

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), yy1@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

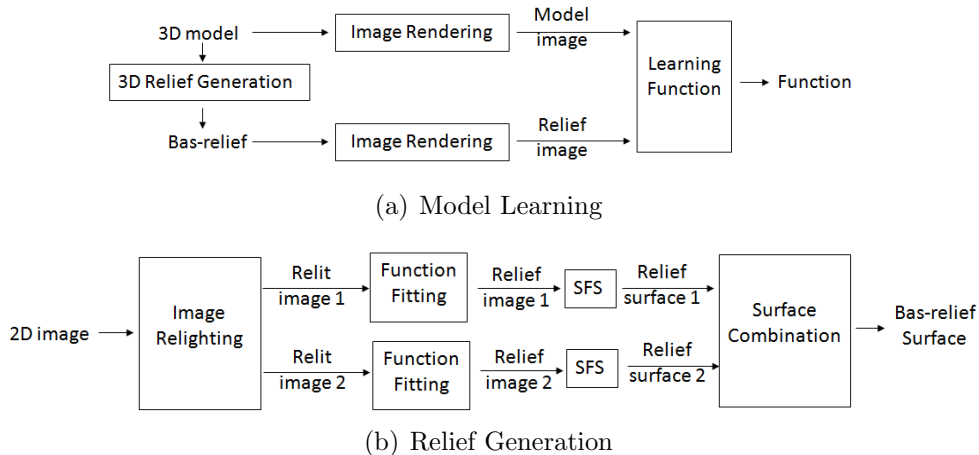


Figure 1: The proposed framework

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

43 Our overall framework has two components, shown in Figure 1. First,
 44 an offline process is used to learn the relationship between an image of a
 45 3D human face and an image of a corresponding 3D bas-relief of that face.
 46 This is done by taking one or more 3D face models, and processing them
 47 using any existing bas-relief generation algorithm to produce corresponding
 48 3D bas-reliefs. Each original 3D model and corresponding bas-relief are then
 49 rendered to give 2D images, using one or more lighting conditions. A learning
 50 algorithm is used to model the relationship between the pixel values in these
 51 images. While here we use an existing 3D bas-relief generation algorithm
 52 for simplicity, an alternative would be to learn the relationship using pho-
 53 tographs of human faces and handcrafted bas-reliefs of those faces derived
 54 from those photographs. This would avoid any deficiencies in existing bas-
 55 relief generation algorithms (but would also necessitate careful registration
 56 of reliefs and photographs).

57 Once we have learnt the model between 2D face images, and 2D face

58 bas-relief images, we can input a new face image, and apply the model to
59 determine what a corresponding bas-relief model should look like. We then
60 apply SFS to recover the bas-relief surface from the generated bas-relief im-
61 age. In practice, we find that if we re-light the input image from several new
62 directions [12], giving multiple versions of the input image, and use each to
63 determine a bas-relief, these can be combined into a more satisfactory final
64 bas-relief.

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

72 **2. Related Work**

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

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

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

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

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

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

129 Our work employs existing SFS techniques, which recover shape from in-
130 tensity variation in an image. A survey of early SFS work can be found in [9].
131 Assuming Lambertian reflectance and a known directional light source, Horn

132 and Brooks [15] gave a variational approach to solve the SFS problem. The
133 energy to be minimised comprises a brightness constraint and a quadratic
134 regularizing term enforcing surface smoothness. However, this method in-
135 volves the choice of a Lagrange multiplier, and the results tend to be over-
136 smoothed. To overcome these deficiencies, Worthington and Hancock [10]
137 proposed a geometric SFS framework which strictly satisfies the brightness
138 constraint at every pixel: surface normals are forced to lie on their irradiance
139 cones during each iterative update. The same authors have also given sev-
140 eral robust regularizers with better smoothing behaviour than the quadratic
141 one [16]. Huang and Smith [11] gave a structure-preserving regularization
142 constraint, which allows smoothing to be performed locally, dependent on
143 the intensities in a local area. We adopt the last method, as it is particularly
144 suited to our requirement to preserve salient facial features.

