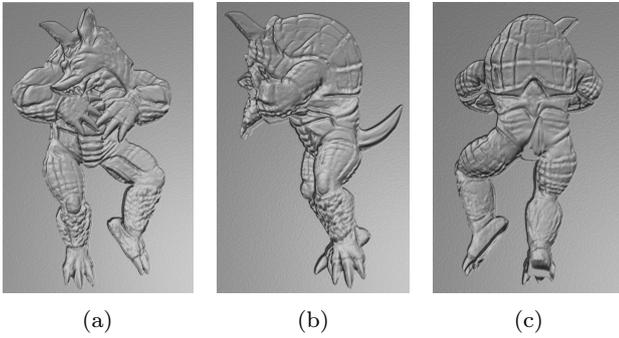
Fig. 11: Relief generation with different projection direction for Pose 1 in Figure 5(a). Each corresponds to the selected projection direction in Figure 5(a) respectively.

Fig. 12: 3D printed reliefs with different projection direction for Pose 1 in Figure 11.

the content and procedure of the study and fill in some background data.

4.4 User Study

User studies play a similar fundamental role in evaluating the accuracy and applicability of the proposed method. Therefore, we conducted a user study to validate and evaluate the outcomes of our proposed method [19, 18]. 100 undergraduate students (50 male and 50 female) are hired as participants in this experiment. Before the experiment, the participants were informed of

4.4.1 Conduction

The user study is designed as two parts, one is for the pose selection, and the other one is for the viewpoint selection according to the selected poses.

(a) Pose selection

Participants were asked to watch 6 animation videos rendered as 24 frames per second and 1920x1080 pixels
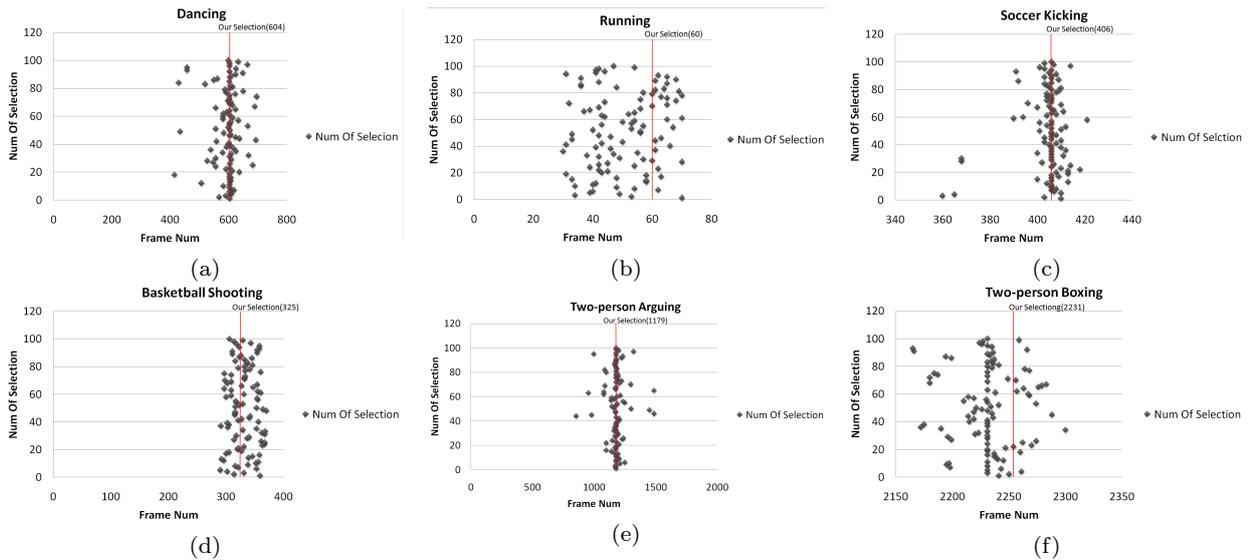
Fig. 13: User study results of pose selection.

Animation video	(a)Dancing	(b)Running	(c)Soccer Kicking	(d)Basketball Shooting	(e)Two-person Arguing	(f)Two-person Boxing
Figure 8	52	35	64	47	54	55
Figure 9	8	33	27	0	39	23
Figure 10	40	32	9	53	7	22

Table 2: User study results of viewpoint selection.

and select the most meaningful pose frame respectively. The experimental results is shown in Figure 13.

(b) Viewpoint selection

Participants were asked to compare Figure 8-10 to select a most meaningful viewpoint according to the selected poses as indicated in Figure 5.

4.4.2 Discussions

(a) Pose selection

It can be seen from Figure 13 that the selected poses comply with the participant's selection in the most video sequences as shown in Figure 5.

