EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2019)
S. Rizvic and K. Rodriguez Echavarria (Editors)

Approximate Reconstruction of 3D Scenes From Bas-Reliefs

P. Casati[1], R. Ronfard[1], and S. Hahmann[1]

[1]Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.

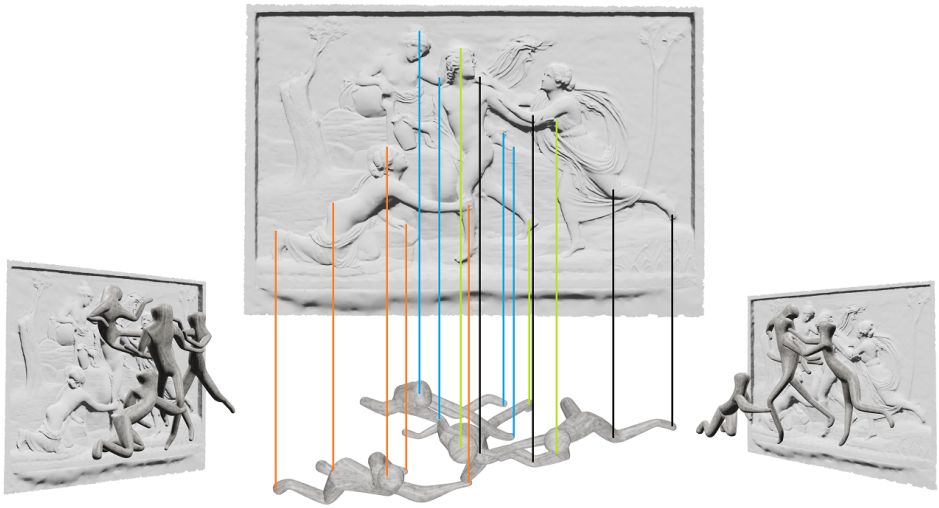


Figure 1: *3D interpretation of the mythological story of Hylas and the Water Nymphs, after a bas-relief marble by Bertel Thorvaldsen (1833). Hylas was sent to fetch water for the camp. Finding a pool in a clearing, he was encircled by water nymphs reaching up to kiss him and there disappeared with them forever. Using hand-drawn silhouette shapes and 2D skeletons of the four characters, we compute a plausible 3D reconstruction of the scene with rigged and skinned models suitable for 3D animation.*

Abstract

For thousands of years, bas-reliefs have been used to depict scenes of everyday life, mythology and historic events. Yet, the precise geometry of those scenes remains difficult to interpret and reconstruct. Over the past decade, methods have been developed for generating bas-reliefs from 3D scenes. In this paper, we investigate the inverse problem of interpreting and reconstructing 3D scenes from their bas-relief depictions. Even approximate reconstructions can be useful for art historians and museum exhibit designers, as a first entry to the complete interpretation of the narratives told in stone or marble. To create such approximate reconstructions, we present methods for extracting 3D base mesh models of all characters depicted in a bas-relief. We take advantages of the bas-relief geometry and high-level knowledge of human body proportions to recover body parts and their three-dimensional structure, even in severe cases of contact and occlusion. We present experimental results for 6 bas-relief depictions of Greek mythological and historical scenes involving 18 characters and draw conclusions for future work.

CCS Concepts

•***Computing methodologies*** → ***Mesh models; Mesh geometry models;***

1. Introduction

A bas-relief is a kind of sculpture that represents one or multiple scenes from the viewpoint of the artist. These scenes are flattened into the background plane, usually composed of marble or stone, where the bas-relief is carved. Alternatively, bas-reliefs can be sculpted with clay. Some parts are going out of the plane, according to the depth order of the scene: a figure that appears to be on the foreground is actually carved with more relief out of the plane than

DOI: 10.2312/gch.20191356

a figure in the background. Seen from the correct perspective, bas-reliefs provide a vivid perception of the 3D scene while using a very narrow range of depth values. Importantly, a bas-relief can always be represented with a single height field relative to the background surfaces (usually a plane). In contrast, high-relief sculptures contain sculpture in the round elements, which cannot be represented as a height field, although they are still attached to the background surface.

Bas-relief sculpture is a universal artistic form which can be found in all ages, parts of the world and civilizations. Bas-reliefs have been used for decorative, abstract or narrative purpose. Bas-reliefs, more than the classic sculptures (called sculpture *'in the round'*), often represent complex scenes, with a complex background, many characters, and the narration of a complete story. Bas-relief sculptures have been used to depict scenes of daily life, mythological stories and historical events carved into marble or sculpted into clay. This has led some authors to recognize bas-relief sculptures as precursors of modern animation [Ben16, Sif18].

In the case of ancient Greek and Roman bas-reliefs, deciphering the story told in a bas-relief may be difficult for the general public. To remedy this situation, museums frequently provide explanatory material such as captions, artistic drawings and sketches.

Some bas-relief sculptures may also allow multiple narrative interpretations. The Parthenon frieze is a good example where art historians dispute whether the story is historical or mythological [Con96] and propose diverse and contradictory interpretations [Joh04].

Reconstructing bas-relief sculptures in full 3D can help understand and explain the story told in the original artwork better than captions and sketches. A good example is the *Parthenon Project Japan* project [Osa16], where students in artistic anatomy sculpted miniature three-dimensional models of the twelve gods of the Parthenon's east frieze. The models were then displayed at the British Museum in front of the actual frieze to explain how the overlapping bodies of the gods filled space in the frieze.

In this paper, we propose a computational approach to reconstruct in 3D the figures depicted in a bas-relief, as an alternative to actually sculpting them for the purpose of explaining and commenting the original artwork.

Figure 2: *Side by side, a bas-relief, and its 3D scene reconstitution*

Presenting bas-reliefs in a museum exhibition presents several challenges. First of all, they can only be watched from a narrow range of viewing directions. More importantly, they provide only a partial view of the 3D scene, and leave a wide margin of interpretation to the imagination of the viewers. For a younger audience, it may be beneficial to augment the original artworks with solid shape representations and 3D animations allowing to better understand the artistic style, story and 3D structure of the scene. Turning around the scene, and visualizing from another point of view could put into perspective some details of the story. It will also open possibilities to create new sculptures.

But figuring out what a bas-relief is narrating and representing is not that easy. Firstly, among the collection of bas-reliefs got by museums, many of them have suffered from the ravages of times. And even without that, interpreting a bas-relief is complex. Only the elements that are facing the spectator can be recovered. Details that are not on the foreground can be overlapped. And it is not always possible to make a difference between a larger figure in the background and a smaller figure in the foreground. The decision is led by the understanding we have of the bas-relief, which is highly subjective.

Our work proposes methods for generating an approximate 3D scene from its bas-relief depiction, using a combination of high-level user input and low-level analysis of the bas-relief surface geometry. Starting from a digital version of the bas-relief, as a textured 3D mesh, the user decomposes the scene into a set of human figures, by drawing their silhouette shapes and 2D skeletal structures, including hidden body parts and joints. We then generate a 3D interpretation of the scene by lifting the 2D skeletons to 3D, reconstructing an approximate geometry for each figure, and placing them correctly in relation to each other.