145 3. Mapping face images to face bas-relief images

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

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

160
161 We thus turned to understanding and modeling the mapping between
162 intensities in images of faces and images of corresponding bas-reliefs. It soon
163 became clear that a simple function is not adequate for this purpose. Some
164 explicit image processing methods, such as image embossing, can produce an
165 image with a bas-relief-like effect. However, these methods usually change
166 the reflectance properties of the surface, and the lighting conditions in the
167 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.

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

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

181 *3.1. Generating Bas-reliefs for Training*

182 To learn the mapping from images of faces to images of bas-reliefs of
183 faces, we need corresponding pairs of images. Given one or more 3D face
184 models, we need to generate corresponding 3D bas reliefs. We do so using
185 Sun’s method [5], which we briefly summarise. Starting from a height map
186 of the face (i.e. a range image), it performs histogram equalization of heights
187 within a local neighborhood for each point. Two modifications are applied
188 to this local histogram equalization. First, the calculation of the histogram
189 is weighted by the gradient magnitude after applying a non-linear transfor-
190 mation, in order to preserve small shape details. The second modification
191 applies an iterative clipping and redistribution procedure to the local his-
192 tograms, limiting their content. This prevents too many counts in any one

193 histogram bin, which would result in shape distortion and increased noise.
194 A scaling factor l controls this limit for each bin’s content. To generate the
195 final bas-relief surface, the method processes the input height maps using sev-
196 eral different neighborhood sizes, and averages the results. Figure 2 shows
197 a scanned head of Julius Caesar and the final bas-relief produced using the
198 method.

199 3.2. Saliency Map Calculation

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

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

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

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

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

222 Examples of saliency maps calculated from images rendered using mesh
223 models, and from photographs, are shown in Figure 3; they have resolutions
224 of 596×852 and 701×841 respectively. We use 256 equal-sized bins during
225 histogram equalization, and a radius of 3 for the circular averaging filter.
226 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.

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

229 3.3. Feedforward Network Training

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

239 In our experiments, we used a 3D model of Julius Caesar and a corre-
 240 sponding generated bas-relief (as shown in section 3.1) to generate the train-
 241 ing model images and bas-relief images. We generated two pairs of corre-
 242 sponding training images using Lambertian reflectance and parallel lighting,
 243 from lighting directions, $(1, 1, 1)$ and $(-1, 1, 1)$, respectively (with z towards
 244 the model), as shown in Figure 4. For each pair of training images, our
 245 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 $(-1, 1, 1)$.

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

249 4. Generating Bas-relief Images

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

259 4.1. Image Relighting

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

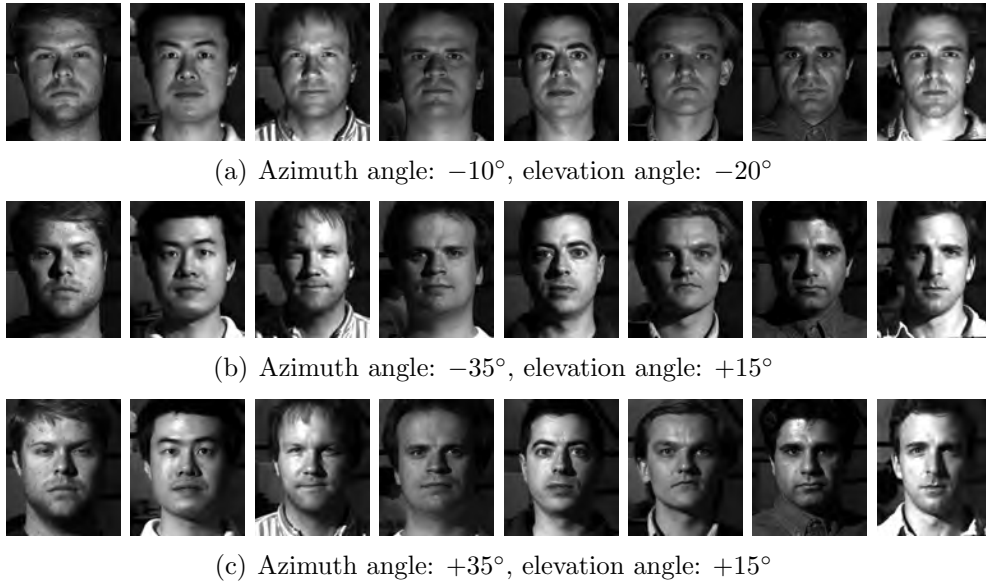


Figure 5: Bootstrap set for image relighting.

