Making Bas-reliefs from Photographs of Human Faces

J. Wu, R.R. Martin, P.L. Rosin, X.-F. Sun, F.C. Langbein, Y.-K. Lai,
A.D. Marshall[a], Y.-H. Liu[b]

[a] *School of Computer Science & Informatics, Cardiff University, Cardiff CF24 3AA, UK*
[b] *Department of Computer Science, Aberystwyth University, Aberystwyth SY23 3DB, UK*

Abstract

Bas-reliefs are a form of flattened artwork, part-way between 3D sculpture and 2D painting. Recent research has considered automatic bas-relief generation from 3D scenes. However, little work has addressed the generation of bas-reliefs from 2D images. In this paper, we propose a method to automatically generate bas-relief surfaces from frontal photographs of human faces, with potential applications to e.g. coinage and commemorative medals.

Our method has two steps. Starting from a photograph of a human face, we first generate a plausible *image* of a bas-relief of the same face. Secondly, we apply shape-from-shading to this generated bas-relief image to determine the 3D shape of the final bas-relief. To model the mapping from an input photograph to the image of a corresponding bas-relief, we use a feedforward network. The training data comprises images generated from an input 3D model of a face, and images generated from a corresponding bas-relief; the latter is produced by an existing 3D model-to-bas-relief algorithm. A saliency map of the face controls both model building, and bas-relief generation.

Our experimental results demonstrate that the generated bas-relief surfaces are smooth and plausible, with correct global geometric nature, the latter giving them a stable appearance under changes of viewing direction and illumination.

Keywords:
Bas-relief, photograph, feedforward network, image relighting, shape from shading

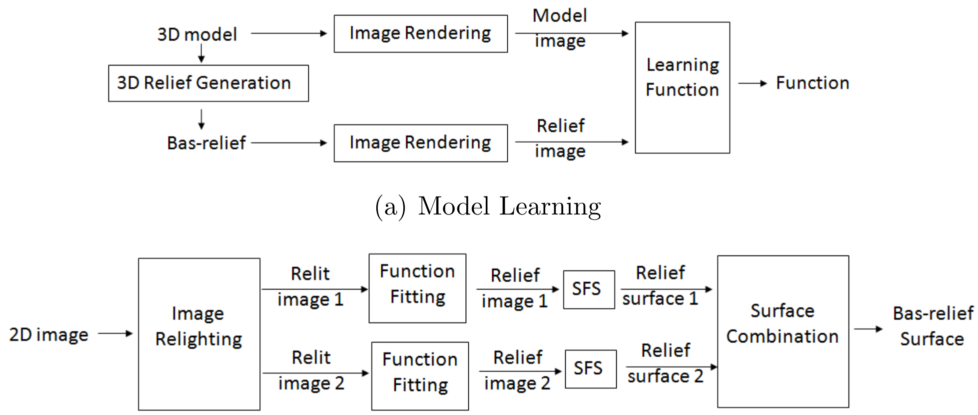
Email addresses: `{J.Wu, ralph, paul, Xianfang.Sun, F.C.Langbein, Yukun.Lai, dave}@cs.cardiff.ac.uk` (J. Wu, R.R. Martin, P.L. Rosin, X.-F. Sun, F.C. Langbein, Y.-K. Lai, A.D. Marshall), `yyl@aber.ac.uk` (Y.-H. Liu)

1. Introduction

Bas-reliefs are a form of flattened sculpture applied to a base surface. Compared to high-reliefs, bas-reliefs have a limited height above the background, and no part is undercut. They can be considered to be part way between sculpture and painting. Bas-reliefs have been used for centuries in art and architectural decoration, for example as portraits on coins. In modern times, they are also popular in industrial design, for example for branding packaging. However, the production of bas-reliefs requires considerable artistic skill and manual effort. In the fields of computer aided design and computer graphics, recent research [1, 2, 3, 4, 5, 6] has considered automatic bas-relief generation from *3D scenes*. However, as such methods are based on 3D input data, this restricts their range of application, as the necessary 3D input models require specialised and expensive equipment for capture, or must be created laboriously by hand. An alternative approach, with potentially much wider application, is to generate bas-reliefs from *2D images*. However, little work has addressed this problem [7, 8].

Here, we consider a specific problem: the production of a bas-relief from a single frontal photograph of a human face. We focus on human faces, since the face is of special interest in bas-reliefs, especially for coinage and commemorative medals. We mainly address frontal faces here as they are somewhat simpler to process, even though applications often also use profile or semi-profile views. Frontal faces have fixed head pose, and eliminate the necessity of head pose estimation for face images with semi-profile views. Moreover, many frontal face databases exist, facilitating experiments, for example on image relighting. Nevertheless, as we do not use any specific attributes of frontal faces (such as symmetry), our method can in principle be extended to other views. Indeed, our experiments, demonstrate an example using a non-frontal face too.

Our approach is based on shape-from-shading (SFS) [9, 10, 11], a standard technique to recover 3D shape from a single image of an object, based on a model of variation of reflected intensities as a function of surface orientation. However, generating a bas-relief surface from a human face image is not straightforward. One approach would be to use SFS to directly recover the 3D shape of the face as a depth map, and then process that with one of the existing bas-relief production algorithms given above. We do not take



(a) Model Learning

(b) Relief Generation

Figure 1: The proposed framework

this approach because the results would be dependent on any deficiencies in the chosen bas-relief production algorithm. Instead, we take an alternative path: we first generate a new image from the face photograph; this new image corresponds to the expected appearance of the bas-relief. We then apply SFS to this image to recover the shape of the bas-relief. This potentially allows us to base our approach on high-quality hand-crafted bas-reliefs, rather than algorithmically generated ones, as we now discuss.

Our overall framework has two components, shown in Figure 1. First, an offline process is used to learn the relationship between an image of a 3D human face and an image of a corresponding 3D bas-relief of that face. This is done by taking one or more 3D face models, and processing them using any existing bas-relief generation algorithm to produce corresponding 3D bas-reliefs. Each original 3D model and corresponding bas-relief are then rendered to give 2D images, using one or more lighting conditions. A learning algorithm is used to model the relationship between the pixel values in these images. While here we use an existing 3D bas-relief generation algorithm for simplicity, an alternative would be to learn the relationship using photographs of human faces and handcrafted bas-reliefs of those faces derived from those photographs. This would avoid any deficiencies in existing bas-relief generation algorithms (but would also necessitate careful registration of reliefs and photographs).

Once we have learnt the model between 2D face images, and 2D face bas-relief images, we can input a new face image, and apply the model to determine what a corresponding bas-relief model should look like. We then apply SFS to recover the bas-relief surface from the generated bas-relief image. In practice, we find that if we re-light the input image from several new directions [12], giving multiple versions of the input image, and use each to determine a bas-relief, these can be combined into a more satisfactory final bas-relief.