To summarize, we propose:

- to estimate a 3D skeletal pose for a figure, based on the 2D pose, some proportions, and the bas-relief geometry.
- to parametrize the generation of a mesh, around the 3D skeleton.
- a method to estimate the depth position of each figure in the scene.

In Section 2, we discuss related work on bas-relief generation and 3D reconstruction from images. Then, we present an overview of our method in Section 3. We describe in more details how we extract 3D data to reconstruct human figures in Section 4, and how we arrange them to compose the 3D scene in Section 5. Finally, in Section 6, we present experimental results on a varied set of bas-reliefs with increasing complexity.

2. Related Work

There is a vast literature in computer generation of bas-relief sculptures from digital 3D scenes [CMS97, WDB*07] and we refer the reader to a recent survey for a full review of those methods [KWC*12]. Since 2012, new methods have been proposed for generating bas-reliefs from 3D scenes in real time [ZZZY13], or with improved textures [MSF*18], or using normal maps [JMS13, WTP*18], or using photographs [WMR*13].

To the best of our knowledge, the inverse problem of reconstructing 3D scenes from bas-reliefs by computer has not been addressed directly in previous work. Rossi [Ros17] describes the painful process of manually reconstructing an antique city based on its bas-relief depiction, but does not investigate the possibility of doing it automatically. In this paper, we attempt to fill this void.

Other authors [SLL07, ZTS09, CCL*11] have described methods for reverse engineering bas-reliefs, i.e. extracting geometric models

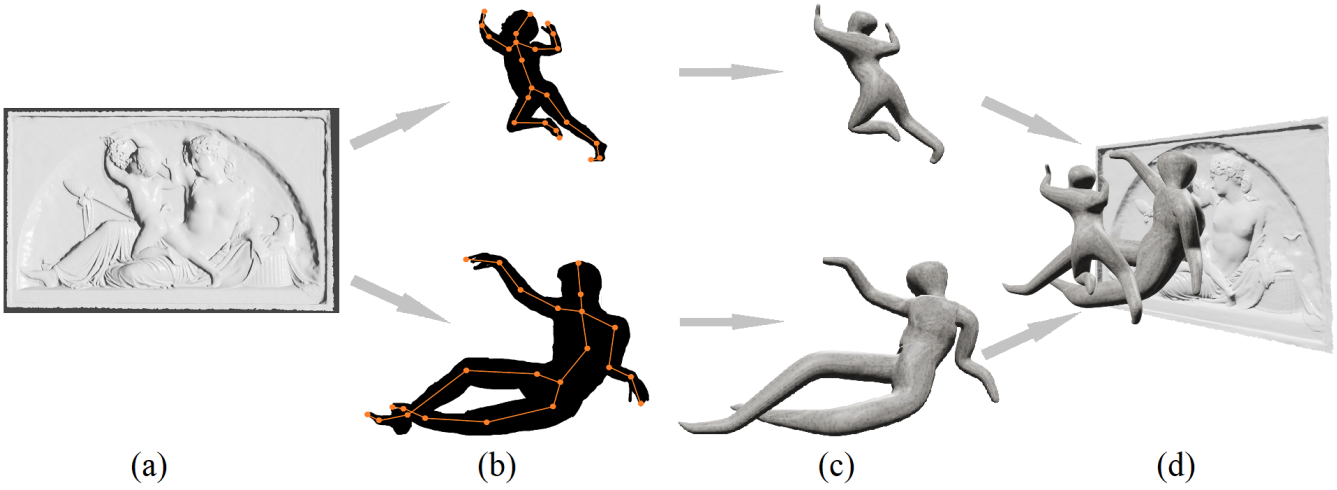


Figure 3: *Overview of our approach. (a) Our system takes as input a digitized bas-relief. (b) The user traces the silhouettes of all human figures in the scene and adds the skeletons. (c) 3D base mesh is computed (d) Final result.*

from existing reliefs superimposed on an underlying surface, such that they be pasted to novel surfaces. But they did not address the problem of recovering the 3D structure of the scenes represented in the reliefs.

The related problem of reconstructing 3D figures from artworks such as drawings or painting has been addressed using geometric methods [ZFL*10, HDK07, BVS16] and machine learning methods [AECOKC17, WCKS19]. Bogo et al. used the SMPL parameterized body model for extracting human figures from images and recovering their 3D structure and proportions [BKL*16].

While this previous work is generally successful for recovering isolated human figures with very impressive and convincing results, it cannot easily handle the more complex cases of overlapping human figures occluding each other in a cluttered background which are typical of narrative bas-reliefs (see Fig. 1 and 16). Thus, in this more difficult case, it seems reasonable to take advantage of the geometry of the bas-reliefs, rather than their 2D image.

In archaeological drawings, reliefs are drawn by hand to aid the archaeologists to visualize the artifacts and compare them without actually holding them in their hands. Previous work has proposed methods for automatically creating such drawings from relief height maps [KST08]. That method was later improved by Zatzarinni et al. [ZTS09] who suggested instead to draw special iso-contours of the height function called prominent curves. In our case, we use hand-drawn archaeological drawings outlining the different figures in the bas-relief separately and uncovering hidden parts.

3. Overview of our method

Our method takes as input a digitized bas-relief in form of a height function, a photograph of the bas-relief on which the user paints shapes (in black on Figure 3 b.) to delimit the contours of the scene elements of interest. We assume the relief can be decomposed into figures, where each figure represents a single scene object. We also ask users to add a predefined 2D skeleton of the silhouettes (in orange on Figure 3 b.). 2D Skeletons are defined as a set of 2D coordinates for joints and pairs of joints for the edges.

While automatic methods have been proposed to compute skeletons of closed contour curves automatically [SP09], we found that these annotations are easy to provide. The user's reliable perception of the pose allows to deal easily with occlusions as well. These annotations enable to use the same prescribed topological structure of the skeleton for all figures, which greatly simplifies the subsequent figure reconstruction. We additionally assume that the photograph is approximately orthographic and taken from an informative viewpoint with little foreshortening.

In this work, we abstract all the details of the human body, such as hair, face, fingers, as we focus on reconstructing the scene in terms of depth and proportion. All we need for this is a rather abstract representation of the human through a type of base-mesh. Taking into account the reconstructions of these so-called details from a bas-relief are left for future work.

Figure 3 illustrates user annotations and main steps of our approach. The first stage of our pipeline estimates a 3D pose of all figures individually. For each figure, using predefined proportions of the skeleton's bones, we apply a combinatorial algorithm to determine all possible 3D skeletons, whose re-projection onto the bas-relief match the user defined 2D skeleton. By integrating the relative depth provided by the digitized bas-relief combined with anatomical constraints on the skeleton joints, we are able to reduce drastically the theoretically exponential number of 3D skeletons to a very small set of plausible solutions. We further propose to solve ambiguities due to partial occlusions present in the bas-relief.

The second step consists in reproducing the staging of the bas-relief. For this, each element of the scene is positioned relative to the others by varying its position relative to the depth of the scene. It should be noted that, in general, the relief map of the bas-relief does not make it possible to establish a total order between all the

elements. Our algorithm then aims to reproduce the relative order of depth given by the bas-relief by minimizing the distance between all the elements while avoiding any collision.

