Master Thesis

3D Reconstruction of Niniveh

Naphur van Apeldoorn

Master Thesis

3D Reconstruction of Niniveh

by

Naphur van Apeldoorn

in partial fulfilment of the requirements for the degree of

Master of Science
in Mechanical Engineering

at the Delft University of Technology,
to be defended publicly on Monday June 26, 2017 at 14:00 PM.

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Contents

3D Reconstruction of Niniveh

A casestudy on the Bas-reliefs of the Southwest Palace of Sennacherib with monocular images

Naphur van Apeldoorn

Delft University of Technology

June 20, 2017

Abstract

Parts of the Southwest Palace of Sennacherib (700 BC), which were located in Nineveh (Iraq), are now destroyed by the ongoing conflict in Iraq and Syria. The palace rooms used to be decorated with numerous Assyrian reliefs. Luckily, digital pictures of the site are available thanks to an Italian expedition in 2002. The goal of this research is to physically reproduce the destroyed reliefs using this photo database. First photogrammetry was used to rebuild the global dimensions of the reliefs. The details are reconstructed by fusing a highly detailed depth map retrieved with Shape from Luminance (SFL), together with a coarser Artificial Reconstruction. The fused reconstruction was evaluated with the eigenvalues of the Hessian. The eigenvalues give an intuitive measure about the shapes of the reliefs. The RMSE of the results improve up to 44.4% compared to the usage of an intensity (grey) image. The final 3D models are used to produce full size reproductions with 3D printing and CNC milling.

I Introduction

Preservation of cultural heritage has always been a delicate art. With the rise of 3D technologies in art preservation, a new set of tools became available to recover lost cultural heritage.

The current developments in the Middle East have led to a systematic destruction of cultural heritage in parts of Syria and Iraq. Heritage which survived for centuries, has been erased in years. This created an intriguing test case for advanced digital restoration techniques.

In this research we focused on the city of Niniveh in Iraq. Niniveh has suffered from the destruction of cultural heritage. Parts of the Southwest Palace of Sennacherib, built around 700 BC, which were still exhibited at their original location in Iraq, have now been destroyed by terror groups. Figure 1 shows the satellite pictures of the palace before and after destruction.

Figure 1: Left: Image taken by Digital Globe/ASOR on May 2, 2016 showing the Southwest Palace missing its roof but with reliefs still in place. Right: Image taken by Digital Globe/ASOR on May 9, 2016 showing the reliefs are gone and most internal walls have been destroyed.

The Rijksmuseum van Oudheden (RMO) collaborated with the Delft University of Technology, to create a replica of the destroyed palace rooms similar to their condition in 2002. The palace rooms used to be decorated with bas-reliefs, that were considered as some of the greatest stone carvings from antiquity [5]. The carvings often illustrate scenes of conquests or the hunt by Assyrian royals. A famous example is the Lion Hunt, displayed in the British Museum, Figure 2. Although the reliefs are relatively flat, the key-features are in the details. The main focus of this research: recovering the fine details and shape of the sculptures. Even though the original reliefs are lost, there are still pictures available from the site thanks to an expedition in 2002 by a group of Italian researchers [21]. Unfortunately these photos were never meant to be used for reconstruction. The challenge is to recover the original shape and details of the reliefs from uncalibrated monocular pictures. In this paper we introduce a new method for reconstructing the original 3D shape of Assyrian bas-reliefs from the region of Niniveh from monocular 2D images, where original is defined as the shape in 2002.

Figure 2: A close-up of the Lion Hunt relief at the British museum. It displays a royal fighting a lion, while his servant holds his weapon equipment. Original image by Naphur van Apeldoorn

i Related Literature

Digital Art Preservation, 3D imaging technology has been used before in art preservation. In 1999, Levoy et al. took the challenge to digitize 10 statues of Michelangelo as an application for 3D scanning of external shape and reflectance at Stanford University, [19]. In addition, Levoy et al. experienced noise in the range data due to the scattering of laser light on the marble structure. Ahmon expended the application of 3D scanning by producing a full-sized colour replica of $16m^2$ of the tomb of Seti in 2004, [2]. The 3D data was obtained with a short-range laser scanner. More recently, the Delft University of Technology managed to scan and reproduce paintings by Rembrandt and Van Gogh with a stereo camera set-up and the 3D printer of Océ [7].

The previous preservation projects only work when the objects are still physically accessible. In our case, the Assyrian reliefs have been destroyed. One of the most promising projects for the reconstruction of lost heritage is Rekrei former "Project Mosul" [29]. Rekrei uses photogrammetry with an online photo database to recover the shape of objects. Photogrammetry uses the difference in overlap (disparity) between multiple pictures of the same scenery to extract the 3D shape). There are several image pairs available within our database. However, these pairs fail to capture the fine details of the objects. This is often caused by the large distance between the camera and object. Fortunately, there are numerous close-ups available in our database, which can provide the details of the reliefs. This brings us to the field of monocular reconstructions.

Monocular image reconstruction, the main clue in monocular reconstruction is the illumination. One of the earliest applications of surface reconstruction from a monocular image was in 1966 by Rindfleisch [26]. He was able to obtain the lunar topography by using shading as depth cue. Horn [13] formulated a more general approach, which can recover the surface shape once the reflectivity function and the position of the light sources are known. In the survey of Zhang et al. [31], multiple Shape from Shading (SFS) algorithms have been reviewed on both synthetic and real images. The different methods all have their own strengths, but fail to make a reconstruction of more complex images, as for example a human face.

Recently, Delft University of Technology and QdepQ patented a technology which uses the luminance to transform a monocular 2D photo into 3D. The method is originally designed to produce an output on a 3D display. However, the 3D data is still too noisy for a reproduction. The difference between the display and the physical reproduction is in the number of viewpoints: while only a few (typically 2) viewpoints are needed for a display, a physical reproduction requires the reconstruction to be consistent from any point of view (including hidden parts). Moreover, the usage of a single photo also results in a loss of metrics (units of length), which are necessary for the physical reproduction.