In the following, Section 2 reviews related work on bas-relief generation and shape from shading. Sections 3, 4, and 5 give detailed descriptions of the model building step, bas-relief image generation, and shape from shading. Section 6 describes how multiple renderings may be combined to give a final bas-relief surface. Section 7 presents examples, while Section 8 considers several alternative strategies in our methods. Section 9 gives conclusions and discusses possible improvements.

2. Related Work

The earliest attempt to generate bas-reliefs by computer was given in [1]. The authors summarized various basic attributes of artistic bas-reliefs, in particular noting that more distant objects undergo greater depth compression than nearer ones. Based on this finding, the authors applied a standard perspective transformation to the height fields of a 3D scene. Although the results generally adhered to the principles of creating bas-relief, the results only weakly preserved detailed features.

More recent work [2, 4, 3] was inspired by techniques used in high dynamic range (HDR) imaging, where a wide range of intensities is compressed to use a lower intensity range in a way that retains important visual features. In relief processing, depths replace intensities. The method in [4] performs depth compression in the gradient domain, using a non-linear scaling [13] of gradient magnitudes; the aim is to preserve small gradients while attenuating large ones. The approaches in [2] and [3] both make use of unsharp masking to emphasize salient features, before using linear scaling for compression. The former works in differential coordinates, while the latter works in the gradient domain. The results in [3] were improved in [14] by replacing linear scaling with non-linear scaling techniques during compression. Further work of a similar kind [6] also applies non-linear scaling, but uses bilateral filtering to decompose the gradient into coarse and fine components, enabling careful manipulation of detail.

A different kind of approach is based on the concept of adaptive histogram equalization from image processing [5]; depth compression works directly on the height field. The authors demonstrate good results for various scenes and objects, including human faces, and we use it as a basis for our learning process.

The above methods start with a *depth-map* of a 3D scene, and selectively compress depths to create the bas-relief surface. Two recent papers [7, 8] use *images* as input. A two-level (low frequency component and high frequency detail) approach is given in [8] to restore brick and stone reliefs from images taken as rubbings. The authors have also applied their approach to photographs, but, as they note, it is only suitable for objects made of homogeneous materials with relatively little texture and low albedo. An experiment on a photograph of Picasso showed that the approach provided poor results for portrait photographs.

More pertinent to our work is [7], which aims to create relief surfaces that approximate desired images under known directional lighting. The authors first adjust the input images to match their average radiance to that of a relief plane. They then apply a modified SFS method with height constraints to this adjusted image to create the relief surface. The authors note that the integrability constraint enforced by SFS constrains the radiance for each element of a recovered surface. To use this observation, they associate each pixel with not just one, but several, surface elements. Unfortunately, the increased numbers of degrees of freedom also increases the sensitivity of the generated bas-relief surfaces to changes in viewing direction and illumination.

An important observation that we have made is that images of real bas-reliefs, such as heads on coins, *do not* approximate images of the corresponding 3D objects (photographs of heads). Instead, they *enhance* the salient features. Thus, we do not follow the aims of [7], but instead try to make bas-relief surfaces with the same appearance as bas-reliefs created by an artist. Trying to approximate an original photograph is an unrealistic goal given that the bas-relief surface must be relatively flat. This different emphasis of approach has a further advantage that the results are not strongly view dependent, and the global geometric nature of each generated bas-relief surface is consistent with human perception, giving them a stable appearance under changes of viewing direction and illumination.

Our work employs existing SFS techniques, which recover shape from intensity variation in an image. A survey of early SFS work can be found in [9]. Assuming Lambertian reflectance and a known directional light source, Horn and Brooks [15] gave a variational approach to solve the SFS problem. The energy to be minimised comprises a brightness constraint and a quadratic regularizing term enforcing surface smoothness. However, this method involves the choice of a Lagrange multiplier, and the results tend to be oversmoothed. To overcome these deficiencies, Worthington and Hancock [10] proposed a geometric SFS framework which strictly satisfies the brightness constraint at every pixel: surface normals are forced to lie on their irradiance cones during each iterative update. The same authors have also given several robust regularizers with better smoothing behaviour than the quadratic one [16]. Huang and Smith [11] gave a structure-preserving regularization constraint, which allows smoothing to be performed locally, dependent on the intensities in a local area. We adopt the last method, as it is particularly suited to our requirement to preserve salient facial features.

3. Mapping face images to face bas-relief images

As shown in Figure 1, the first step of our framework is to learn the relationship between a 2D frontal image (photograph) of a human face and a 2D image of a corresponding bas-relief of the same face. The idea is that if we know the mapping, we can generate bas-relief images from *new* input face images without requiring corresponding 3D models.

Initially, we tried an alternative approach (with similar goals to [7]): to use the 2D frontal image as a basis for *directly* producing a relief using shape-from-shading, with extra constraints to enforce the result to have very low height: the aim was to produce a relief which *looks as similar as possible* to the input face. It soon became obvious that this does not give satisfactory results. On analysing images of artistic bas-reliefs, while they are recognisably related to images of the original object, they are also quite clearly different from them. Figure 2 shows an example of a bas-relief generated using an existing 3D bas-relief generation method [5], clearly demonstrating this point.

We thus turned to understanding and modeling the mapping between intensities in images of faces and images of corresponding bas-reliefs. It soon became clear that a simple function is not adequate for this purpose. Some explicit image processing methods, such as image embossing, can produce an image with a bas-relief-like effect. However, these methods usually change the reflectance properties of the surface, and the lighting conditions in the original image, which increases the difficulty of applying shape-from-shading

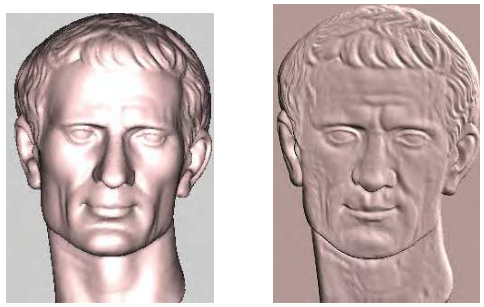


Figure 2: Two images rendered under the same conditions: a 3D model, and a bas-relief generated from it using the method in [5]. Note that these images are very different.

in the subsequent steps of our process. Instead, then, we take a different strategy, and learn the mapping by training a feedforward network.

For training, computer generated 2D frontal images of a 3D face model and a corresponding 3D bas-relief model are produced, using the same rendering setup—the same reflectance model and lighting conditions. We make use of this consistency of rendering during the shape from shading step. We take the 3D face models as given; during the learning process, to generate corresponding bas-reliefs, we use an existing algorithm chosen for its good performance on faces [5]. (As noted, better results are likely to be obtained using high-quality bas-relief models produced by a sculptor.) We also use a saliency map to guide the selection of the training data, so that the more salient areas are more likely to be selected during training (and hence better modelled). We now give further details.