4. Figure reconstruction

Most bas-reliefs, in particular the narrative ones, represent scenes or stories with humans, animals, or human-like figures, which are, technically spoken, articulated objects. Skeleton-based mesh representations are thus most appropriate, since they would even allow "bringing them to life" later on. Without loss of generality let us focus on human figures.

4.1. User input and notations

In our digital system, a template human skeleton, see Figure 5-left, is provided as a graph with $n+1$ *nodes* (joints) $\{J_1,\dots,J_{n+1}\}$ and n *edges* (bones). The user is then asked to place it onto a figure in a displayed image of the bas-relief by moving the nodes in right positions and thus, adapting the length of all 2D skeleton edges to the figure, see a result in Figure 3(b). Let us denote the 2D coordinates of node i in the (x,y) image frame of the bas-relief by $\mathbf{x}_i = (x_i, y_i)$, $i = 1,\dots,23$, its 3D counterpart by $\mathbf{X}_i = (X_i, Y_i, Z_i)$ and the edges by the set of distinct indices (k,l) with $k,l \in \{1,\dots,22\}$, see Figure 5-left. The lengths l_{kl} of the 2D edges are known from the user's input in the image.
The user also paints the silhouette of the human body, see Figure 3(b), by estimating as accurately as possible the partly occluded parts of the body. The silhouette is defined as a binary image S. From the digitized bas-relief, given as a 3D mesh, we compute a height map

$$z = h(x,y), \tag{1}$$

sampled on a regular fine grid. We assume that there exists a monotone increasing function between scene depth and relief depth.

The goal is now to compute a plausible 3D pose matching the 2D skeleton under orthographic view. Without any further requirements, the problem is ill-posed, i.e. an infinite number of solutions exists. We therefore impose the following "plausibility" constraints: the 3D skeleton bones must match human proportions and their pose be anatomically feasible while minimizing the lengthening of the image bones that is observed in the 3D pose.

4.2. Skeleton estimation

The problem of reconstructing a 3D human pose from a single image is a well studied problem, see [BKL*16, RS18] and literature cited herein. Our problem states a bit differently and simpler. Remember that we do want to estimate a skeleton's pose from a bas-relief and not an image. A bas-relief provides precious height information that sophisticated methods do generally not have available. We also do not need to stick to a perspective camera model, which is required when dealing with photographs. For bas-reliefs, a simple orthographic model is sufficient, since most scenes are generally not viewed from an arbitrary viewpoint, but from front. Note further, that the proportions of the characters in bas-reliefs may not be exact for artistic reasons or perception reasons, thus making estimation of a perspective camera model difficult. We also do not want to limit our method to human poses, but want to be able to reconstruct any articulated figure with our method.

2D human pose estimation from a single image is another well studied problem [CHS*18] and it would be interesting to check whether it is applicable or not in this case. It might be used to propose a 2D skeleton instead of letting the user to define the whole skeleton. But it could not be generalized to other figures that are not human.

Our problem is to estimate 3D human pose from the depth and texture images of a bas-relief. This is a very similar problem to 3D human pose from RGBD images provided by kinect sensors. A crucial difference is that the color channel is in grey scales, and the depth channel is highly stylized though bas-relief compression.

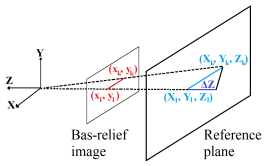


Figure 4: *Orthographic projection of a bas-relief.*

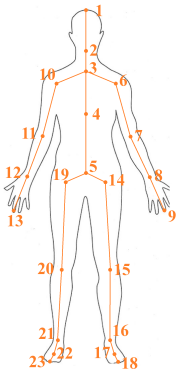
Our algorithm to recover a 3D skeleton for a figure is based on Taylor's combinatorial algorithm [Tay00], because unlike others, it allows to directly integrate the specificities of the bas-relief. Given a 2D skeleton of the figure with n edges, a set of generic relative lengths L_{kl} of the edges in a human figure, and an orthographic scale factor s, the algorithm computes all the 2^n 3D configurations. For all n edges it is indeed possible to choose either end points of the 3D edge to have the smaller Z-coordinate, while projecting on the same image edge.

Let us consider only one edge (k,l) in the bas-relief image plane and its projection onto an image under scaled orthographic projection, the algorithm computes an offset ΔZ_{kl}, which can be applied either to the point $\mathbf{X}_k$ or $\mathbf{X}_l$ (2 choices), depending which one is selected to stick to the reference frame, in order to compute the 3D edge, see 4. Once a scale factor s and length L_{kl} of the 3D edge is given, one obtains

$$\Delta Z_{kl} = \sqrt{L_{kl}^2 - \frac{(x_k - x_l)^2 + (y_k - y_l)^2}{s^2}}. \tag{2}$$

For an entire skeleton, one computes the relative depths for all

1. Head
2. Neck
3. Spine
4. Sternum
5. Pelvis
6. Left Shoulder
7. Left Elbow
8. Left Wrist
9. Left Hand
10. Right Shoulder
11. Right Elbow
12. Right Wrist
13. Right Hand
14. Left Hip
15. Left Knee
16. Left Ankle
17. Left Foot
18. Left Forefoot
19. Right Hip
20. Right Knee
21. Right Ankle
22. Right Foot
23. Right Forefoot



Bone	Relative lengths L_{kl} (in % of the height)
Head (1, 2)	6,6
Neck (2, 3)	4,4
Torso (3, 4)	13,3
Abdomen (4, 5)	16,6
Clavicle (3, 6) (3, 10)	13,3
Arm (6, 7) (10, 11)	13,3
Forearm (7, 8) (11, 12)	13,3
Hand (8, 9) (12, 13)	6,6
Hip (5, 14) (5, 19)	8,8
Thigh (14, 15) (19, 20)	26,6
Calf (15, 16) (20, 21)	23,3
Foot (16, 17) (21, 22)	6,6
Forefoot (17, 18) (22, 23)	3,3

Figure 5: *Human skeleton and relative length L_{kl}.*

edges, chooses a reference point among all nodes and computes the Z-coordinates of all nodes relative to that feature.

Since the size of the figures in the bas-relief is unknown, relative target lengths for the 3D edges are used, whereas the absolute lengths are absorbed by the scale factor s. More details are given in [Tay00]. In our implementation, we use for a human body the relative lengths shown in Figure 5, which are average proportions given in character drawing [ZK14].