For (a)Dancing, some samples are distributed near 430 frames because the ballet dancer has a slight leap in this period. While the ballet dancer spinning starts from 580 frames and ends at about 620 frames, so that most samples fall into this interval. However, The dancers showed the most stretch gesture at about 604 frames that is complying with our selection. For (b)Running, this group imitating running motion with high repeatability, samples are evenly dispersed in the movement interval of the runner. For (c)Soccer Kicking, it imitates someone before and after playing soccer. A few samples fall into 350 frames to 370 frame, because this person during this period of time to make a soc-

cer sprinting before the gesture. While most samples fall into the interval before and after playing soccer by the body, that is between 400 to 420 frames. However, the body started in touch with soccer around 406 frames, thus the sample also gathers near 406 frames. For (d)Basketball Shooting, this group of actions imitates someone before and after pitching. An action showing knee bend and jumping happens before and after pitching. In instances of the animation, these main actions occur between 300 and 350 frames, so most samples also center on this interval. For (e)Two-person Arguing, the group of actions imitates two figures in quarrel. As two figures begin to approach and the distance shortens for the first time at about 850 frames, a few of people choose this frame. When two figures have relatively large amplitude of motions between 1150 and 1380 frames, most samples fall into this interval. Physical touch by two figures starts at 1240 frames. Due to large hidden part, most people don't choose motions after 1240 frames, while two figures are some distance apart around 1470 fames with good posture extending, so the survey finds there are still a few people choosing this period. For (f)Two-person Boxing, this group of actions imitates two figures in fighting. The whole process demonstrates obvious motions and rich body language, so samples distribute evenly. Owing to fierce

interaction between two figures near 2231 frames showing obvious knee bend and shake fist performance, etc, many samples select 2231 frames as our selection.

(b) Viewpoint selection

It can be seen from Table 2 that our proposed method is outperformed than [21,26] except for (b)Running and (d)Basketball Shooting. Three methods all work fine for (b)Running. For (d)Basketball Shooting, Figure 9(b) only shows the back side, while Figure 9(a) and Figure 9(c) are both fine to show the optimal pose, for the perspective viewpoint selection, it is largely depended on the environment whether the character is shooting the ball or passing the ball. For (a)Dancing, Figure 8(a) presents the side of the head, which can provide more information than the front view as shown in Figure 10(a). For (c)Soccer Kicking, Figure 9(c) does not show the optimal pose, while Figure 10(c) shows the back side. For (e)Two-person Arguing and (f)Two-person Boxing, Figure 8(a) are able to distinguish the poses, while the other two part of the body are occluded.

5 Conclusion and Future work

This paper proposed a method to select an informative pose and projection direction from an animation sequence. A concept- Action Snapshot was put forward to obtain the optimal poses of characters and viewpoint direction which are maximally meaningful during the information transmission of an animation sequence. Several animation sequences were tested shown that our method was applicable to scenarios involving one or two characters and performing different types of activities.

A user study validated the accuracy and effectiveness was conducted to experimentally compare the outcome of computer-selected poses and viewpoints with participants' selection. The results showed that the performance of the proposed method is outperformed than others.

To demonstrate the usage of our proposed method, we apply our meaningful Action Snapshot results for bas-relief modeling. Because relief creation is, at its heart, an artistic process, it is difficult to provide a uniform criterion for its aesthetics. The goal of this work is not to replace the creative process of professionals artists, but rather an approach to allow nonprofessionals and industry practitioners to fast-prototype their design. As 3D printing becomes more popular and widely available, personalized artworks and souvenirs can be quickly prototyped for industry products.

Our methods can be further developed to handle a number of characters playing complicated activities.

One of the most challenge works is to XXX. It will be the focus of our future research.