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

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

282 Figure 6 shows examples of applying image relighting using this training
 283 data. Two images of the same person are shown under different lighting.
 284 Apart from shadows, the quotient images are quite similar, and approxi-
 285 mately invariant to changes in light source as hoped. The quotient image
 286 technique unfortunately cannot take shadows into account. Relighting im-
 287 ages 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 $(-1, 1, 1)$.

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

292 4.2. Generating the Bas-relief Images

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

301 Examples of generated bas-relief images are shown in Figure 7(The inten-
 302 sity of the relief images are linearly stretched for showing purpose.). Salient
 303 facial features are preserved in the generated images, giving these images
 304 recognizable bas-relief appearance. The lighting directions used in the re-
 305 lit model images are also evident in the bas-relief images, and are utilized
 306 directly in the following shape-from-shading step.

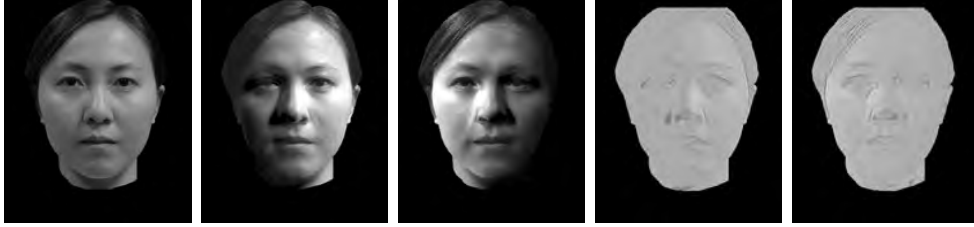


Figure 7: A generated bas-relief image. Left to right: original image, two relit images, and corresponding bas-relief images.

307 5. Finding the Relief using Shape-from-shading

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

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

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

319 Our SFS method improves upon this formulation in two ways. First, we
 320 satisfy intensity closeness as a hard constraint using the method of Wor-
 321 thington and Hancock [10]. The aim is to preserve the appearance of the
 322 image, which is important in our application. Secondly, we use a modified
 323 version of Huang and Smith's [11] structure-preserving regularization con-
 324 straint, which helps to preserve salient facial features. Our SFS method is
 325 iterative. In each iteration, the surface normals are updated to first satisfy
 326 the regularizing term, and secondly to satisfy the brightness constraint. Fi-
 327 nally, we use the algorithm of Frankot and Chellappa [22] to integrate the
 328 field of recovered surface normals to generate the bas-relief surface. We now

329 give further details.

330 5.1. Brightness Constraint

331 For Lambertian surfaces, satisfying the intensity closeness as a hard con-
332 straint is equivalent [10] to enforcing

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

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

339 5.2. Regularization Constraint

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

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

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

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

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

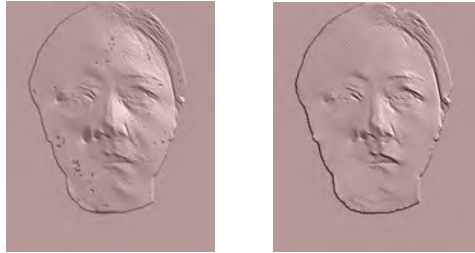


Figure 8: Surface normal adjustment. Left: result before adjustment; right: after adjustment.