The scale factor s, for which a lower bound exists, $s \geq \max_{k,l}\{L_{kl}/\sqrt{(x_k-x_l)^2+(y_k-y_l)^2}\}$, is the only parameter and has to be chosen carefully. Due to the non-linear relation (2) between s and all L_{kl}, there is a trade-off between increasing degree of foreshortening of 3D edges and decreasing fidelity to predefined proportions L_{kl} of the skeleton. In Figure 6 we show the error curves of the characters of the **Hylas and the Water Nymphs** bas-relief, $E(s) = \frac{1}{n}\sum \frac{\|L^*_{kl}-L_{kl}\|}{L^*_{kl}}$ and $D(s) = \sum \Delta Z_{kl}$, where L^*_{kl} are the lengths of the resulting 3D skeletons. These curves are however not comparable, so we can't take the intersection as optimal value. Nevertheless, since the initial 2D edges are only roughly estimated by the user, we favour minimizing foreshortening $E(s)$ and use the minimum allowable s value, i.e. a factor α in (3), which gives acceptable results

$$s = \alpha \cdot \max_{k,l}\{L_{kl}/\sqrt{(x_k-x_l)^2+(y_k-y_l)^2}\} \quad \alpha \in R. \qquad (3)$$

In all examples tested, the observed error is of order 10^{-1} for the minimal value of $\alpha = 1$ and decreases rapidly for $\alpha \geq 2$, see Figure 6. Plausible results are obtained for $1.5 \leq \alpha \leq 2.5$.

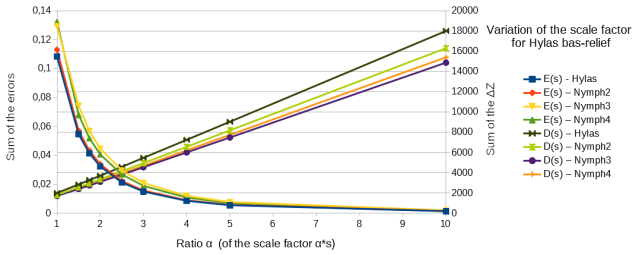


Figure 6: *Evolution of the errors in proportions and depth of reconstructed figure.*

4.3. Choice of 3D configuration

The difficulty of lifting a $2D$ skeleton to $3D$ using Taylor's algorithm is that the orientation of the 22 body parts relative to the viewing plane is left unresolved, which leaves $2^{22} \approx 4million$ possible interpretations. A unique configuration exists, when the orientation of each body part is known. A possibility is to let the user specify which end of each edge in the figure is closer to the observer. This may however be a fastidious task, and in case of occlusions the user would have to make difficult decisions, especially in cases of multiple occlusions. Instead, we choose to propose the most plausible interpretation of the figure given the input annotation, by considering both the geometry of the bas-relief and some anatomic constraints on the recovered joint angles. In those rare cases where the proposed interpretation is incorrect, users can quickly toggle the erroneous body part orientations.

We therefore propose to estimate automatically the orientation by exploiting the height map of the digitized bas-relief, wherever possible and reduce further the set of solutions by imposing anatomical constraints, so that only a few plausible 3D configurations remain. Note, that in case, where all orientations can be determined, a unique solution remains.

In order to estimate the orientation of the edge, i.e. to estimate which joint is closer to the viewer, our insight is to exploit the silhouette together with the bas-relief height map. The latter alone is not sufficient, because we are interested into the orientation of the bone inside the body and not on the skin which is the visible surface on the bas-relief. We therefore clear up the height values with some thickness values, which we derive from the silhouette. For each point on the edge in the image plane, we compute its shortest distance to the silhouette, called Chamfer distance [Bor86], and use it as approximate thickness value. Given a feature curve (our silhouette) in an image, the Chamfer distance transform, well-known in Computer Vision, computes a distance image, where each pixel contains the real-valued distance from this pixel to the nearest boundary point in the silhouette. Using the Chamfer distance to estimate the body parts' thickness is a rough approximation, since it assumes that the body parts are near to be generalized cylinder, which is not the case everywhere. However, since we do not need to recover exact bone position in 3D, but only it's orientation, this assumption is sufficient.

To this end we sample the edge between the joints and perform a least squares fit of their height values h minus the nearest distance d to the silhouette (Chamfer distance). Let us consider a skeleton edge (k,l) defined by its two joints J_k and J_l. Because of the details of the bas-relief (like drapery, wrinkles, and ornaments) partially occluding the edge, a large set of sampling points is required. We define N points along the edge in the (x,y)-frame. N cannot exceed the resolution of the height map h, stored as a gray scale image (1). To each sample point we associate an adjusted height value $z_i = h(x_i,y_i) - d(x_i,y_i)$. With a change of basis, each point is now given by $(x_i,0,z_i)$. We are looking for the equation $z = mx+b$ approximating the function values in a least-squares sense [PTVF07] and derive the edge's orientation from $sign(m)$ (Figure 7a).

As already mentioned above, the edges may be partly occluded, since the least-squares fit may fail along these edges. We therefore accept the proposed orientation for an edge only, if the correlation coefficient is $\in [0.5,1]$, i.e. if it indicates a reliable result. At the end of this step, we have a reliable orientation given for most edges, only a very few number of edges remains undecided. In Figure 7 we illustrate 4 different cases and compare to the naive approach, which computes the orientation of the edge only from the values at the end points.

In the rare cases where the relative depths of joints across a bone remain undetermined, we use anatomical and physical constraints to choose the most plausible solution taking into account the limits of angular movements at the joint and the possible collisions between body parts [LC85, CL03]. We also apply symmetric constraints to preserve the coherency of the body (e.g. the two thighs should have the same length).

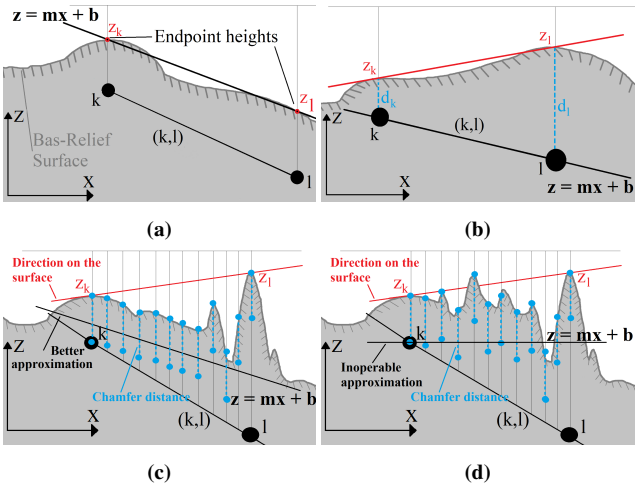


Figure 7: *Application of least square fitting (LSF) to figure out the direction of a bone. (a) In this simple case, comparing the height values at both joints is sufficient for resolving the depth ambiguity. LSF method also gives the right solution. (b) This case can still be handled by the naive method of comparing the height values at both joints, if we take into account the estimated radii (Chamfer distance) of the spherical joints. LSF also finds the right solution in this case. (c) In this more complex case, the naive method fails because at least one of the joints is covered by clothing. LSF is performed on the depth across the entire bone and it finds a correct solution. (d) In the worst case, where the entire bone is covered by clothing, neither our method nor the naive method can be assured to find the right solution.*

Finally, in the case where a bone is almost in the plane of the bas-relief, i.e. $\frac{L^*_{kl}}{L_{kl}} \leq 10^{-2}$, choosing a direction would modify the orientation of the bone of a really small angle. So, it would be hard to figure out if the direction should be backward or forward. We prefer the solution to keep that bone in its original direction, without stretching on the z-axis. As the bone is almost in the plane, its length is already close to the targeted proportion, according to the scale factor chosen.