Line drawings, in archaeology it is common to make line drawings of artefacts during excavation. These line drawings can be used as extra depth cue besides our main cue: illumination. Kolomenkin et al. [17] were able to make an impression of relief objects from only line drawings. They described the details of relief as a height function on a base. Where the lines indicate changes in the height function and the gradient in height is considered to be strong near the lines. Furthermore, they assume that the base shape is known and consist of a simple geometrical object as for example a plane or cylinder. Next, the height function of the lines is

projected over the base shape. The summation of both shapes is the reconstruction. Zeng et al. [30] combined the line drawings with the photo into a single solution. These line drawings were extracted from the input image and the resulting feature line image was used for segmentation, where each region was assigned its own depth value. Next, a base surface was generated from the region layers and converted into a 2D depth image. This depth map is then augmented with the grayscale image and gradient image of the original input image. In this way Zeng et al. combined the smoothness of the line reconstructions with the fine details of the original photo. We elaborated on this idea. However, instead of taking the gray image, we combined the SFL depth map with an Artificial depth map. Moreover, we merged both depth maps with a local fusion method.

ii Dataset

The dataset has been acquired from an Italian excursion in 2002 [21]. The dataset only contains photos from the slabs in room V. Most of the good quality reliefs from the other rooms are in possession of different museums around the world. One of these reliefs is part of the collection of the **RMO** and they granted us access to the relief for further research. We have 3D scanned this relief as reference material for our results. Furthermore, the archaeologists of the museum have provided us with the line drawings of the pictures.

The reliefs in the database are classified as bas-reliefs. The word bas-relief probably originates from the Italian word ''Bassorilievo'', which means low contrast. The name already reveals some of the characteristics of the reliefs:

- Flat base
- Lines of the sculptures define changes in height
- There are no overlapping parts within the relief. This implies that the complete structure is visible from a frontal view.
- The fine details are a key-feature of Assyrian reliefs.

The characteristics of the database are described as follow:

- The quality of the photos is limited to 6 megapixel (MP).
- The database consists of monocular 2D images, which means a loss of 3D information.
- Some photogrammetry measurements have been performed. Although these are not accurate enough to obtain the depth maps of the details within the relief, they can provide the absolute measurements for the outlines of the slabs.
- Most of the measurements are not compatible with the current software any more, which resulted in a loss of calibration data.

- The reliefs on the photos were already severely damaged by fire and wear at the moment of the expedition in 2002.
- The original objects are now destroyed.

he characteristics of the relief make it possible to use the SFL method, since there are no hidden parts on the photos.

The size of 6MP of the photos is small compared to average camera of current mobile phones. This affects the resolution of final reproduction. The photo size of 6MP is sufficient for close up. However, it will lead to a poor resolution for more distant photos and the details will be barely visible. Moreover, the fire and wear have will also have a negative effect on the SFL results.

The relation between the lines and changes in height make it possible to create an artificial reconstruction. The artificial reconstruction is used to compensate for negativ

iii Overview

Our research is divided into 3 parts. First, the novel reconstruction method is introduced. Second, an evaluation method for our digital reconstruction is proposed, which is used to tune the parameters and validate our reconstruction method. At last, the used reproduction methods are revealed and their results.

Our reconstruction method consists of 3 steps. The basic idea is to fuse the high detailed depth map from **Shape from Luminance** (SFL) with the smooth artificial reconstruction from the line drawings. The first part consists of transforming the RGB input image into a depth map with the Shape from Luminance algorithm. Second, a mathematical model of the reliefs is derived, which is used to convert the line drawings into an **Artificial Reconstruction**. For the third step, **Data Fusion**, the two different depth maps are fused together with a locally adaptive fusion algorithm. This combines the smoothness of the artificial reconstruction and the detailed SFL, by fusing the depth maps on their local strengths.

To determine the quality of the proposed method, a digital reconstruction is made of one of the reliefs at the RMO. A surface comparison, based on the eigenvalues of the Hessian, is proposed in **Evaluation Method** and next an **Overview of the Results** is given.

The last step consists of making a physical **Reproduction** of the 3D reconstruction. We have used 3 different digital manufacturing methods, CNC milling, 3D printing with sandstone and elevated printing. Each reproduction method is briefly examined.

II   Reconstruction Method

i   Shape from Luminance

The Assyrian reliefs are considered as relatively flat, but rich in details. In order to reconstruct these details, Shape from Luminance (SFL) is used. In short, the SFL algorithm changes the luminance signal of the input photo into moon-like conditions, so that a direct relation between the luminance gradient and the surface shape can be found. Note the usage of the word "surface", since it is impossible to obtain an absolute depth map with this method. The final result will be a relative depth map, where the surface is expressed in values between 0 and 1. In more detail, we divide the SFL algorithm in 4 steps:

1. Extraction
2. Decomposition
3. Selection
4. Reconstruction

Extraction, ideally you would like to use a camera set-up with an additional sensor which directly displays the luminance per pixel or the High Dynamic Range (HDR) option on most camera's. Unfortunately our photo database consists of RGB photos with a limited dynamic range of illumination. However, the luminance (Figure 5ii) can be extracted from the colour components of each pixel in the image (Figure 5i) by using their weighted value:

$$L(x,y) = 0.2116R(x,y) + 0.7152G(x,y) + 0.0722B(x,y) \tag{1}$$

The luminance is described as the intensity of light received by the observer and can be described as [15]:

$$L(x,y) = L_e + \int_\Omega f_r(\Theta_i, \Theta_o) L_i(\Theta_i) \cos \Theta_i d\omega_i \tag{2}$$

L_e is the luminance emitted by an object, which is zero except for light sources. f_r is the reflectance function, which describes the way the object's material interacts with light. $\Theta_{i,o}$ represents respectively the incoming and outgoing angle of light, L_i is the reflectance map and ω_i is the subtended angle.