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

358 5.3. Surface Normal Adjustment

359 After the surface normals have been recovered from the image by it-
 360 eratively satisfying the above regularization constraint and brightness con-
 361 straint, we apply a further step of postprocessing. Suppose at position (x, y) ,
 362 the angle between the recovered surface normal and the light source direction
 363 is $\theta(x, y) = \cos^{-1}(\mathbf{n}(x, y) \cdot \mathbf{s})$, and the saliency value normalized to $[0, 1]$ is
 364 $w(x, y)$. Then, we adjust the angle to be

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

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

372 6. Combination of Relief Surfaces

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



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

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

387 7. Experimental Results and Discussion

388 We now present various results obtained using our method. Various issues
 389 should be considered when deciding if the results are satisfactory. The first
 390 is whether the salient features are distinct and well-preserved, making the
 391 face recognisable, and can be best assessed by visual inspection of the results.
 392 The second is whether the geometry of the generated bas-relief is appropriate,
 393 so that the relief’s appearance is stable under changes of viewing and
 394 illuminating directions. We show height maps of the generated bas-reliefs to
 395 reveal their overall geometries. (As shape-from-shading is an ill-posed prob-
 396 lem, it is possible to recover a shape which looks correct from the original
 397 viewing direction, but is clearly the wrong shape when viewed from another
 398 direction—for example, it is well-known that convexity and concavity can be
 399 reversed [23]). A third issue is that the results should not contain unwanted
 400 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 .

401 In the first experiment, we examine how varying the scaling factor l
 402 in the saliency map calculation affects the amount of detail in the gener-
 403 ated bas-relief surfaces. Figure 10 shows bas-relief surfaces generated using
 404 $l = 1, 4, 8, 16, 32$; as l increases, the surfaces show more detail, but also
 405 more noise. When $l = 1$, salient features are not clearly revealed. For
 406 $l = 4, 8, 16, 32$, the differences between the surfaces are more subtle. A suit-
 407 able compromise seems to be $l = 8$, which we used in other experiments.
 408 We note that real reliefs on coins often prefer smoothness of the relief at the
 409 expense of fine detail.

410 In the second experiment, we assess the overall geometry of the generated
 411 bas-relief surfaces, and their appearance under different lighting directions.
 412 Figure 11 shows generated bas-relief surfaces using $l = 8$, together with
 413 their height fields which help to reveal their overall geometry. We also give
 414 views of the surfaces when illuminated under four different lighting directions:
 415 $(1, 1, 1)$, $(-1, 1, 1)$, $(-1, -1, 1)$, and $(1, -1, 1)$. We can see that the generated
 416 bas-relief surfaces are smooth and maintain the salient facial features in each
 417 case. The overall geometry of each bas-relief is globally of the desired shape,
 418 which ensures that its appearance is as expected under changes of viewing
 419 and lighting directions. One drawback is that the lips are surprisingly and
 420 somewhat undesirably lower than the surrounding area. This is because
 421 these areas are typically dark in the face, but in the SFS process, we have
 422 assumed constant albedo without taking such coloration into account. The
 423 SFS method can only produce the coloration by a geometric adjustment, and
 424 in doing so, the dark area poses the concave / convex ambiguity problem. On
 425 the other hand, the same effect is beneficial elsewhere in the image: eyebrows
 426 in particular are clearly visible in the result, even though geometrically they
 427 are close to the underlying face. A possible improvement could be obtained