4.4. Mesh generation

At this point, we have a 3D skeleton with the proper proportions. We now describe how we create a base mesh approximation of the solid shape for each figure. Some methods already propose to generate a 3D base mesh from a skeleton [TGB13, JLW10]. Other methods have approximated a human body mesh, using a generic model [BKL*16, ASK*05], but they can not be generalized to any figure that could be found in a bas-relief, such as animals, trees, and so on.

Our goal being to generate an animatable model, the generated mesh has to be as clean as possible. B-mesh [JLW10] has demonstrated that it is able to generate a quadrangular mesh, with a good edge flow, to improve the deformation during animation. Herein, each joint in the skeleton is approximated by a sphere. Along each edge a set of sampled spheres with a linearly varying radius is computed. The final mesh is obtained from the convex hull of all these spheres. In practice, we use the implementation "skin modifier" provided by Blender [BB11]. The B-mesh structure has the advantage that it can be easily converted into an animation rig, using the sphere radii to generate skinning weights.

For each joint in the skeleton, we estimate the radius of the sphere directly from the the Chamfer distance, described in Section 4.3. We assume that the radius of a bone is evolving linearly between its endpoints. Then, we can measure the radius along the bone, using the Chamfer distance image.

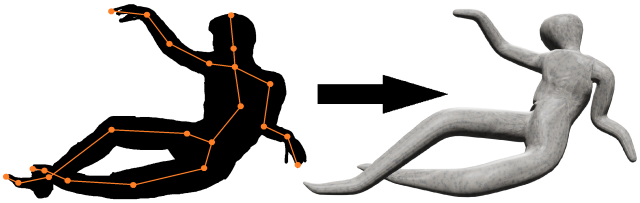


Figure 8: *B-Mesh from a 3D skeleton*

5. Scene reconstruction

After reconstructing each figure separately, we now explain how we choose the remaining free parameters associated with each figure, namely their scale and distance to the bas-relief plane. This can be a tedious process if performed manually. To ease the work of the art historians, we describe a method which computes the distances of all figures relative to the background given their scales, such that the relative distances between the figures is minimal and no two figures intersect. The scale is chosen to keep well-proportioned characters.

5.1. Relative scales

The scale factor is different for each figure of the scene. It figures out the size of a character, how tall it is. Because of the orthographic projection we assume, the figures are fixed on the XY-plane. Then, as the figure is resized only on z-axis, it looks like a stretching on this axis. But increasing the size of each edge of the skeleton also has the benefit of approaching the target proportions. Then, we look for an equilibrium between correct proportions and correct pose. According to Figure 9, we establish that, in our case, the figures look well-proportioned for scale factor included between 1.5 times and 2.0 times its minimal value. Indeed, it neither looks flattened as it seems to be for $\alpha = 1$, nor does it look stretched, such as figures for $\alpha = 2.5$. For a scale factor around $\alpha = 1.75$, we get a proportion error around 3% (cf. Figure 6). For all figures, the factor seems to be in the same range, but as said before, figure scalings are independent, and we wish to estimate an accurate factor based on the morphology of the character.

In the human skeleton we base on, the head is the only bone whose the radius is not impacted by muscular or fat mass. It should be noted that the head bone we talk about consists in the upper of the human head, without the jaw and the chin, and stop at the root of the spine. Then, we can consider that the proportion of the radius does not change a lot from a character to another. It is a

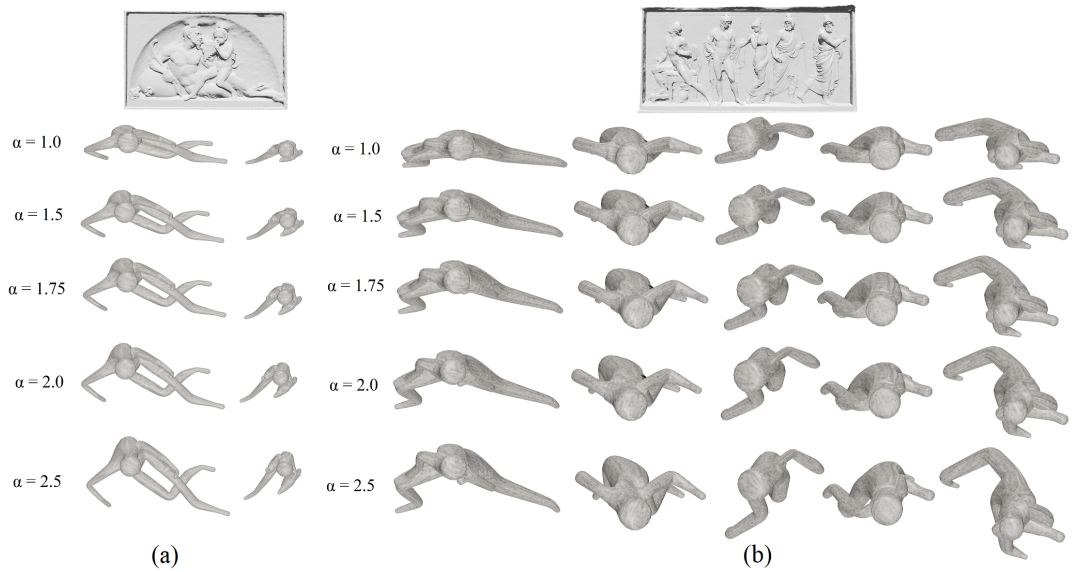


Figure 9: *Effect of the scale factor* α *on the 3D reconstruction of figures. (a)* ***Pan and a satyr*** *bas-relief. (b)* ***Briseis and Achilles*** *bas-relief.*

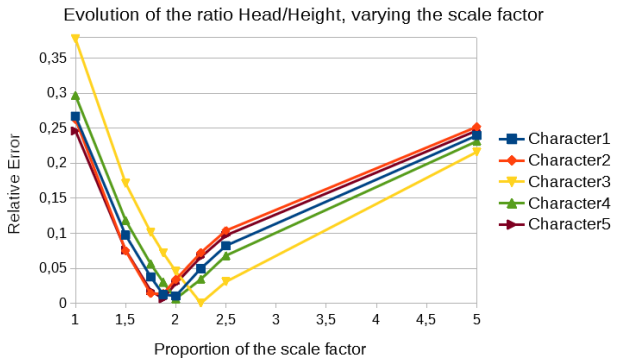


Figure 10: *Evolution of the error of the proportion Head/Height for the characters in* ***Briseis and Achilles*** *bas-relief, while the scale factor is varying.*

good indicator to estimate the proportions of the rest of the body. This proportion can differ a bit according to the specificities of the character (e.g. a baby should have a bigger head, proportionally to the rest of its body).

For a given value of α, we compute the scaled skeleton to be able to estimate the ratio between the size of the head and the height of the character. We consider the height of the character as the longest chain of complete bones: $L_{height} = L_{head} + L_{neck} + L_{torso} + L_{abdomen} + L_{thigh} + L_{calf}$. Finally we compare the computed ratio $\rho_h = \frac{L_{head}}{L_{height}}$ to the ratio ρ_p of the proportions given in Figure 5.