3.1. Generating Bas-reliefs for Training

To learn the mapping from images of faces to images of bas-reliefs of faces, we need corresponding pairs of images. Given one or more 3D face models, we need to generate corresponding 3D bas reliefs. We do so using Sun's method [5], which we briefly summarise. Starting from a height map of the face (i.e. a range image), it performs histogram equalization of heights within a local neighborhood for each point. Two modifications are applied to this local histogram equalization. First, the calculation of the histogram is weighted by the gradient magnitude after applying a non-linear transformation, in order to preserve small shape details. The second modification applies an iterative clipping and redistribution procedure to the local histograms, limiting their content. This prevents too many counts in any one histogram bin, which would result in shape distortion and increased noise. A scaling factor l controls this limit for each bin's content. To generate the final bas-relief surface, the method processes the input height maps using several different neighborhood sizes, and averages the results. Figure 2 shows a scanned head of Julius Caesar and the final bas-relief produced using the method.

3.2. Saliency Map Calculation

When producing a bas-relief, it is more important to preserve details in some areas of the face than others. We define and use a saliency map for this purpose. It is used to guide the learning process so that more salient areas are more likely to be selected during training. It is also used again later in the shape-from-shading process in order to preserve salient facial features (see Section 5).

The saliency map is computed from the input image; during training we also determine saliency maps for the training images. Photographs of faces often contain noise, partly due to data acquisition errors, but also both because of skin blemishes—small local changes in skin colour not due to a change in surface shape. Images of faces generated from 3D mesh models may also contain systematic noise due to low mesh resolution. Thus, before calculating the saliency map, we use bilateral filtering [17] to smooth the image while still preserving the shapes of features.

From this bilaterally-filtered image I, we calculate the image gradient magnitude:

$$g(x, y) = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}. \tag{1}$$

Next, we apply histogram equalization to g to enhance contrast. The same clipping and redistributing procedure described in [5] is also applied to this histogram, again using the scaling factor l to control the level of detail retained—retaining too much detail also retains noise. A final, smoothed, saliency map is found by applying an averaging filter with a circular neighbourhood to the result.

Examples of saliency maps calculated from images rendered using mesh models, and from photographs, are shown in Figure 3; they have resolutions of 596×852 and 701×841 respectively. We use 256 equal-sized bins during histogram equalization, and a radius of 3 for the circular averaging filter. Results are shown in Figure 3 for varying scaling factors l; the saliency maps

(a) saliency maps of a image generated from Julius Caesar model

(b) saliency maps of a real-world image

Figure 3: Examples of saliency maps. Left to right: original images, and saliency maps with $l = 1, 4, 8, 16, 32$ respectively.

bring out more detail with increasing l. A reasonable balance between feature details and noise occurs when $l = 8$.

3.3. Feedforward Network Training

Given a 3D face model and a corresponding (algorithmically generated) bas-relief surface, we now compute an image of each in the same position, using the same lighting conditions and reflectance models. We assume that the intensity of each pixel in the bas-relief image is determined by the intensities in a local neighborhood around the same pixel in the corresponding 3D model image. To learn the relationship between these local neighborhoods and the bas-relief pixel values, we use a feedforward neural network [18] for its simplicity. Other neural networks or learning algorithms could also be used.

In our experiments, we used a 3D model of Julius Caesar and a corresponding generated bas-relief (as shown in section 3.1) to generate the training model images and bas-relief images. We generated two pairs of corresponding training images using Lambertian reflectance and parallel lighting, from lighting directions, $(1, 1, 1)$ and $(-1, 1, 1)$, respectively (with z towards the model), as shown in Figure 4. For each pair of training images, our feedforward network has one hidden layer with 30 neurons. Each network is

Figure 4: Model images and corresponding bas-relief images used for training. Left pair: light direction $(1, 1, 1)$, right pair: light direction $(-1, 1, 1)$.

trained for up to 1000 epochs and to a mean-square error goal of 0.001. Once the error goal is reached, a cross-validation technique is used to determine the performance and decide whether to stop training.

4. Generating Bas-relief Images

Having learnt a mapping from a face image to a bas-relief image, we can apply it to new images of faces to generate corresponding bas-relief images. However, the images used for training are illuminated under specific lighting conditions. Given a new image, for the learnt mapping to be applicable, it should be illuminated from the same lighting direction as the training images. Various methods exist in the literature which take an image under one set of illumination conditions, and re-light it to produce a corresponding image under different illumination conditions. We make use of the quotient image technique [12] for this purpose.

4.1. Image Relighting

Three images of the same object under linearly independent light sources are sufficient to generate the image space resulting from varying lighting directions [19, 20]. The basic idea of the quotient image technique is to apply the image space generated from one object to other objects of the same kind. The key is to find the quotient image, which is defined as the quotient between the objects' albedos. The quotient image is independent of illumination, and once it has been determined, the whole image space of the new object can be generated from three images of the base object. In [12], the authors show how to obtain the quotient image Q_y given an image y_s of

(a) Azimuth angle: $-10°$, elevation angle: $-20°$

(b) Azimuth angle: $-35°$, elevation angle: $+15°$

(c) Azimuth angle: $+35°$, elevation angle: $+15°$

Figure 5: Bootstrap set for image relighting.

object y under a certain light source s, based on a bootstrap set of training objects $A_1, \ldots, A_N$. Each A_i is a matrix whose columns are the three images of a base object a_i. The use of a bootstrap set instead of a single object allows for variation of albedos. The albedos of the N training objects are expected to span the albedo of the novel object. Increasing N in principle gives more freedom to represent novel objects, although experiments in [12] show little difference as N varies from 2 to 10.

In our experiments, we used a bootstrap set of images of 8 faces from Yale Face Database B [21]. The three images of each face are all frontal, being illuminated from three lighting directions with azimuth and elevation angles of $(-10°, -20°)$, $(-35°, +15°)$, and $(+35°, +15°)$ respectively. The images are coarsely aligned using the tip of the nose and the centers of the eyes. The aligned bootstrap set is shown in Figure 5.

Figure 6 shows examples of applying image relighting using this training data. Two images of the same person are shown under different lighting. Apart from shadows, the quotient images are quite similar, and approximately invariant to changes in light source as hoped. The quotient image technique unfortunately cannot take shadows into account. Relighting images without shadows produces results with a realistic appearance (top row,

Figure 6: Image relighting results, for 2 images of the same person taken under different lighting. Left to right: original image, quotient image, and images relit from directions $(1, 1, 1)$ and $(-1, 1, 1)$.