The moon-like conditions are described as follow:

- if objects are not emitting light: $L_e = 0$
- if the objects are diffuse: L_i only depends on the angle between the incident lights and the surface normals.
- if the light source is unique and placed far away from the scene: The angle of light on the scene is constant, $\cos \Theta_i = c$

With these assumptions we have a direct relation between the surface shape and the luminance. To calculate the shape, the luminance signal has to be converted into these moon-like conditions, in order to make these assumptions valid.

Decomposition, the Emperical Mode Decomposition (EMD) is used for the conversion to moon-like conditions. For the sake of simplicity, we discuss the 1D case of the decomposition. The luminance signal, $L(t)$ is decomposed into a finite and small number of components, c_i and a residual r_n:

$$L(t) = \sum_n^{i=1} c_i(t) + r_n(t) \tag{3}$$

Where $t \in \Omega$, with Ω being a discrete multidimensional domain. In our 2D case it would be a grid of pixels in x and y direction.

The decomposition is performed with the EMD and the resulting components are called Intrinsic Mode Functions (IMF). The EMD has been developed to analyse adaptive data. This makes it possible to analyse data with non-linear and non-stationary properties, which is common in data obtained without a controlled environment. If the decomposition is performed correctly, each IMF will have a physical interpretation. For the readers which are not familiar with the EMD, a brief description is given based on the paper of Huang [14]. The results of the decomposition are illustrated in Figure 5iii - 5v.

The EMD is built on the assumption that, "At any given time, the data may have many coexisting simple oscillatory modes of significantly different frequencies, superimposed on each other." [14] Each component (IMF) satisfies two conditions:

1. The number of extrema and zero-crossings must either be equal or differ at most by one.
2. At any data point, the mean value of the envelopes defined by the local maxima and minima needs to be zero.

The process of obtaining an IMF is called sifting. The first step is to identify all local maxima and minima in signal $L(t)$, as shown in Figure 3b. Next, the upper and lower envelope are build to connect the local maxima and minima. The cubic spline is a popular method for building these envelopes in 1D. Their mean, m_1, is subtracted from the original signal to create the first protomode, h_1:

$$h_1 = L(t) - m_1 \tag{4}$$

Ideally, h_1 is expected to be an IMF (meet the pre-stated conditions). However, in reality this is often not the case. Therefore, the sifting procedure is duplicated and h_1 is treated as proto-IMF, Figure 3e:

$$h_{11} = h_1 - m_{11} \tag{5}$$

This procedure is repeated k number of iterations:

$$h_{1k} = h_{1(k-1)} - m_{1k} \tag{6}$$

until an IMF is found:

$$c_1 = h_{1k} \tag{7}$$

The sifting process is stopped once the normalized squared difference between two successive sifting operations is smaller than a predefined value, [14]:

$$SD_k = \frac{\sum\limits_{t=1}^{T} |h_{k-1}(t) - h_k(t)|^2}{\sum\limits_{t=1}^{T} |h_{1k}(t)|^2} \tag{8}$$

Once an IMF is obtained, the IMF is subtracted from signal $L(t)$, and the sifting procedure is repeated again, until no IMF's are present anymore. Which leaves the residual $r_n(t)$.

As mentioned in Equation 3, the luminance signal consists of several IMF's, c_i and a residual, r_n. The residual is the remainder of the original signal $L(t)$, once there are not enough extrema left in the signal to build the envelopes.

As for the decomposition of the 2D case, the procedure is mostly the same. The difficulty is in building the 2D envelopes. The solution can be found in the usage of thin plate splines or low pass filters.
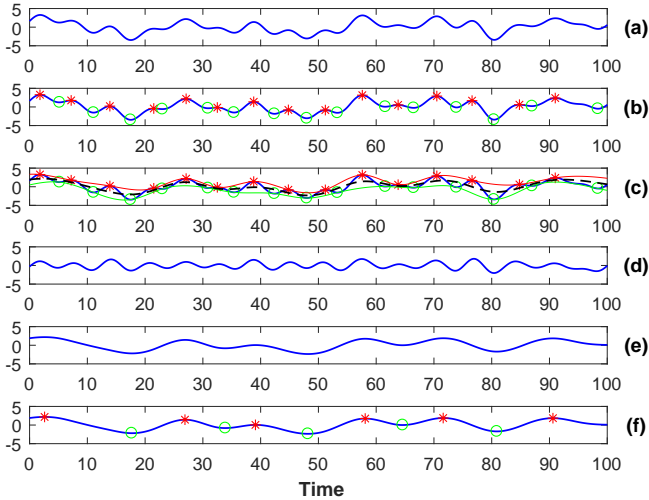
Figure 3: Sifting process of the empirical mode decomposition: **(a)** an arbitrary input; **(b)** identified maxima (red stars) and minima (green circles) superimposed on the input; **(c)** upper envelope (red) and lower envelope (green thin solid lines) and their mean (black dashed line); **(d)** intrinsic mode function (IMF),the difference between the bold blue solid line and the dashed line in subfigure **(c)**; **(e)** the remainder after an IMF is subtracted from the input; **(f)** and the sifting procedure is started again by looking for extrema. Figure is inspired by the paper of Huang. [14]

Selection, after decomposition, a selection of IMF's is used for the reconstruction. The number of IMF's depends on the separation criterion. The level of details decreases in subsequent IMF's, which is clearly visible in the decomposition results: Figure 5iii - 5v. The first mode optimally contains the information about textures and edges, while the remaining modes have more information about the shape of the objects. The residual consists of the illumination direction in the scene. This bias is caused by luminance which does not fit the assumption of the lunar conditions: a unique light source, placed far away from the scene. Cutting [6] found out that the human vision system uses a mixture of depth cues with weighted components depending on the scenery. This can be mimicked in our algorithm by selecting which IMF's to use and adding weights to each of them. After selection we end up with a luminance signal, Figure 5vi, which follows the moon-like conditions.