Finally, we minimize the relative error $e_h = \frac{\|\rho_h - \rho_p\|}{\rho_h}$, varying the scale factor α, see Figure 10. It gives us, for each character, the best scale factor to get a well-proportioned character, according to its head size.

5.2. Relative distances

Now that each characters are scaled correctly in the scene, we are able to work on the staging itself. Because of the orthographic projection, figures were already fixed in the XY-plane, locally and globally. But they are still placed at the origin on the z-axis. Indeed, we moved locally, the joints of the characters but we never moved the complete character in the scene, relatively to the others.

We propose to sort the figures on the z-axis according to the information we get from the bas-relief. We cannot use the z position of the figures in the bas-relief, to estimate their distance to each other, as this z measure is compressed, compared to the equivalent 3D scene. But, considering that a bas-relief conserves the depth order (see [WDB*07]), we can figure out if a figure is before or behind another one.

Sorting articulated figures directly is not possible. One joint of a character could be behind another character, while a second joint is before (such as the two characters on the right in Figure 1). Then, we sort all the joint on the z-axis, according to their depth position in the bas-relief (Figure 11). Using the local position of the joint we computed previously, we can estimate an interval of location for each joint of a character, relatively to the other characters. Then, we translate the complete figure, according to the intervals of location of each joints, to recover a depth order similar to the one observed in the bas-relief.

We also consider that each collision or contact in the 3D scene also exists in the bas-relief. Indeed, we suppose that two characters that seem to touch each other in a bas-relief, should do the same in the 3D scene, such as the third and the fourth characters (from the left) are holding their hands in the **Briseis and Achilles** bas-relief (Figure 16).

Working on a scene composed of N characters $P_1 \dots P_N$, we define J^i_j to be the joint j of the character P_i, with $1 \leq j \leq 23$. Given

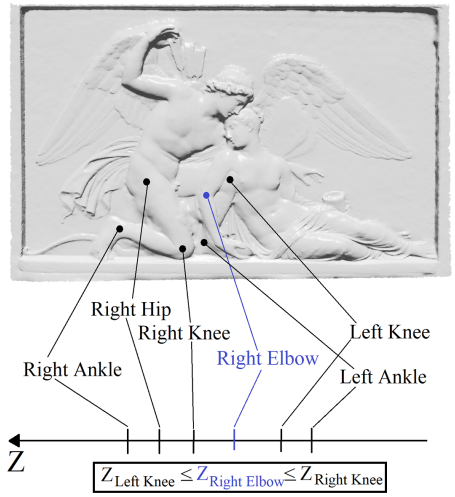


Figure 11: *On **Cupid & Psyche** bas-relief, we estimates the depth position of Psyche's right elbow relatively to the joints of Cupid.*

z^i_j, the z coordinate in the bas-relief system, and Z^i_j the z coordinate in the 3D scene system. Each figure is placed on $Z = 0$ in the 3D scene system. The Chamfer Distance on the location of a joint J^i_j is denoted by $r^i_j = d(x^i_j, y^i_j)$.

In the bas-relief, we compare the position of joints, two at a time, on the z-axis. If one joint J^i_k is placed before another J^j_l, we expect to have the same order in the 3D scene. For a joint J^i_k of the figure P_i, the set of joints J^j_l of the figure P_j that are placed behind in a bas-relief is notated $J^-_{i,k\Rightarrow j}$. By contrast, the set of joints J^j_l placed before is notated $J^+_{i,k\Rightarrow j}$.

Then, we can establish the lower and upper bounds of the interval, where the joint J^i_k should be place on the z-axis:

$$\begin{aligned} Z^{low}_{i,k\Rightarrow j} &= \max_l (Z^j_l + r^j_l), \text{ with } J^j_l \in J^-_{i,k\Rightarrow j} \\ Z^{up}_{i,k\Rightarrow j} &= \min_l (Z^j_l + r^j_l), \text{ with } J^j_l \in J^+_{i,k\Rightarrow j} \end{aligned} \tag{4}$$

So, we have the relationship $Z^{low}_{i,k\Rightarrow j} \leq Z^i_k + r^i_k \leq Z^{up}_{i,k\Rightarrow j}$. Let $\tau_{i=>j}$ be the translation on the z axis of P_i relatively to P_j.

$$\tau_{i=>j} = \bigcap_k \left[Z^i_k + r^i_k - Z^{low}_{i,k\Rightarrow j}, Z^i_k + r^i_k - Z^{up}_{i,k\Rightarrow j} \right] \tag{5}$$

This interval exists only if the proportions of the figures are realistic. So, they have to be similar to what the scene is depicts, and the bas-relief itself has to be realistic. For the first point, existence of interval can be guaranteed by the choice of the scale factor in the previous section. However, the second point cannot be guaranteed.

The global translation of P_i relatively to its original position is:

$$T_i = min \bigcap_j \tau_{i\Rightarrow j} \tag{6}$$

For the specific case where the translation is null, then, no translation was needed, or, there is absolutely no translation that respects all the defined constraints. In the case, we are not able to find a correct translation, we try to move the characters as close as possible, without any collision.

Table 1: *Number of body part orientation errors per character. Basics: direct reading of bas-relief depth at 2D joint locations. LSF: unconstrained least-squares fit over body parts. Constraints: least-squares fit with anatomic constraints.*

Bacchante	Basic	LSF	Constraints
Character1	8	4	2
Character2	4	3	2

Briseis	Basic	LSF	Constraints
Character1	3	1	0
Character2	4	2	2
Character3	7	5	4
Character4	7	4	2
Character5	7	6	5

Childhood	Basic	LSF	Constraints
Character1	6	3	2
Character2	4	2	2
Character3	8	6	3

Cupid_Psyche	Basic	LSF	Constraints
Character1	8	5	3
Character2	7	4	3

Hylas	Basic	LSF	Constraints
Character1	5	3	3
Character2	9	7	5
Character3	7	4	2
Character4	10	7	4

Pan	Basic	LSF	Constraints
Character1	5	3	3
Character2	5	4	3

6. Experimental results

We tested our method on six digital bas-reliefs, created from scans of original artworks. Note, that we do not dispose the scan data, but only the meshes. These bas-reliefs depict scenes with increasing complexity in the number of characters, interactions, and contacts.

Entering the silhouette and the skeletons took less than 2 minutes per figure for a human skeleton of 23 joints. When a character is not facing the camera or has hidden parts, several attempts are needed before a satisfactory annotation could be obtained. This process is highly subjective, and demands a though understanding of the artwork. Ideally, this annotation should be handled by domain experts, i.e. art historians or museum curators.

We obtained realistic results in recovering the 3D poses of the six bas-reliefs after manual correction of minor errors. This was done interactively by switching bone directions. On average, this was needed for 3 bones per character, with a worst case of 5 bones (see Table 1 for complete results). In most cases, those bones where partly or fully overlapped. As a typical example, we manually corrected the direction of the left arm of Cupid, on the left in Figure 13, because his left elbow is behind Psyche. In other cases, errors could be traced back to imprecise user input causing misplaced joints or disproportionate silhouettes.