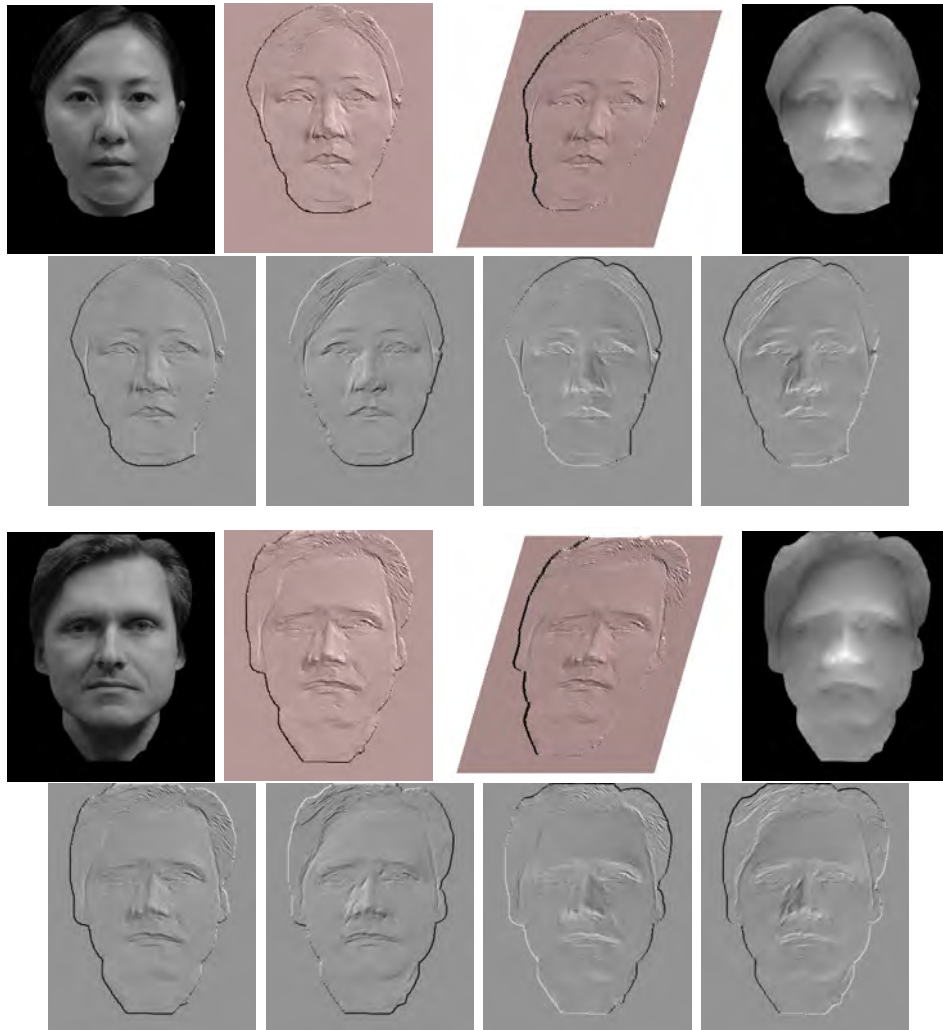


Figure 11: Output bas-relief surfaces. Rows 1, 3: original photograph, relief surface viewed from 2 angles, and the corresponding height fields. Rows 2, 4: views of the relief surface using four different lighting directions: $(1, 1, 1)$, $(-1, 1, 1)$, $(-1, -1, 1)$, and $(1, -1, 1)$.

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

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

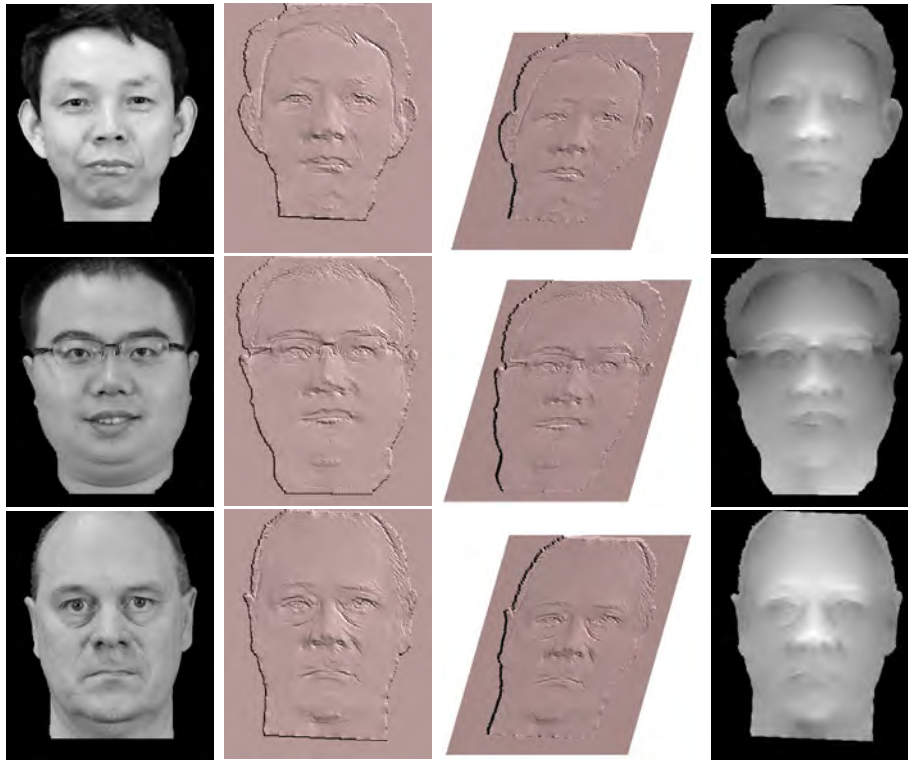


Figure 12: Further reliefs produced from photographs under ambient light.