Reconstruction, the last step is to convert the filtered luminance map into depth values. Equation 9 is used to reconstruct the depth. This equation mimics the human photoreceptor response [12].

$$\mathrm{D_{SFL}}(x, y) = \frac{L_f^n(x, y)}{L_f^n(x, y) + \sigma^n} \tag{9}$$

with σ being the correction for illumination sensitivity:

$$\sigma^n = \exp\left(\frac{1}{N} \sum_{x,y} \log(L_f(x, y))\right) \tag{10}$$

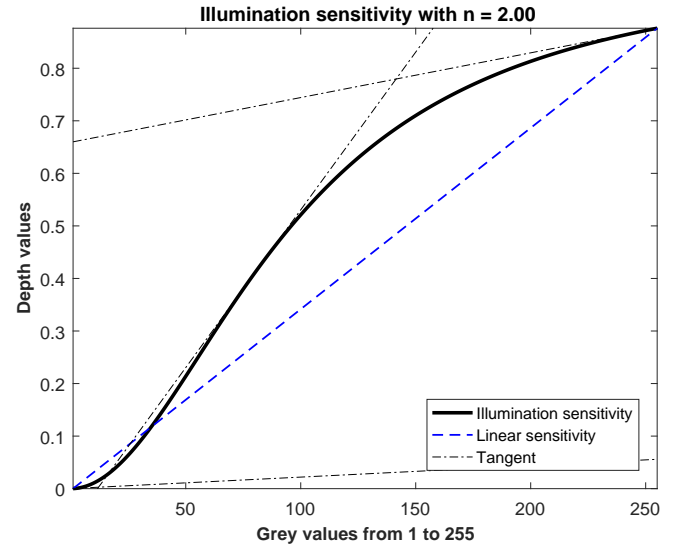The shape of the function is illustrated in Figure 4.

Figure 4: The black bold line display the corrected illumination sensitivity. The blue dashed line is the original linear sensitivity. The dashdot lines represent the tangent lines for the high, mid and low grey values.

The graph clearly shows the saturation effects at high[1]

---

[1] What happens when you look into the sun?

and low[2] luminance values and an approximated linear behaviour in the mid range. However, humans are still great at distinguishing objects under a wide variety of illumination circumstances. This is partly caused by adapting the sensitivity of the photoreceptor. Therefore the user has the option to locally adapt σ in Equation 10 to the luminance map, instead of the summation over the complete map. This is recommended for photos with large variation in the luminance map. An extended review about the illumination sensitivity can be found in Appendix C.

The SFL reconstruction, 5vii, is constructed by transforming the selected luminance signal, 5vi, with Equation 9. It should be emphasized that the depth map is expressed in relative values between 0 and 1,

---

[2] Or a dark room?

(i) The original RGB image

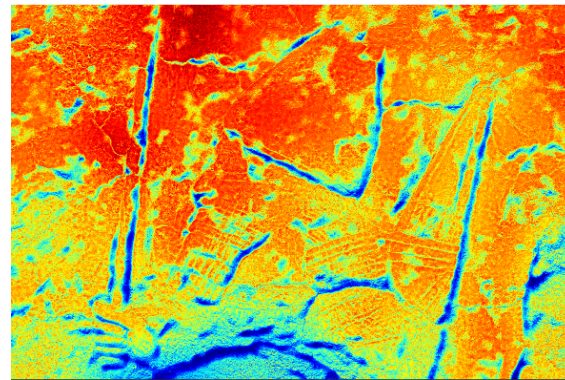(ii) The luminance signal
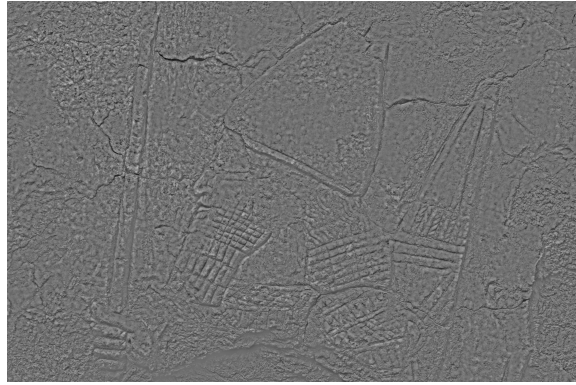
(iii) The first IMF
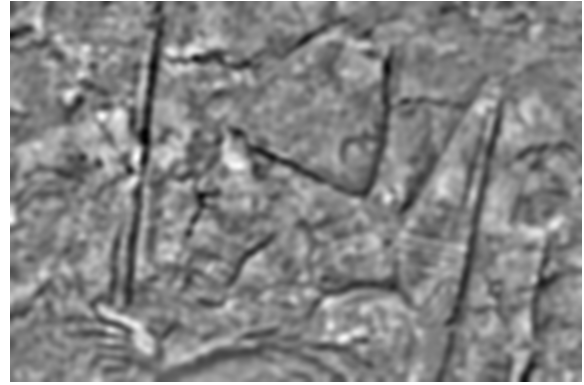
(iv) The second IMF

(v) The residual of the luminance
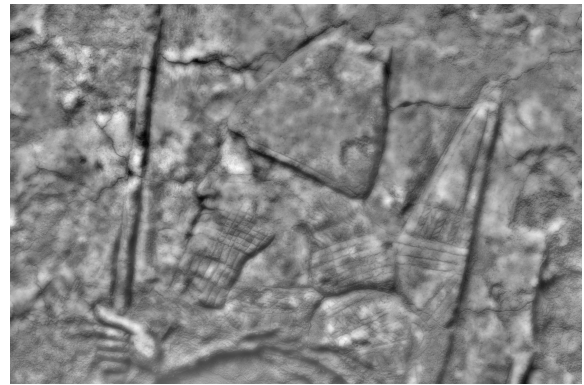
(vi) The weighted combination of the IMF's

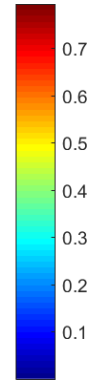
(vii) Depth map retrieved with SFL

Figure 5: An overview of the SFL method and the different process steps, with **i** being the **input**, **ii** is the result the **Extraction** step, **iii - v** of the **Decomposition** with the EMD, **vi** the **Selection** and **vii** is the depth **Reconstruction**.