We also reconstructed the layout of the six scenes, by placing the character in the 3D space, according to their original bas-relief depictions. All the staging were simple enough to be evaluated, from the left to the right, one character after the other.

7. Limitations

Our current implementation is limited to the human body. Other articulated figures could easily be implemented along the same line, by allowing to change the topology of the input 2*D* skeletons and providing the proper proportions and symmetry constraints. Other object categories such as musical instruments, jugs, trees, wheels, togas, scarves and rivers would require other methods.

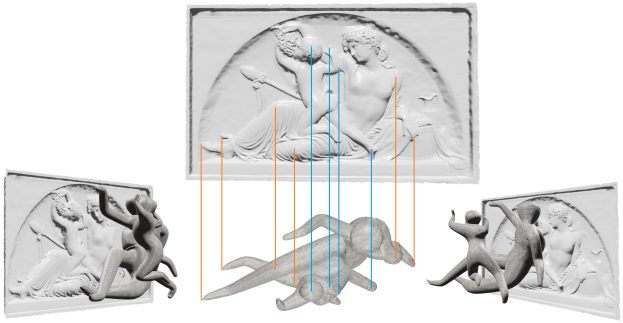


Figure 12: *Bacchante and a Satyr, by Bertel Thorsvalden (1833). The woman playfully holds up a cluster of grapes in front of a small satyr.*

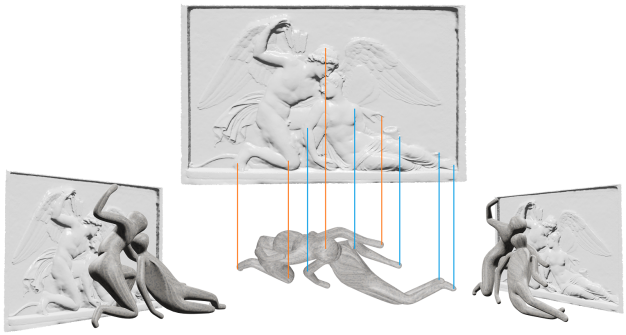


Figure 13: *Cupid revives Psyche, relief in marble by Thorvaldsen (1810) Psyche has poisoned herself with a jar from the underworld, and Cupid brings her to life again.*

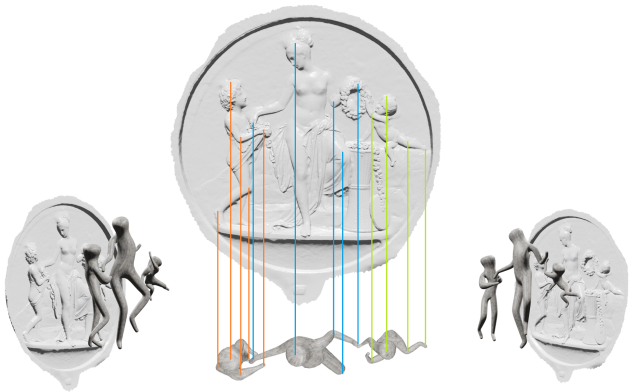


Figure 14: *Childhood or spring, by Bertel Thorvaldsen (1836).*

Also, the mesh generation is based on the Blender Skin Modifier, an implementation of the B-mesh method, which uses spherical joints. For some human joints, like the wrists or the ankles, ellipsoids controlled by two or even three radii would provide a better approximation.

The number of body part orientation errors could be further reduced by flagging the input joints as visible or occluded and by segmenting the bas-relief surfaces into semantic units such as skin and clothing. This would further improve the robustness and precision of our interpretation.

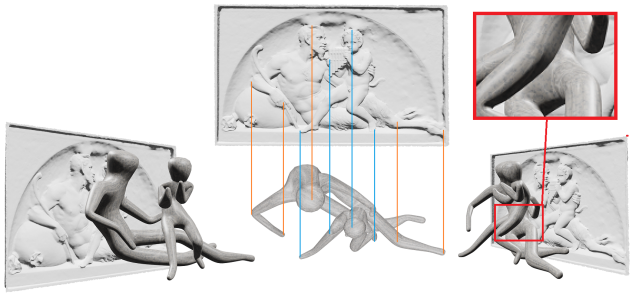


Figure 15: *Pan and a satyr playing flute by Bertel Thorvaldsen (1831).*

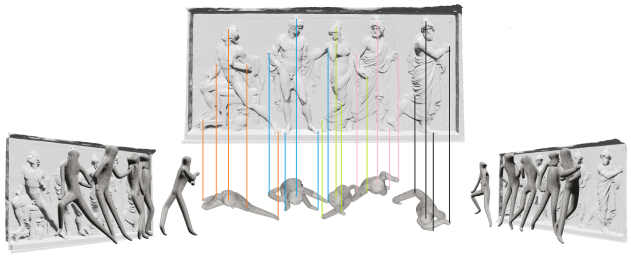


Figure 16: *Briseis and Achilles, marble bas-relief by Bertel Thorvaldsen (1803). Achilles had kidnapped Briseis during a raid in the region of Troy, but was forced by the gods to hand her over to Agammemnon. His wrath or pain is expressed by his tense posture, unusual for Thorvaldsen.*

Finally, we can not guarantee to maintain the proper contact relations between figures. Our method approximates the silhouette of the characters through a base mesh reconstruction, without taking care of all the details (like the draperies, the muscles, etc ...). Even if the body proportions are well-estimated, the base mesh can not keep the exact shape of the body away from the joint locations. Figure 15 shows an example that precisely describes this problem. The small satyr is supposed to be seated on the thigh of Pan, but there already is a collision between its leg, and we can not reproduce this contact point precisely.

8. Conclusion and future work

We have introduced the new problem of reconstructing 3D scenes from bas-relief depictions of those scenes. We have presented a method that takes as input the height field of a digital bas-relief annotated with silhouettes and 2*D* skeletons of human figures and generates plausible 3D scenes, that can then easily be edited and refined to correct mistakes.

Our work opens several new avenues of research. First, we would like to transfer and scale the fine details of the bas-relief to the base meshes in the 3D scenes to obtain a more realistic view of the 3D

scene. We also would like to apply inpainting to hidden parts of the foreground figures and backgrounds. The resulting 3D models could then be animated to reveal the stories suggested by the bas-relief. Automatic annotation of all figures in a bas-relief, with their silhouettes and 2D skeletons, is another topic for future research.

8.1. Acknowledgements

Bas-reliefs used in this paper were scanned and digitized from the originals in the Thorvaldsen Museum (Copenhagen, Denmark) by Geoffrey Marchal using Memento beta (now ReMake) from AutoDesk and obtained from Sketchfab under a Creative Commons licence Attribution 4.0 International (CC BY 4.0). This work was partially funded by the ANR e-ROMA (ANR-16-CE38-0009) from Lugdunum - Gallo-Roman Museum of Lyon, France.