Figure 6). Due to the simple coarse alignment used, some minor artifacts can be seen in the relit images around the eyes and hair. This could be improved by applying a more sophisticated pointwise alignment method. We return to the problem of shadows later.

4.2. Generating the Bas-relief Images

We are now ready to generate the bas-relief image from the input face image. We first relight it from each of the same lighting directions as the training images, using the quotient image technique. Next, the original image and relit images are scaled, according to the distance between the eyes, to be a similar size to the training images. A saliency map is then calculated from the resized original image, for use later. Next, we apply the learnt feedforward networks to the relit images, to get the pixel values in the bas-relief images from pixel neighborhoods in the relit images.

Examples of generated bas-relief images are shown in Figure 7(The intensity of the relief images are linearly stretched for showing purpose.). Salient facial features are preserved in the generated images, giving these images recognizable bas-relief appearance. The lighting directions used in the relit model images are also evident in the bas-relief images, and are utilized directly in the following shape-from-shading step.

Figure 7: A generated bas-relief image. Left to right: original image, two relit images, and corresponding bas-relief images.

5. Finding the Relief using Shape-from-shading

We now apply shape-from-shading (SFS) to each constructed relief image, to determine the geometry of the relief surface. SFS recovers shape from variation of intensities in the image. Most popular SFS methods solve the problem by minimizing an energy function, which usually includes an intensity constraint (that the surface orientation should lead to the observed intensity) and a regularizing term (enforcing surface smoothness). A basic energy function for Lambertian surfaces is given in [15]:

$$I = \int\int \underbrace{(E(x,y) - \mathbf{n}(x,y)\cdot\mathbf{s})^2}_{\text{Brightness Error}} + \underbrace{\lambda\left(\left|\frac{\partial\mathbf{n}(x,y)}{\partial x}\right|^2 + \left|\frac{\partial\mathbf{n}(x,y)}{\partial y}\right|^2\right)}_{\text{Regularizing Term}} dxdy, \tag{2}$$

where $E(x,y)$ and $\mathbf{n}(x,y)$ are respectively the image intensity and the surface normal at pixel location (x,y), $\mathbf{s}$ is the direction of the light source, and λ balances intensity fidelity against surface smoothness. In practice, surfaces recovered using this formulation are often over-smoothed.

Our SFS method improves upon this formulation in two ways. First, we satisfy intensity closeness as a hard constraint using the method of Worthington and Hancock [10]. The aim is to preserve the appearance of the image, which is important in our application. Secondly, we use a modified version of Huang and Smith's [11] structure-preserving regularization constraint, which helps to preserve salient facial features. Our SFS method is iterative. In each iteration, the surface normals are updated to first satisfy the regularizing term, and secondly to satisfy the brightness constraint. Finally, we use the algorithm of Frankot and Chellappa [22] to integrate the field of recovered surface normals to generate the bas-relief surface. We now give further details.

5.1. Brightness Constraint

For Lambertian surfaces, satisfying the intensity closeness as a hard constraint is equivalent [10] to enforcing

$$\int\int (E(x,y) - \mathbf{n}(x,y) \cdot \mathbf{s})^2 dxdy = 0. \tag{3}$$

This causes the surface normal at pixel (x, y) to lie on a cone whose axis is in the light source direction $\mathbf{s}$ and whose opening angle is $\alpha = \cos^{-1} E(x, y)$. During each iteration of SFS, after updating the surface normals according to the regularizing term, the updated surface normals usually do not lie on the cone. Then, we need to rotate them back to their closest on-cone positions to enforce the brightness constraint.

5.2. Regularization Constraint

Enforcing the regularizing constraint in Equation (2) during each iteration of SFS can be done by updating the surface normals using

$$\mathbf{n}^{(t+1)}(x,y) = \frac{1}{4} \sum_{(i,j)\in\Omega(x,y)} \mathbf{n}^{(t)}(i,j), \tag{4}$$

where $\Omega(x, y) = \{(x+1, y), (x-1, y), (x, y+1), (x, y-1)\}$ is the local neighborhood. The structure preserving regularization constraints in [11] modify Equation (4) by introducing a weighting scheme. The idea is that adjacent pixels with closer intensities are more likely to have similar surface normal directions. Instead, surface normals are updated using

$$\mathbf{n}^{(t+1)}(x,y) = \frac{\sum_{(i,j)\in\Omega(x,y)} W(i,j)\mathbf{n}^{(t)}(i,j)}{\| \sum_{(i,j)\in\Omega(x,y)} W(i,j)\mathbf{n}^{(t)}(i,j)\|}, \tag{5}$$

where $W(i, j)$ is a normalized measure of the intensity similarity between pixel (i, j) and the current pixel (x, y). It provides surface smoothness when adjacent pixels have similar intensities, but smoothing is reduced when there are large differences in intensities. During each SFS iteration, this weighted updating of surface normals is iterated until convergence (the angular difference between $\mathbf{n}^{(t)}$ and $\mathbf{n}^{(t+1)}$ is less than a predefined ξ) or a predefined maximum number of iterations (set to 200 in our experiments).

Figure 8: Surface normal adjustment. Left: result before adjustment; right: after adjustment.

Our variant of this approach replaces the weight $W(i,j)$ in Equation (5) with the saliency value at location (i,j). Thus, updated surface normals are more determined by positions with high saliency values than with low saliency values, which helps to preserve salient facial features.

5.3. Surface Normal Adjustment

After the surface normals have been recovered from the image by iteratively satisfying the above regularization constraint and brightness constraint, we apply a further step of postprocessing. Suppose at position (x,y), the angle between the recovered surface normal and the light source direction is $\theta(x,y) = \cos^{-1}(\mathbf{n}(x,y)\cdot\mathbf{s})$, and the saliency value normalized to $[0,1]$ is $w(x,y)$. Then, we adjust the angle to be

$$\hat{\theta}(x,y) = w(x,y)\theta(x,y). \tag{6}$$

Together with the light source direction $\mathbf{s}$, this defines a new cone at position (x,y). We rotate $\mathbf{n}(x,y)$ to its closest on-cone position. Adjusted in this way, we reduce differences of surface normals in areas with low saliency values, while increasing differences between areas with low saliency values and areas with high saliency values. As a result, we achieve a smoother surface with more prominent features. An example of relief surfaces generated with and without this adjustment step are shown in Figure 8.

6. Combination of Relief Surfaces

Our whole process (training, generating bas-relief images, and shape-from-shading) is based on predefined lighting directions. We use lighting from above (as this is natural), and to one side, to emphasize facial features.