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

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

447 Our prototype implementation using MATLAB 7.9.0. Approximate com-
 448 putational times taken by each step of our method are shown in Table 7, for

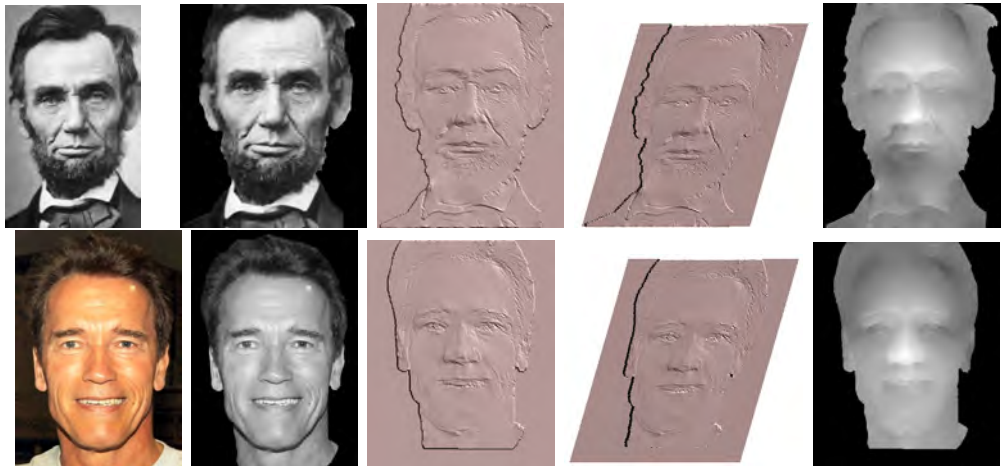


Figure 13: Reliefs of famous people. The first two columns show the input photograph, and the aligned grayscale image derived from it.

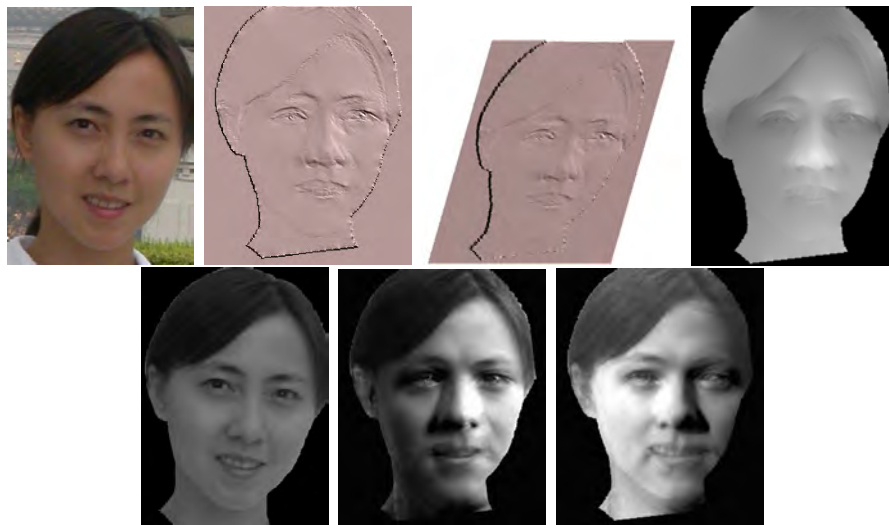


Figure 14: Results on photographs of a non-frontal face. Top: photograph and relief, bottom: relit images.