ii Artificial Reconstruction

Although the Shape from Luminance algorithm is able to capture the fine details from the reliefs, the luminance reconstruction is noisy for large and flat surfaces. The artificial reconstruction counteracts this problem. The idea is to recreate the concept behind a relief: the exact depth values are not important within the reliefs, but being able to differentiate between the figures and the textural details is. The artificial reconstruction is done in 3 steps:

1. Preparation
2. Segmentation
3. Elevation

Preparation, the first step is to compute the gradient as preparation for the image segmentation. The segmentation is important since the line drawings, (Figure 7ii) do not define closed contours. These closed contours are necessary for the elevation. We start with calculating the Euclidean distance (Figure 7iii) for every pixel to the closest line: $D_{linear}(x, y)$. Next, the euclidean distance is converted into greyscale between 0 and 1 and inverted to obtain $D_{inverted}(x, y)$, Figure 7iv. This makes the watershed result more natural as shown in Appendix D.

Segmentation. The segmentation process is conducted with the Watershed algorithm of Meyer [22]. The benefit of using the Watershed transform is that it provides us with closed contours [3]. The name of the algorithm already reveals its working principle. Let us consider the distance map as a topographic map, which will be flooded. The local minima are used as sources of this flood. The flood occurs with a constant vertical speed. When two distinct basements merge, a dam is erected to prevent mixing (Watershed lines) [22]. Meyer warns that the local minima are often too numerous and lead to oversegmentation. In our pictures we observed some oversegmentation. This is visible as several sets of lines appointed by the black circle in Figure 7v. Instead of reducing the number of minima, we chose to merge regions smaller than certain amount of pixels with their smallest neighbour. The merging process is repeated until every region complies with the minimum size requirement. This makes it more obvious for the user which parts are merged in the process. The results are shown in Figure 7vi.

Elevation. The last step is to elevate the segmented regions. Every region in the Watershed result corresponds with a positive integer and colour, which serves as a label. The zeros are reserved for the Watershed lines. The user defines which regions have the same height level by assigning a uniform colour to every elevation level. In Figure 7vii, we have defined green as the upper layer and red as background. Next, the Euclidean distance for every pixel to the contour lines (where different colours meet) is computed. This distance function is used for the step transformation.

Evaluation of 3D scanned relief led to the conclusion that there are 2 typical transformations: *Step* and *Carving* as shown in Figure 6. As artificial counterpart we used the Tangens Hyperbolicus ($tanh$) to imitate both the step function:

$$Z_{step}(x, y) = \frac{A_1}{2} * tanh\left(\frac{D_{linear}(x, y)}{W}\right) + \frac{A_1}{2} \quad (11)$$

as the carving function:

$$Z_{carving}(x, y) = \frac{A_2}{2} * tanh\left(\frac{|D_{linear}(x, y)|}{W} + B\right) - \frac{A_2}{2} \quad (12)$$

Where $D_{linear}(x, y)$ is a function of pixels x and y, is considered to be negative for the lower elevation level. A is the desired amplitude of the step or carve, W influences the width of the transition. B shifts the function along the XY plane, influencing the sharpness of the carve around $D_{linear} = 0$. As you might have noticed, $A/2$ is added in the step function and subtracted in the carving function. This ensures that respectively only material is added (completely positive in z-direction) or removed (negative).
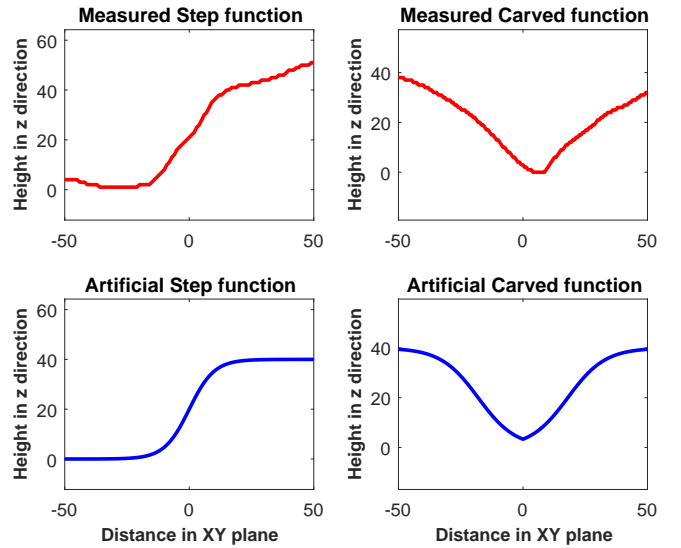

Figure 6: The top 2 plots are an example of a measured Step and Carving function from the 3D scans. The bottom 2 figures are the artificial functions.