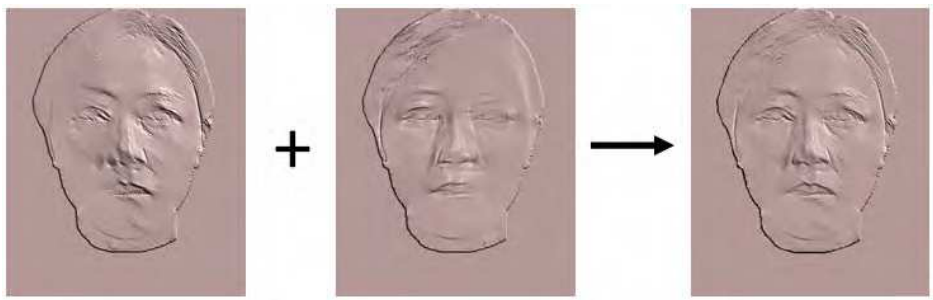


Figure 9: Combination of left- and right-illuminated relief surfaces.

The drawback is that features are revealed in an uneven way. Features inside shadows, and those facing the light, are hard to see, while those in other areas are revealed much better. We overcome this difficulty by repeating the whole model building process *twice* using two symmetric lighting directions from upper right $(1, 1, 1)$ and upper left $(-1, 1, 1)$. *Two* bas-relief surfaces are generated, and we use the average surface as the final output (alternatives to this approach are discussed further later). Figure 9 shows an example of the two bas-relief surfaces generated from the same original photograph, and their average. These two surfaces were recovered from the two generated bas-relief images in Figure 7. The average surface combines features independently revealed by the two surfaces, and further smooths out noise.

7. Experimental Results and Discussion

We now present various results obtained using our method. Various issues should be considered when deciding if the results are satisfactory. The first is whether the salient features are distinct and well-preserved, making the face recognisable, and can be best assessed by visual inspection of the results. The second is whether the geometry of the generated bas-relief is appropriate, so that the relief's appearance is stable under changes of viewing and illuminating directions. We show height maps of the generated bas-reliefs to reveal their overall geometries. (As shape-from-shading is an ill-posed problem, it is possible to recover a shape which looks correct from the original viewing direction, but is clearly the wrong shape when viewed from another direction—for example, it is well-known that convexity and concavity can be reversed [23]). A third issue is that the results should not contain unwanted noise.

Figure 10: Bas-relief surfaces generated using different saliency scaling factors l. Left to right: surfaces using $l = 1, 4, 8, 16$, and 32.

In the first experiment, we examine how varying the scaling factor l in the saliency map calculation affects the amount of detail in the generated bas-relief surfaces. Figure 10 shows bas-relief surfaces generated using $l = 1, 4, 8, 16, 32$; as l increases, the surfaces show more detail, but also more noise. When $l = 1$, salient features are not clearly revealed. For $l = 4, 8, 16, 32$, the differences between the surfaces are more subtle. A suitable compromise seems to be $l = 8$, which we used in other experiments. We note that real reliefs on coins often prefer smoothness of the relief at the expense of fine detail.

In the second experiment, we assess the overall geometry of the generated bas-relief surfaces, and their appearance under different lighting directions. Figure 11 shows generated bas-relief surfaces using $l = 8$, together with their height fields which help to reveal their overall geometry. We also give views of the surfaces when illuminated under four different lighting directions: $(1, 1, 1)$, $(-1, 1, 1)$, $(-1, -1, 1)$, and $(1, -1, 1)$. We can see that the generated bas-relief surfaces are smooth and maintain the salient facial features in each case. The overall geometry of each bas-relief is globally of the desired shape, which ensures that its appearance is as expected under changes of viewing and lighting directions. One drawback is that the lips are surprisingly and somewhat undesirably lower than the surrounding area. This is because these areas are typically dark in the face, but in the SFS process, we have assumed constant albedo without taking such coloration into account. The SFS method can only produce the coloration by a geometric adjustment, and in doing so, the dark area poses the concave / convex ambiguity problem. On the other hand, the same effect is beneficial elsewhere in the image: eyebrows in particular are clearly visible in the result, even though geometrically they are close to the underlying face. A possible improvement could be obtained

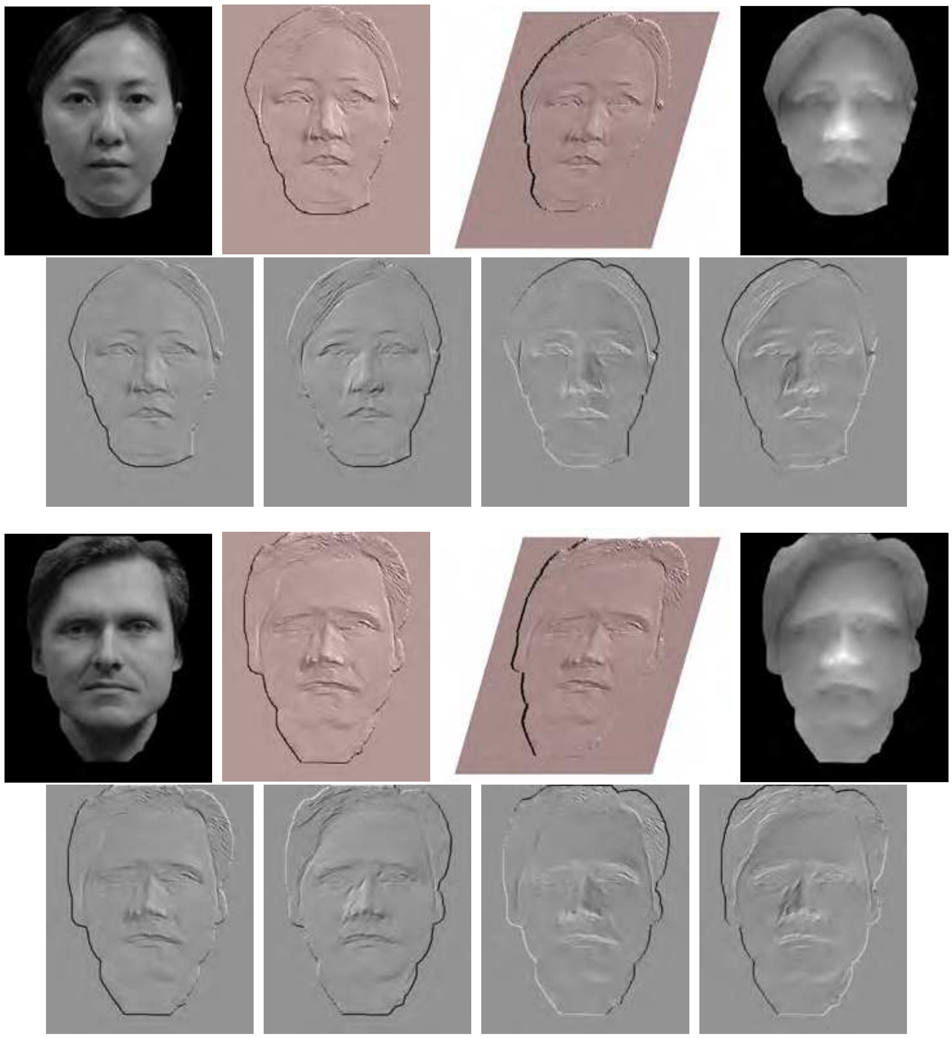


Figure 11: Output bas-relief surfaces. Rows 1, 3: original photograph, relief surface viewed from 2 angles, and the corresponding height fields. Rows 2, 4: views of the relief surface using four different lighting directions: $(1, 1, 1)$, $(-1, 1, 1)$, $(-1, -1, 1)$, and $(1, -1, 1)$.

by taking facial albedo into account during SFS, at least for the lips.