449 images of size 701×841 . Neural network training step took the longest time
 450 (3 hours) but needs doing only once. Given a new photograph, there are five
 451 steps to get the final bas-relief surface, taking about 5 minutes in total; this
 452 could probably be reduced by a high-level language implementation. Note
 453 that the time for image relighting includes the time for manually marking

454 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

455 8. Variants

456 We finish by considering various alternative strategies we have investi-
457 gated, but rejected.

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

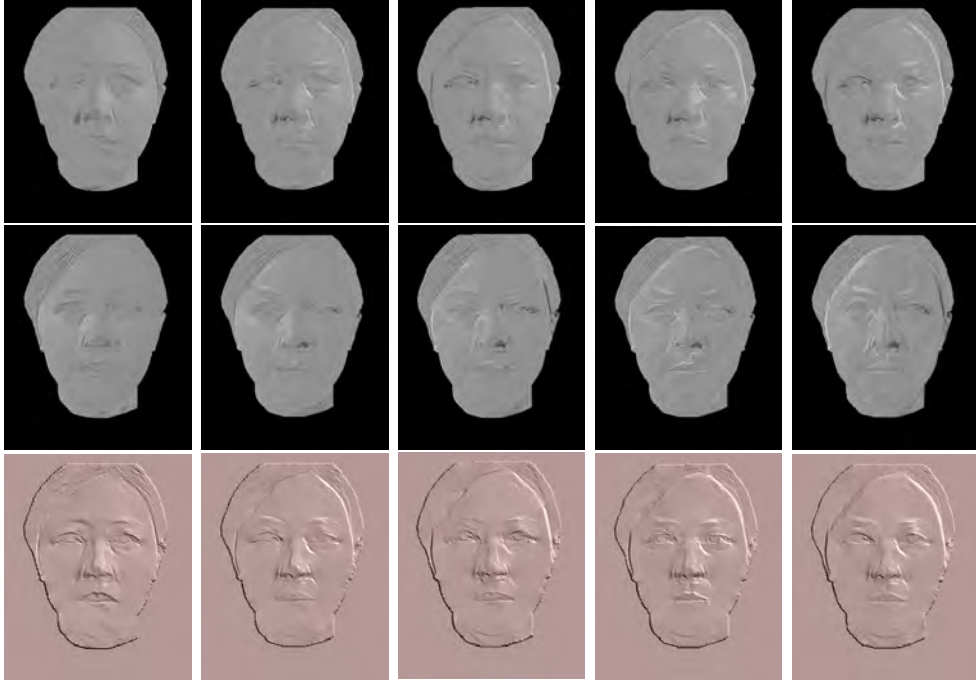


Figure 15: Bas-relief images (with 2 lighting directions) and surfaces generated using 1, 2, 3, 5 and 10 saliency bands.

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

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

489 Then, we define a combination map

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

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

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

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

$$S(x, y) = M^*(x, y)S_1(x, y) + (1 - M^*(x, y))S_2(x, y), \quad (10)$$

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

508 9. Conclusions and future work

509 Bas-reliefs of human faces are of particular interest in art and design. We
510 have given a method, based on neural networks, image relighting, and shape-
511 from-shading techniques to automatically generate bas-reliefs from frontal
512 photographs of faces. Experimental results show that our method is capable
513 of generating reasonable bas-relief surfaces from such photographs, and are
514 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.