The artificial reconstruction is built by stacking up the elevation layer with the carved layer, resulting in Figure 7viii. With the amplitude variable, A, in Equations 11 and 12, the user can add the necessary units to the depth values. In contrast with the unit-less(only relative values between 0 and 1) SFL depth map, this will be resolved in the next section: **Data Fusion**.
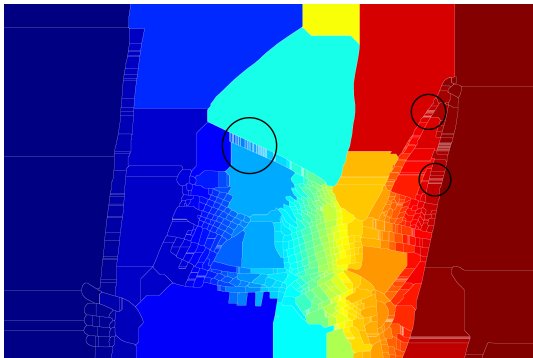
(i) The original RGB image
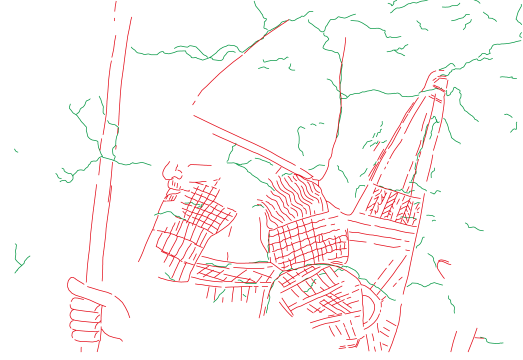
(ii) The line drawing. Red are the lines and Green the large cracks.
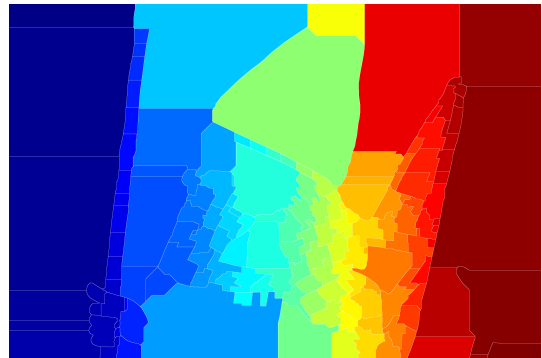
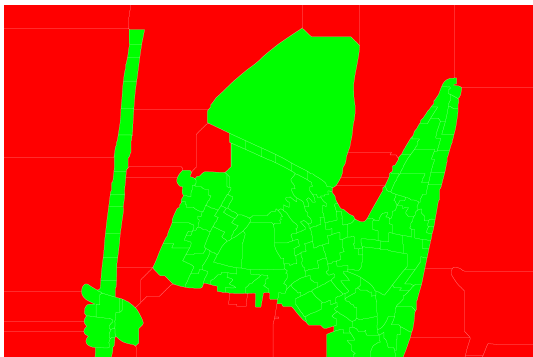(iii) The euclidean distance per pixel, where black is the closest

(iv) The inverted Euclidean distance per pixel, where white is the closest to the lines.

(v) The original Watershed result. The different colours represent the labels for each region.

(vi) The fused Watershed Result

(vii) The user defined height map. Green is the top level and red is the background.

(viii) The depth map by artificial reconstruction, with red corresponds with the highest value and blue with the lowest.

Figure 7: An overview of the artificial reconstruction and results of the different steps: where **i** is the photo used by the RMO, **ii** is the line drawing made by the experts, **iii** and **iv** are the **Preparation** steps for the **Segmentation** results in **v** and the fused Watershed, **vi**. The last step, **Elevation**, combines **vii** with **ii** to create the artificial reconstruction **viii**

iii Data Fusion

The fusion of data combines the textural details of the Shape from Luminance algorithm with the smoothness of the artificially reconstruction. This is realised by using the weighted sum per pixel between the two different depth maps. The computation of the weighted sum is simple. However, it is important to determine the correct weight per pixel for both inputs.

$$\text{D}_{\text{fused}} = \frac{W_{\text{SFL}}\text{D}_{\text{SFL E}} + W_{\text{Artificial}}\text{D}_{\text{Artificial}}}{W_{\text{SFL}} + W_{\text{Artificial}}} \tag{13}$$

We reformulate this problem into locating information in the image, where we define information as tubular structures (lines). Frangi [11] experienced the same problem with his research in multiscale vessel enhancement filtering. It does not require much imagination to see the similarities between vessels and the lines in an image. Therefore we locate the information by using the Hessian based filter by Frangi. By computing the eigenvalues of the Hessian, it is possible to determine the orientation patterns shown in Table 1 and Figure 8. Categorically we divide the Data Fusion into 4 parts:

1. Convolution
2. Analysis
3. Measure
4. Fusion

Where the first 3 parts lead to the construction of the weights per pixel and the last step is the actual data fusion.

Convolution, it is important to use a multiscale approach, since we do not know the size of the tubular structures. Florack et al. [9] and Koenderink [16] used a convolution with derivatives of Gaussians as differentiation, which contains scale factor σ:

$$\frac{\partial}{\partial x_i}L(x_i,\sigma) = \sigma^\gamma L(x_i) * \frac{\partial}{\partial x_i}G(x_i,\sigma) \tag{14}$$

where the D-dimensional Gaussian is defined as:

$$G(x_i,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}^D}\ e^{-\frac{||x_i||^2}{2\sigma^2}} \tag{15}$$

The amplitude along a multiscale derivative decreases in general with the scale. The parameter γ was introduced by Lindeberg [20] to normalize this effect, so the derivatives can be compared. When there is no scale preference, γ is set to 1.

Now we have a multiscale derivative, the Hessian is computed, which is the 2nd order derivative:

$$\frac{\partial}{\partial x_i^2}L(x_i,\sigma) = \sigma^{2\gamma} L(x_i) * \frac{\partial}{\partial x_i^2}G(x_i,\sigma) \tag{16}$$

In our 2 Dimensional case the Hessian becomes:

$$\mathcal{H}_{x,y,\sigma} = \sigma^2\ L(x,y)\ *\begin{bmatrix}\frac{\partial^2}{\partial x^2} & \frac{\partial^2}{\partial x\partial y}\\ \frac{\partial^2}{\partial y\partial x} & \frac{\partial^2}{\partial y^2}\end{bmatrix}G(x,y,\sigma) \tag{17}$$

Analyse, the Hessian is a 2 by 2 matrix where we obtain the eigenvalues by solving the 2nd order polynomial for λ. The eigenvalues of the Hessian are obtained by calculating the following determinant and stating that v is a non-zero vector:

$$\begin{aligned}\mathcal{H}v &= \lambda v\\ (\mathcal{H}-\lambda I)v &= 0\\ \det(\mathcal{H}-\lambda I) &= 0\end{aligned} \tag{18}$$

with:

$$I = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

Where we sort the eigenvalues, λ, according to the following condition:

$$|\lambda_1| \le |\lambda_2| \tag{19}$$

With the eigenvalues of the Hessian, we analyse the local orientation pattern as stated in Table 1 and illustrated in Figure 8. We basically distinguish between 3 different orientation patterns: noise, tubular and blobs. The sign of λ_2 determines whether the orientation is bright or dark coloured compared to its environment. Visual inspection of the SFL depth map shows that the carvings appear as dark tubular structures. The next step is to construct a measure to filter these structures, which we regard as information.