Further results are shown in Figure 12, using photographs captured under ambient (rather than directional) light. Figure 13 shows results from public domain photographs of various famous people. Faces were cropped from backgrounds manually. In each case, reasonable bas-relief surfaces were produced. One limitation is that teeth (last row in Figure 12 and Figure 13)

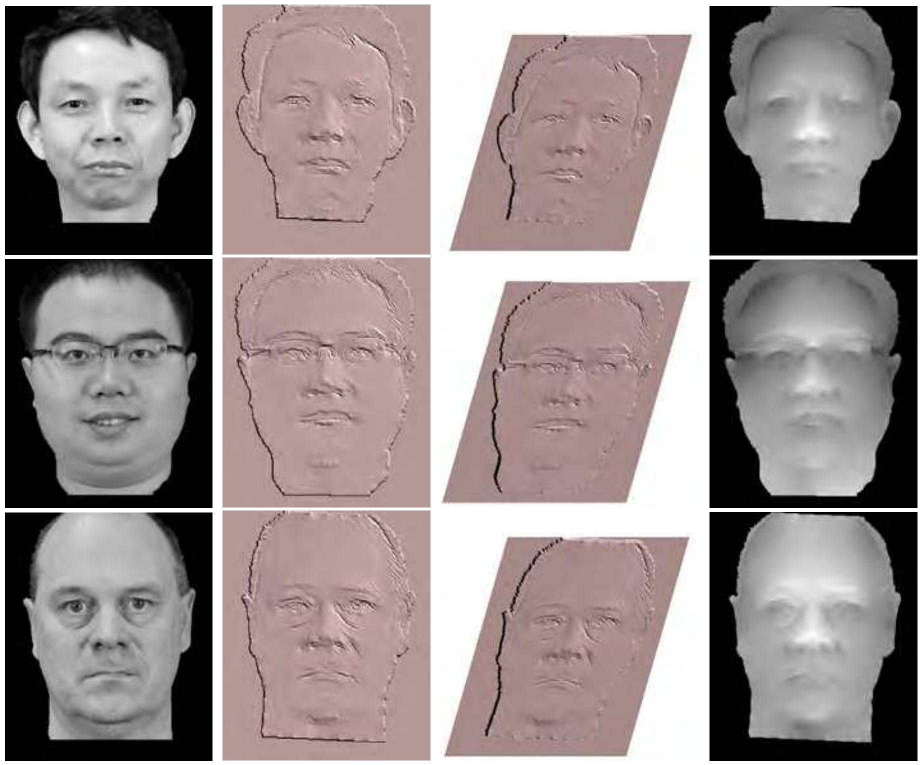


Figure 12: Further reliefs produced from photographs under ambient light.

and extensive hair (first row in Figure 13) are not handled well, because they are not well represented in the relief training data and bootstrap images for relighting. A further possible improvement would be to enlarge the training and bootstrap sets to include various facial albedos and expressions.

Finally, we applied our method to a photograph of a non-frontal face—see Figure 14. The generated bas-relief surface reveals the general shape of the face and maintains the prominent features. However, there are artifacts around the eyes and mouth. Figure 14 makes it clear that the artifacts are introduced during image relighting. The bootstrap set used for image relighting was entirely composed of frontal faces. Our simple alignment procedure did not do a good job of aligning this image to the bootstrap set, causing the artifacts observed. Better fine alignment, or a point-to-point correspondence method is likely to improve the results.

Our prototype implementation using MATLAB 7.9.0. Approximate computational times taken by each step of our method are shown in Table 7, for

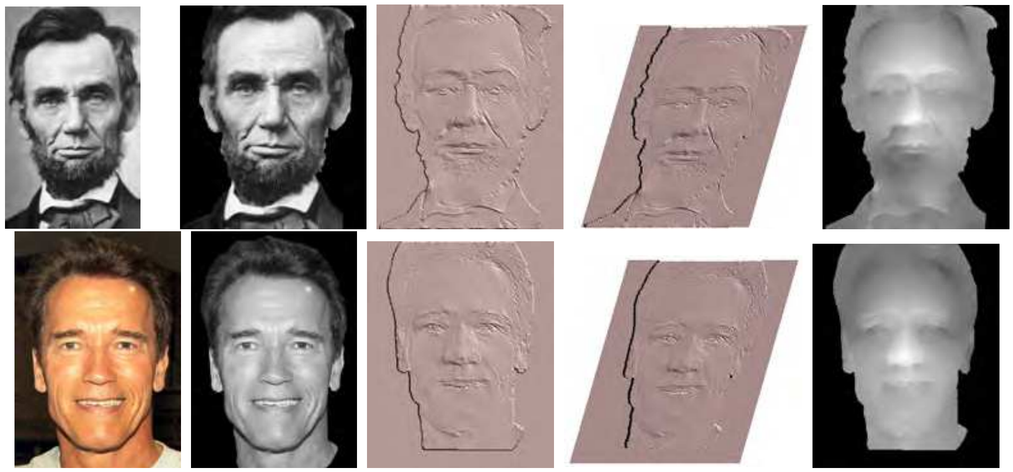


Figure 13: Reliefs of famous people. The first two columns show the input photograph, and the aligned grayscale image derived from it.