References

[AECOKC17] AVERBUCH-ELOR H., COHEN-OR D., KOPF J., COHEN M. F.: Bringing portraits to life. ACM Trans. Graph. 36, 6 (2017). 3

[ASK*05] ANGUELOV D., SRINIVASAN P., KOLLER D., THRUN S., RODGERS J., DAVIS J.: Scape: Shape completion and animation of people. ACM Trans. Graph. 24, 3 (July 2005), 408–416. 6

[BB11] BLENDER FOUNDATION, BISHOP N.: Blender, skin modifier, march 2011. 6

[Ben16] BENDAZZI G.: Animation: A World History: Volume I: Foundations - The Golden Age. CRC Press, 2016. 2

[BKL*16] BOGO F., KANAZAWA A., LASSNER C., GEHLER P., ROMERO J., BLACK M. J.: Keep it smpl: Automatic estimation of 3d human pose and shape from a single image. In European Conference on Computer Vision (2016), Springer, pp. 561–578. 3, 4, 6

[Bor86] BORGEFORS G.: Distance transformation in digital images. Computer Vision, Graphics, and Image Processing 34, 3 (1986), 344 – 371. 5

[BVS16] BESSMELTSEV M., VINING N., SHEFFER A.: Gesture3d: Posing 3d characters via gesture drawings. ACM Trans. Graph. 35, 6 (Nov. 2016). 3

[CCL*11] CHEN Y., CHENG Z.-Q., LI J., MARTIN R. R., WANG Y.-Z.: Relief extraction and editing. Comput. Aided Des. 43, 12 (Dec. 2011). 2

[CHS*18] CAO Z., HIDALGO G., SIMON T., WEI S., SHEIKH Y.: Openpose: Realtime multi-person 2d pose estimation using part affinity fields. CoRR abs/1812.08008 (2018). 4

[CL03] COHEN I., LI H.: Inference of human postures by classification of 3d human body shape. In Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (2003), AMFG '03. 5

[CMS97] CIGNONI P., MONTANI C., SCOPIGNO R.: Computer-assisted generation of bas-and high-reliefs. Journal of Graphics Tools 2, 3 (1997), 15–28. 2

[Con96] CONNELLY J. B.: Parthenon and parthenoi: A mythological interpretation of the parthenon frieze. American Journal of Archaeology 100, 1 (January 1996), 53–80. 2

[HDK07] HORNUNG A., DEKKERS E., KOBBELT L.: Character animation from 2d pictures and 3d motion data. ACM Trans. Graph. 26, 1 (Jan. 2007). 3

[JLW10] JI Z., LIU L., WANG Y.: B-mesh: A modeling system for base meshes of 3d articulated shapes. Computer Graphics Forum 29, 7 (2010), 2169–2177. 6

[JMS13] JI Z., MA W., SUN X.: Bas-relief modeling from normal images with intuitive styles. IEEE transactions on visualization and computer graphics 20, 5 (2013), 675–685. 2

[Joh04] JOHNSON R. B.: The Parthenon Code: Mankind's History In Marble. Solving Light Books, 2004. 2

[KST08] KOLOMENKIN M., SHIMSHONI I., TAL A.: Demarcating curves for shape illustration. ACM Trans. Graph. 27, 5 (Dec. 2008). 3

[KWC*12] KERBER J., WANG M., CHANG J., ZHANG J. J., BELYAEV A., SEIDEL H.-P.: Computer assisted relief generation - a survey. Comput. Graph. Forum 31, 8 (Dec. 2012). 2

[LC85] LEE H.-J., CHEN Z.: Determination of 3d human body postures from a single view. Computer Vision, Graphics, and Image Processing 30, 2 (1985), 148 – 168. 5

[MSF*18] MIAO Y., SUN Y., FANG X., CHEN J., ZHANG X., PAJAROLA R.: Relief generation from 3d scenes guided by geometric texture richness. Computational Visual Media 4, 3 (2018), 209–221. 2

[Osa16] OSADA T.: The Parthenon Frieze. The Ritual Communication between the Goddess and the Polis. Phoibos Verlag,Vienna, 2016. 2

[PTVF07] PRESS W. H., TEUKOLSKY S. A., VETTERLING W. T., FLANNERY B. P.: Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, 2007. 5

[Ros17] ROSSI A.: From the bas-relief to the 3d model: a reconstruction hypothesis for an 8th century bc civil fortification. DISEGNARECON 10, 19 (2017), 12–1. 2

[RS18] ROGEZ G., SCHMID C.: Image-based Synthesis for Deep 3D Human Pose Estimation. International Journal of Computer Vision 126, 9 (2018), 993–1008. 4

[Sif18] SIFIANOS G.: Phidias the animator. movement analysis in the parthenon frieze. In Animafest Zagreb (2018). 2

[SLL07] S.-L. LIU R.R. MARTIN F. L. P. R.: Background surface estimation for reverse engineering of reliefs. International Journal of CAD/CAM 7, 4 (2007). 2

[SP09] SIDDIQI K., PIZER S.: Medial Representations. Mathematics, Algorithms and Applications. Springer, 2009. 3

[Tay00] TAYLOR C. J.: Reconstruction of articulated objects from point correspondences in a single uncalibrated image. Computer Vision and Image Understanding 80 (2000), 349–363. 4, 5

[TGB13] THIERY J.-M., GUY E., BOUBEKEUR T.: Sphere-meshes: Shape approximation using spherical quadric error metrics. ACM Trans. Graph. 32, 6 (Nov. 2013), 178:1–178:12. 6

[WCKS19] WENG C.-Y., CURLESS B., KEMELMACHER-SHLIZERMAN I.: Photo wake-up: 3d character animation from a single photo. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019). 3

[WDB*07] WEYRICH T., DENG J., BARNES C., RUSINKIEWICZ S., FINKELSTEIN A.: Digital bas-relief from 3d scenes. In ACM SIGGRAPH 2007 Papers (2007), SIGGRAPH '07. 2, 7

[WMR*13] WU J., MARTIN R., ROSIN P., SUN X.-F., LANGBEIN F., LAI Y.-K., MARSHALL A., LIU Y.-H.: Making bas-reliefs from photographs of human faces. Computer-Aided Design 45, 3 (2013), 671 – 682. 2

[WTP*18] WEI M., TIAN Y., PANG W.-M., WANG C. C., PANG M.-Y., WANG J., QIN J., HENG P.-A.: Bas-relief modeling from normal layers. IEEE transactions on visualization and computer graphics 25, 4 (2018), 1651–1665. 2

[ZFL*10] ZHOU S., FU H., LIU L., COHEN-OR D., HAN X.: Parametric reshaping of human bodies in images. In ACM SIGGRAPH 2010 Papers (2010), SIGGRAPH '10. 3

[ZK14] ZARINS U., KONDRATS S.: Anatomy for sculptors: understanding the human figure. Exonicus, LLC, 2014. 5

[ZTS09] ZATZARINNI R., TAL A., SHAMIR A.: Relief analysis and extraction. ACM Trans. Graph. 28, 5 (Dec. 2009). 2, 3

[ZZZY13] ZHANG Y.-W., ZHOU Y.-Q., ZHAO X.-F., YU G.: Real-time bas-relief generation from a 3d mesh. Graphical Models 75, 1 (2013), 2–9. 2