References

1. Arpa, S., Süsstrunk, S., Hersch, R.D.: High reliefs from 3d scenes. In: Computer Graphics Forum, vol. 34, pp. 253–263. Wiley Online Library (2015)
2. Bian, Z., Hu, S.M.: Preserving detailed features in digital bas-relief making. Computer Aided Geometric Design **28**(4), 245–256 (2011)
3. Blanz, V., Tarr, M.J., Bülthoff, H.H., Vetter, T.: What object attributes determine canonical views? Perception-London **28**(5), 575–600 (1999)
4. Bo, L., Shenglan, L., Liyan, Z.: Bas-relief generation using manifold harmonics analysis. Journal of Computer-Aided Design & Computer Graphics **2**, 020 (2012)
5. Bo, L., Shenglan, L., Liyan, Z.: Detail-preserving bas-relief on surface from 3d scene. Journal of Computer-Aided Design: Computer Graphics **24**(6), 799–807 (2012)
6. Coleman, P., Bibliowicz, J., Singh, K., Gleicher, M.: 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, pp. 137–146. Eurographics Association (2008)
7. Grodzevich, O., Romanko, O.: Normalization and other topics in multi-objective optimization. In: Proceedings of the Fields–MITACS Industrial Problems Workshop (2006)
8. Halit, C., Capin, T.: Multiscale motion saliency for keyframe extraction from motion capture sequences. Computer Animation and Virtual Worlds **22**(1), 3–14 (2011)
9. Huang, K.S., Chang, C.F., Hsu, Y.Y., Yang, S.N.: Key probe: a technique for animation keyframe extraction. The Visual Computer **21**(8-10), 532–541 (2005)
10. Ji, Z., Ma, W., Sun, X.: Bas-relief modeling from normal images with intuitive styles. Visualization and Computer Graphics, IEEE Transactions on **20**(5), 675–685 (2014)
11. Ji, Z., Sun, X., Li, S., Wang, Y.: Real-time bas-relief generation from depth-and-normal maps on gpu. In: Computer Graphics Forum, vol. 33, pp. 75–83. Wiley Online Library (2014)
12. Kerber, J., Belyaev, A., Seidel, H.P.: Feature preserving depth compression of range images. In: Proceedings of the 23rd spring conference on computer graphics, pp. 101–105. ACM (2007)
13. Kerber, J., Tevs, A., Belyaev, A., Zayer, R., Seidel, H.P.: Feature sensitive bas relief generation. In: Shape Modeling and Applications, 2009. SMI 2009. IEEE International Conference on, pp. 148–154. IEEE (2009)
14. Kerber, J., Tevs, A., Belyaev, A., Zayer, R., Seidel, H.P.: Real-time generation of digital bas-reliefs. Computer-Aided Design and Applications **7**(4), 465–478 (2010)
15. Kerber, J., Wang, M., Chang, J., Zhang, J.J., Belyaev, A., Seidel, H.P.: Computer assisted relief generation—a survey. In: Computer Graphics Forum, vol. 31, pp. 2363–2377. Wiley Online Library (2012)
16. Kwon, J.Y., Lee, I.K.: Determination of camera parameters for character motions using motion area. The Visual Computer **24**(7-9), 475–483 (2008)
17. Li, Z., Wang, S., Yu, J., Ma, K.L.: Restoration of brick and stone relief from single rubbing images. Visualization and Computer Graphics, IEEE Transactions on **18**(2), 177–187 (2012)

18. Raneburger, D., Popp, R., Alonso-Ríos, D., Kaind, H., Falb, J.: A user study with guis tailored for smartphones and tablet pcs. In: Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, pp. 3727–3732. IEEE (2013)
19. Santos, B.S., Dias, P., Silva, S., Ferreira, C., Madeira, J.: Integrating user studies into computer graphics-related courses. Computer Graphics and Applications, IEEE **31**(5), 14–17 (2011)
20. Schüller, C., Panozzo, D., Sorkine-Hornung, O.: Appearance-mimicking surfaces. ACM Transactions on Graphics (TOG) **33**(6), 216 (2014)
21. Secord, A., Lu, J., Finkelstein, A., Singh, M., Nealen, A.: Perceptual models of viewpoint preference. ACM Transactions on Graphics (TOG) **30**(5), 109 (2011)
22. Song, W., Belyaev, A., Seidel, H.P.: Automatic generation of bas-reliefs from 3d shapes. In: Shape Modeling and Applications, 2007. SMI'07. IEEE International Conference on, pp. 211–214. IEEE (2007)
23. Sun, X., Rosin, P.L., Martin, R.R., Langbein, F.C.: Bas-relief generation using adaptive histogram equalization. Visualization and Computer Graphics, IEEE Transactions on **15**(4), 642–653 (2009)
24. Vázquez, P.P., Feixas, M., Sbert, M., Heidrich, W.: Viewpoint selection using viewpoint entropy. In: VMV, vol. 1, pp. 273–280 (2001)
25. Wang, M., Guo, S., Zhang, H., He, D., Chang, J., Zhang, J.J.: Saliency-based relief generation. IETE Technical Review **30**(6), 454–480 (2013)
26. Wang, W., Gao, T.: Constructing canonical regions for fast and effective view selection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4114–4122 (2016)
27. Weyrich, T., Deng, J., Barnes, C., Rusinkiewicz, S., Finkelstein, A.: Digital bas-relief from 3d scenes. In: ACM Transactions on Graphics (TOG), vol. 26, p. 32. ACM (2007)
28. Williams, R.: The Animator's Survival Kit: A Manual of Methods, Principles and Formulas for Classical, Computer, Games, Stop Motion and Internet Animators. Macmillan (2009)
29. Zhang, Y.W., Zhou, Y.Q., Li, X.L., Liu, H., Zhang, L.L.: Bas-relief generation and shape editing through gradient-based mesh deformation. Visualization and Computer Graphics, IEEE Transactions on **21**(3), 328–338 (2015)
30. Zhang, Y.W., Zhou, Y.Q., Zhao, X.F., Yu, G.: Real-time bas-relief generation from a 3d mesh. Graphical Models **75**(1), 2–9 (2013)
31. Zhou, S., Liu, L.: Realtime digital bas-relief modeling. Journal of Computer-Aided Design & Computer Graphics **22**(3), 434–439 (2010)