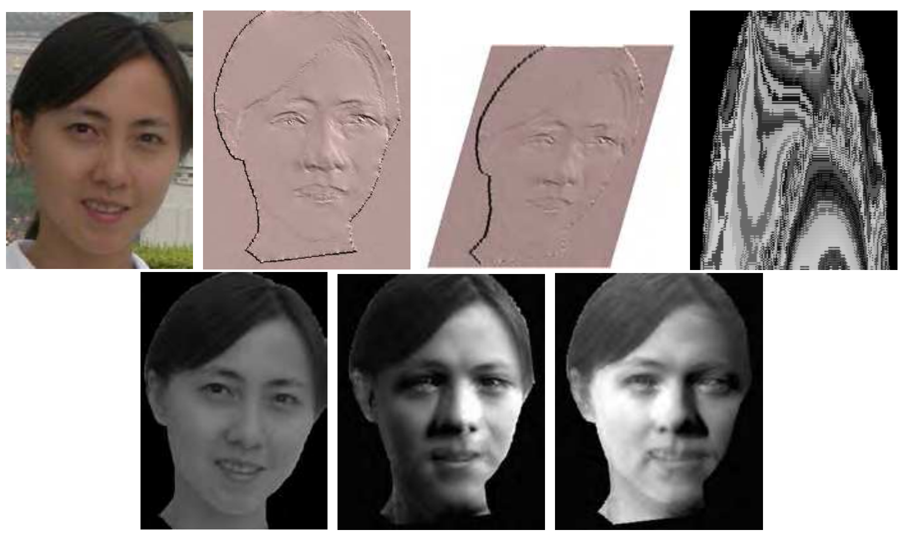


Figure 14: Results on photographs of a non-frontal face. Top: photograph and relief, bottom: relit images.

images of size 701×841. Neural network training step took the longest time (3 hours) but needs doing only once. Given a new photograph, there are five steps to get the final bas-relief surface, taking about 5 minutes in total; this could probably be reduced by a high-level language implementation. Note that the time for image relighting includes the time for manually marking landmarks to perform coarse alignment.

Table 1: Approximate timings.

Step	Time
Neural Network Training	3 hours
Saliency Map Calculation	16 seconds
Image Relighting	16 seconds
Generating Relief Images	8 seconds
Shape from Shading	4 minutes
Surface Combination	0.05 seconds

8. Variants

We finish by considering various alternative strategies we have investigated, but rejected.

First, in the network training process, we train a single neural network from the training data. However, to generate a plausible bas-relief surface, areas with low saliency and high saliency should be compressed in different ways. *Identical* local neighborhoods in the input image may lead to pixels with *different* values in the relief image, in places of different saliency. To allow for this, we considered an alternative strategy during neural network training. We divided the input image into several bands according to the saliency value of each pixel, and trained a separate network for each band. We perform experiments using 2, 3, 5, and 10 bands, and compare the results with using a single band (as described earlier). The generated bas-relief images and corresponding bas-relief surfaces are shown in Figure 15. It is clear that greater intensity variation occurs in the generated bas-relief images when using more bands, and the salient features are more pronounced than when using one band. These more strongly emphasized areas protrude more in the final bas-relief surfaces. However, whether such protruding features are desired in bas-relief creation remains an open question. We can see no obvious reason for preferring the results using multiple bands, and indeed, in places they can look worse—e.g. the hair line looks less natural in these examples.

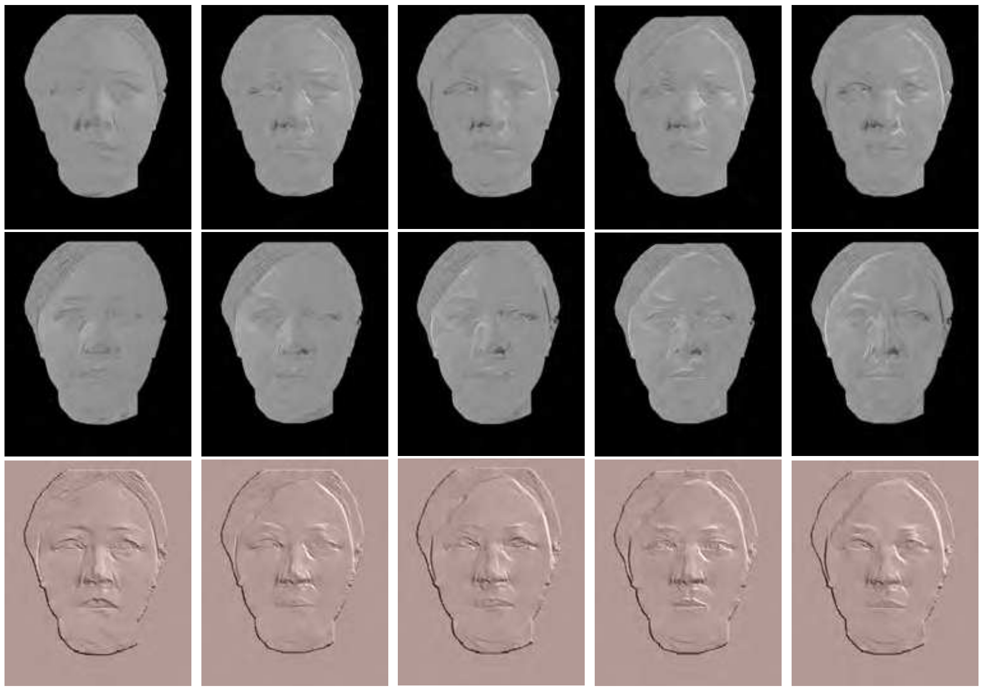


Figure 15: Bas-relief images (with 2 lighting directions) and surfaces generated using 1, 2, 3, 5 and 10 saliency bands.

Secondly, in the surface combination step, we average the two surfaces S_1 and S_2, which are recovered under two lighting directions, to get the final bas-relief surface. However, as we have noted earlier, each image contains some areas in shadow, or with highlights, which lead to poor shape recovery, and it is plausible that rather than simply *averaging* the two relief surfaces produced, we should use some sort of *selection* procedure to locally choose the good parts from each. Shadows and highlights have intensities far from the mean intensity, so we should preferentially use shape information from the image whose intensity is closest to the mean intensity. Suppose I_1 and I_2 are the two relit images under lighting directions $(1, 1, 1)$ and $(-1, 1, 1)$ and $\bar{I} = (I_1 + I_2)/2$ is the mean intensity value. We compute the absolute difference between the two images and the mean value, i.e.

$$\Delta_1(x, y) = |I_1(x, y) - \bar{I}|, \quad \Delta_2(x, y) = |I_2(x, y) - \bar{I}|. \tag{7}$$

Then, we define a combination map

$$M(x,y) = \begin{cases} 1 & \Delta_1 \leq \Delta_2 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

The top left image in Figure 16 illustrates this combination map. An alternative, to avoid abrupt transitions is to use a weighted version M' of M (see the bottom left image in Figure 16):

$$M'(x,y) = \frac{\Delta_2(x,y)}{\Delta_1(x,y) + \Delta_2(x,y)}. \tag{9}$$

The final bas-relief surface S is now produced from S_1 and S_2 using the combination map:

$$S(x,y) = M^*(x,y)S_1(x,y) + \left(1 - M^*(x,y)\right)S_2(x,y), \tag{10}$$

where M^* is either M or M'. The middle column of Figure 16 shows the combined bas-relief surfaces using combination maps M (top row) and M' (bottom row). It is clear that when using combination map M, there are discontinuities where the two surfaces meet. Using the weighted combination map M' mitigates this problem, but the output surface is still noisy. An alternative to further avoid this issue is to use the weighted combination map to take surface normals values from S_1 and S_2, and integrate them using the algorithm of Frankot and Chellappa [22]. The bottom right image in Figure 16 shows the resulting bas-relief surface. Compared to the bas-relief surface combined using simple averaging (the top right image in Figure 16), the final bas-relief emphasises features more strongly, but is perhaps less aesthetically pleasing as defects are also more obvious. This last approach is also somewhat more computationally expensive.