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

530 **References**

- 531 [1] P. Cignoni, C. Montani, R. Scopigno, Computer assisted generation of
532 bas- and high-reliefs, *Journal of Graphics Tools* 2 (3) (1997) 15–28.
- 533 [2] W. Song, A. Belyaev, H. Seidel, Automatic generation of bas-reliefs from
534 3d shapes, in: *Proceedings of IEEE International Conference on Shape
535 Modeling and Applications*, 2007, pp. 211–214.
- 536 [3] J. Kerber, A. Belyaev, H. Seidel, Feature preserving depth compres-
537 sion of range images, in: *Proceedings of the 23rd Spring Conference on
538 Computer Graphics*, 2007, pp. 110–114.
- 539 [4] T. Weyrich, J. Deng, C. Barnes, S. Rusinkiewicz, A. Finkelstein, Digital
540 bas-relief from 3d scenes, in: *ACM Transactions on Graphics (TOG) -
541 Proceedings of ACM SIGGRAPH 2007*, Vol. 26, 2007.
- 542 [5] X. Sun, P. L. Rosin, R. R. Martin, F. C. Langbein, Bas-relief generation
543 using adaptive histogram equalization, *IEEE Transactions on Visualiza-
544 tion and Computer Graphics* 15 (4) (2009) 642–653.
- 545 [6] J. Kerber, A. Tevs, A. Belyaev, R. Zayer, H.-P. Seidel, Feature sensitive
546 bas relief generation, in: *Proceedings of IEEE International Conference
547 on Shape Modeling and Applications (SMI)*, 2009, pp. 148–154.
- 548 [7] M. Alexa, W. Matusik, Reliefs as images, in: *ACM Transactions on
549 Graphics (TOG) - Proceedings of ACM SIGGRAPH 2010*, Vol. 29, 2010.
- 550 [8] Z. Li, S. Wang, J. Yu, K.-L. Ma, Restoration of brick and stone relief
551 from single rubbing images, *IEEE Transactions on Visualization and
552 Computer Graphics*.
553 URL <http://doi.ieeecomputersociety.org/10.1109/TVCG.2011.26>
- 554 [9] R. Zhang, P.-S. Tsai, J. Cryer, M. Shah, Shape from shading: A survey,
555 *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (8)
556 (1999) 690–706.
- 557 [10] P. Worthington, E. Hancock, New constraints on data-closeness and
558 needle map consistency for shape-from-shading, *IEEE Transactions on
559 Pattern Analysis and Machine Intelligence* 21 (12) (1999) 1250–1267.

- 560 [11] R. Huang, W. Smith, Structure-preserving regularisation constraints for
561 shape-from-shading, in: International Conference on Computer Analysis
562 of Images and Patterns, 2009, pp. 865–872.
- 563 [12] T. Riklin-Raviv, A. Shashua, The quotient image: Class based re-
564 rendering and recognition with varying illuminations, IEEE Transac-
565 tions on Pattern Analysis and Machine Intelligence 23 (2) (2001) 129–
566 139.
- 567 [13] R. Fattal, D. Lischinski, M. Werman, Gradient domain high dynamic
568 range compression, ACM Transactions on Graphics 21 (3) (2002) 249–
569 256.
- 570 [14] J. Kerber, Digital art of bas-relief sculpting, Master’s thesis, University
571 of Saarland, Saarbrücken, Germany (2007).
- 572 [15] B. Horn, M. Brooks, The variational approach to shape from shading,
573 Computer Vision, Graphics, and Image Processing 33 (2) (1986) 174–
574 208.
- 575 [16] P. Worthington, E. Hancock, Needle map recovery using robust regular-
576 izers, Image and Vision Computing 17 (8) (1999) 545–558.
- 577 [17] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images,
578 in: In Proceedings of the IEEE International Conference on Computer
579 Vision, 1998.
- 580 [18] T. Fine, Feedforward neural network methodology, Springer Verlag,
581 1999.
- 582 [19] A. Shashua, Illumination and view position in 3d visual recognition, in:
583 In Proceedings of the fourth annual conference on Advances in Neural
584 Information Processing Systems, 1991, pp. 404–411.
- 585 [20] A. Shashua, On photometric issues in 3d visual recognition from a single
586 2d image, International Journal of Computer Vision 21 (1997) 99–122.
- 587 [21] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: Illu-
588 mination cone models for face recognition under variable lighting and
589 pose, IEEE Transactions on Pattern Analysis and Machine Intelligence
590 23 (6) (2001) 643–660.

- 591 [22] R. Frankot, R. Chellappa, A method for enforcing integrability in shape
592 from shading algorithms, IEEE Transactions on Pattern Analysis and
593 Machine Intelligence 10 (4) (1988) 439–451.
- 594 [23] E. Prados, O. Faugeras, Shape from shading: a well-posed problem?,
595 in: In Proceedings of IEEE Computer Society Conference on Computer
596 Vision and Pattern Recognition, Vol. 2, 2005, pp. 870–877.