9. Conclusions and future work

Bas-reliefs of human faces are of prticular interest in art and design. We have given a method, based on neural networks, image relighting, and shape-from-shading techniques to automatically generate bas-reliefs from frontal photographs of faces. Experimental results show that our method is capable of generating reasonable bas-relief surfaces from such photographs, and are a first step towards automating this process to assist artists.

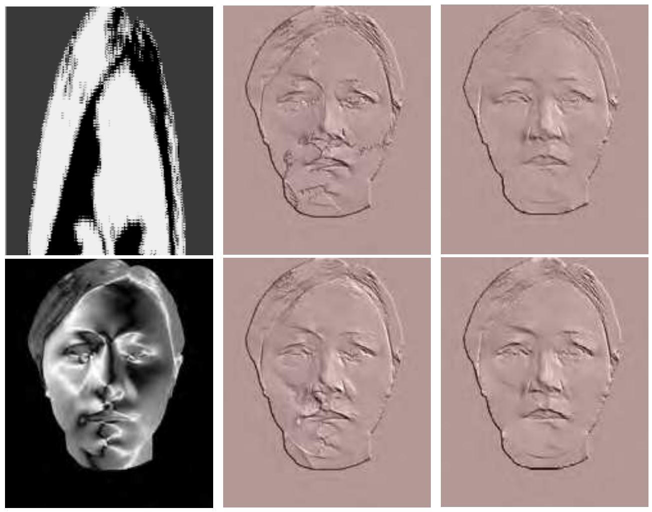


Figure 16: Alternative surface combination methods. Top: 0–1 combination map, relief from 0–1 map, relief using default averaging approach. Bottom: Weighted combination map, relief from weighted map, relief using weighted map to produce normals and integrating.

While we have already experimented with some variants of our approach, there is clearly room for improvement, and we suggest a few avenues that could improve our method further. In image relighting, the simple coarse alignment method used results in various artifacts which are visible in the final output, especially when applying the method to semi-profile faces. Better fine alignment, or a more sophisticated point-to-point correspondence method could reduce this problem. Improvements could be made by taking into account facial albedo information during the SFS step, and other reflectance models than the simple Lambertian model used here may also further improve the results. Clearly, in the function learning process, more than one training image, and training images from real face models, could also improve our results. An enlarged bootstrap set in the image relighting process could better span the space of facial albedos, and as a result, could also improve the results. Finally, practical applications demand extension of our method to faces seen in profile, and to a wider class of objects.

References

[1] P. Cignoni, C. Montani, R. Scopigno, Computer assisted generation of bas- and high-reliefs, Journal of Graphics Tools 2 (3) (1997) 15–28.

[2] W. Song, A. Belyaev, H. Seidel, Automatic generation of bas-reliefs from 3d shapes, in: Proceedings of IEEE International Conference on Shape Modeling and Applications, 2007, pp. 211–214.

[3] J. Kerber, A. Belyaev, H. Seidel, Feature preserving depth compression of range images, in: Proceedings of the 23rd Spring Conference on Computer Graphics, 2007, pp. 110–114.

[4] T. Weyrich, J. Deng, C. Barnes, S. Rusinkiewicz, A. Finkelstein, Digital bas-relief from 3d scenes, in: ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2007, Vol. 26, 2007.

[5] X. Sun, P. L. Rosin, R. R. Martin, F. C. Langbein, Bas-relief generation using adaptive histogram equalization, IEEE Transactions on Visualization and Computer Graphics 15 (4) (2009) 642–653.

[6] J. Kerber, A. Tevs, A. Belyaev, R. Zayer, H.-P. Seidel, Feature sensitive bas relief generation, in: Proceedings of IEEE International Conference on Shape Modeling and Applications (SMI), 2009, pp. 148–154.

[7] M. Alexa, W. Matusik, Reliefs as images, in: ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2010, Vol. 29, 2010.

[8] Z. Li, S. Wang, J. Yu, K.-L. Ma, Restoration of brick and stone relief from single rubbing images, IEEE Transactions on Visualization and Computer Graphics.
URL `http://doi.ieeecomputersociety.org/10.1109/TVCG.2011.26`

[9] R. Zhang, P.-S. Tsai, J. Cryer, M. Shah, Shape from shading: A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (8) (1999) 690–706.

[10] P. Worthington, E. Hancock, New constraints on data-closeness and needle map consistency for shape-from-shading, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (12) (1999) 1250–1267.

[11] R. Huang, W. Smith, Structure-preserving regularisation constraints for shape-from-shading, in: International Conference on Computer Analysis of Images and Patterns, 2009, pp. 865–872.

[12] T. Riklin-Raviv, A. Shashua, The quotient image: Class based re-rendering and recognition with varying illuminations, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2) (2001) 129–139.

[13] R. Fattal, D. Lischinski, M. Werman, Gradient domain high dynamic range compression, ACM Transactions on Graphics 21 (3) (2002) 249–256.

[14] J. Kerber, Digital art of bas-relief sculpting, Master's thesis, University of Saarland, Saarbrucken, Germany (2007).

[15] B. Horn, M. Brooks, The variational approach to shape from shading, Computer Vision, Graphics, and Image Processing 33 (2) (1986) 174–208.

[16] P. Worthington, E. Hancock, Needle map recovery using robust regularizers, Image and Vision Computing 17 (8) (1999) 545–558.

[17] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: In Proceedings of the IEEE International Conference on Computer Vision, 1998.

[18] T. Fine, Feedforward neural network methodology, Springer Verlag, 1999.

[19] A. Shashua, Illumination and view position in 3d visual recognition, in: In Proceedings of the fourth annual conference on Advances in Neural Information Processing Systems, 1991, pp. 404–411.

[20] A. Shashua, On photometric issues in 3d visual recognition from a single 2d image, International Journal of Computer Vision 21 (1997) 99–122.

[21] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: Illumination cone models for face recognition under variable lighting and pose, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (6) (2001) 643–660.

[22] R. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 10 (4) (1988) 439–451.

[23] E. Prados, O. Faugeras, Shape from shading: a well-posed problem?, in: In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, 2005, pp. 870–877.