Table 1: The different possibilities for the eigenvalues and the corresponding orientation patterns

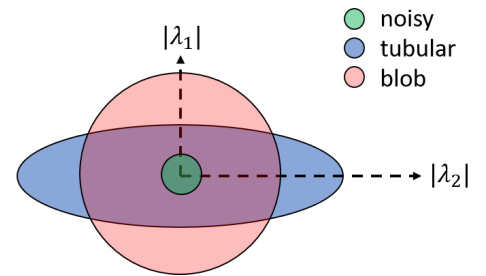| λ_1	λ_2	orientation pattern
Low	Low	noisy, no preferred direction
Low	High-	tubular structure (bright)
Low	High+	tubular structure (dark)
High-	High-	blob-like structure (bright)
High+	High+	blob-like structure (dark)

Figure 8: A visualisation of the different orientation patterns regarding the eigenvalues.

Measure, instead of evaluating each eigenvalue individually, Frangi [11] has introduced a measure which combines the eigenvalues into a single measure:

$$\mathcal{V}(x,y,\sigma) = \begin{cases}0 & \text{if } \lambda_2 > 0\\ \exp\left(-\frac{\mathcal{R}_\beta^2}{2\beta^2}\right)\left(1-\exp\left(-\frac{S^2}{2c^2}\right)\right) & \end{cases} \tag{20}$$

The measure contains two parameter, β and c , which are used to tune the the measure.

Threshold β adjusts the sensitivity of the "Vesselness" measure (Equation 20) to blob shaped orientation patterns. The blobness of a structure is defined with $\mathcal{R}_\beta$ (Equation 21), which is 1 for a perfect blob (circle) and becomes smaller for a more tubular structure.

$$\mathcal{R}_\beta = \frac{\lambda_1}{\lambda_2} \qquad (21)$$

Parameter c is used to tune the sensitivity of the "Vesselness" measure to the magnitude of the eigenvalues, λ_1 and λ_2. The magnitude of the eigenvalues of the Hessian is defined as the structureness (Equation 22). With parameter c and the structureness, noise can be filtered from the "Vesselness" response.

$$S = \sqrt{\sum_{j \leq D} \lambda_j^2} \qquad (22)$$

Figure 9 illustrates the nature of Equation 20.In Appendix E several heatmaps are shown to visualize the effect of thresholds β and c.
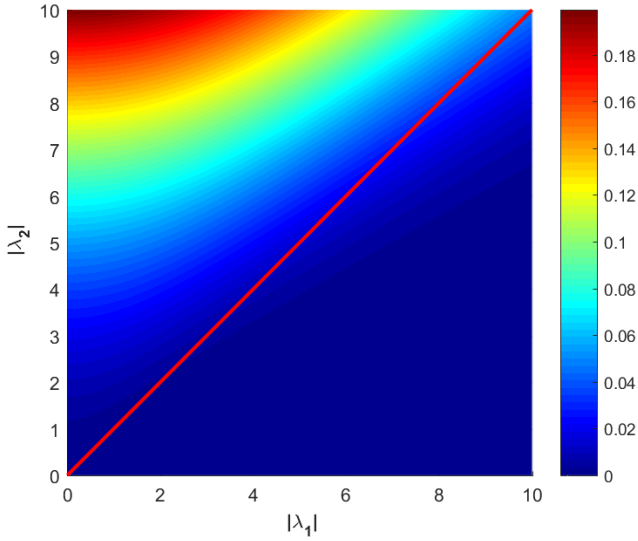
Figure 9: A heatmap visualizing the "Vesselness" score. We only review the values left of the red line, because of the criterion: $|\lambda_1| \leq |\lambda_2|$

In the first step, **convolution**, we discussed the importance of a multiscale approach. Therefore, the vesselness, $\mathcal{V}$, is calculated along the pre-set range of σ. After which the maximum score per pixel is selected:

$$\mathcal{V}(x,y) = \max_{\sigma_{\min} < \sigma < \sigma_{\max}} V(x,y,\sigma) \qquad (23)$$

The final weights for W_{SFL} and $W_{\text{Artificial}}$ are calculated with:

$$W_{\text{SFL}} = (1 - W_{\text{mix}}) \frac{\mathcal{V}(x,y) - \min(\mathcal{V})}{\max(\mathcal{V}) - \min(\mathcal{V})} + W_{\text{mix}} \qquad (24)$$

$$W_{\text{Artificial}} = 1 - W_{\text{SFL}}$$

where W_{mixed} allows the user to mix both depth maps with a certain offset.

Fusion, given that the weights are constructed, the actual fusion can be performed. Recall that the SFL depth map contains only relative values between 0 and 1, where the artificial depth map has a unit of length. Moreover, the artificial depth map has an elevated base. To counteract these problems, we multiply the SFL map with the same amplitude, A_1 as used in Equation 11, after which we remove the mean and add artificial reconstruction to the results:

$$\begin{aligned} D_{\text{SFL1}} &= A_1 * D_{\text{SFL}} \\ D_{\text{SFL2}} &= D_{\text{SFL1}} - \text{mean}(D_{\text{SFL1}}) \qquad (25) \\ D_{\text{SFL E}} &= D_{\text{SFL2}} + D_{\text{Artificial}} \end{aligned}$$
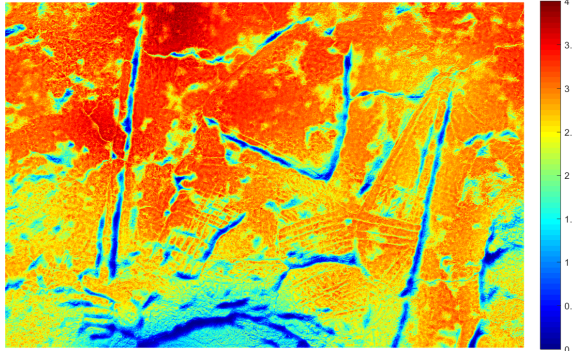
The addition is needed to ensure that both depth maps are in the same order of magnitude and the mean is removed to prevent any translation bias. At last, we fuse both depth maps with Equation 13.
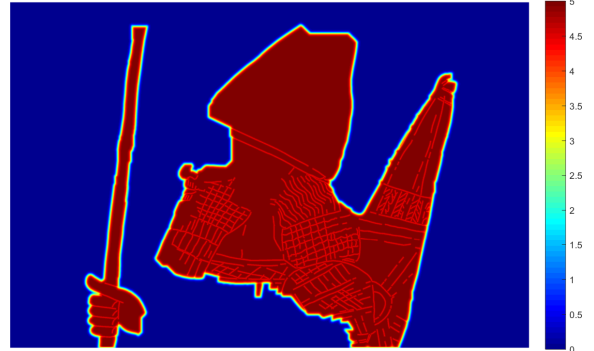
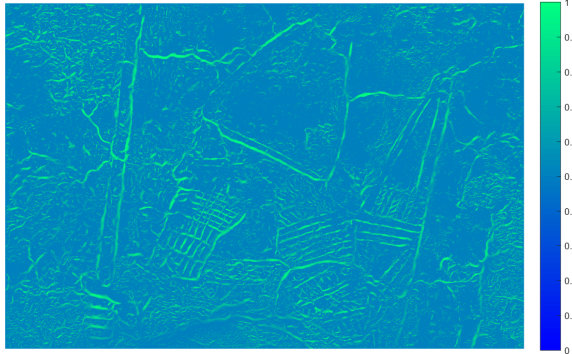(i) The SFL depth map after elevation
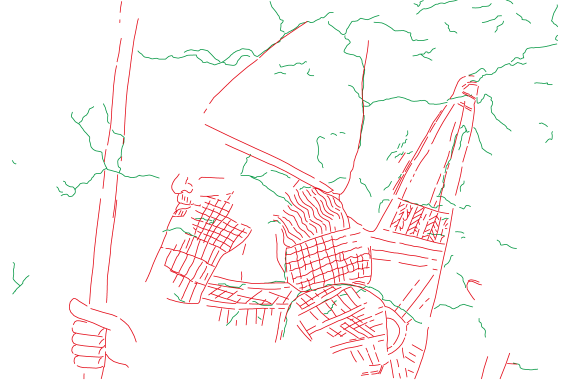
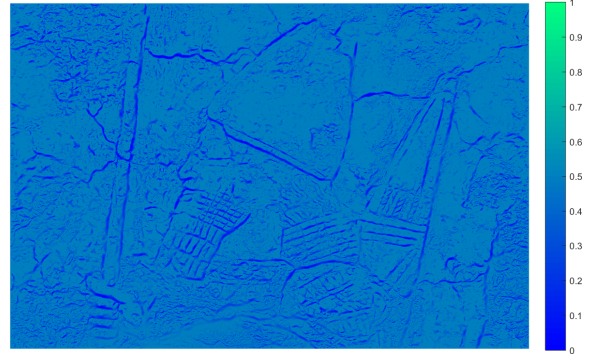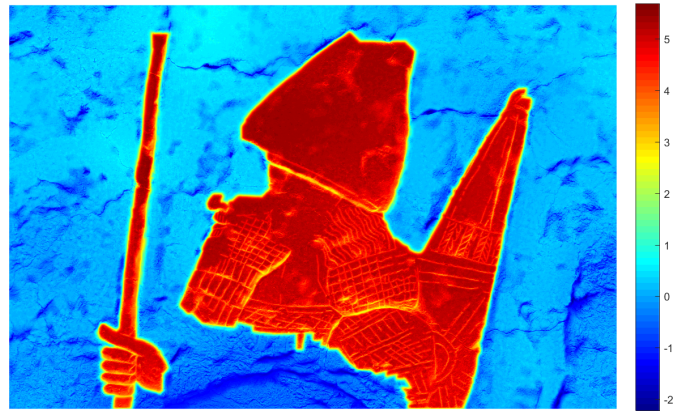
(ii) The line drawings

(iii) SFL depth map

(iv) Artificial depth map

(v) Confidence map for SFL. Green is the highest confidence value and blue the lowest.

(vi) Confidence map for the Artificial depth map.

(vii) Fusion result. Red is the largest depth value in mm and blue the lowest.

Figure 10: An overview of the Data Fusion method and the different process steps, **i** and **ii** are the original inputs. **iii** and **iv** are the depth maps respectively obtained with SFL and the line drawings. **vii** is the fused depth map.

III Evaluation

i Method

In the previous sections we established a monocular 3D reconstruction method. The next step is to verify our approach and set the parameters of the algorithm. Therefore, we need an evaluation method to compare our results with the 3D scan from the museum, which serves as ground truth. The simplest way would be to compute the difference in depth for each pixel. However, the base of our reconstruction differs substantially from the 3D scan, namely, the 3D scan contains a multi-directional drift in the relief base visible in Figure 12b. While in our reconstruction, we made the assumption that the base is flat (Figure 12a and 12c). Simply subtracting the depth values from each other will result in a large error, which does not represent the quality of the reproduction.

Furthermore, the SFL algorithm is not capable of obtaining the exact depth values, since we use uncalibrated monocular images only. It is therefore more logical to evaluate the shape instead of the exact depth values. Again, the eigenvalues of the Hessian are used. As shown in Table 1 the eigenvalues of the Hessian represent different orientation patterns. In Section iii we only looked for dark tubular structures, which accordingly corresponded to information in the pictures. For the evaluation we use all the mentioned orientation patterns. The evaluation method contains 3 different elements:

1. Align
2. Convert
3. Compare

Align, the first challenge is to align the RGB photo with the 3D scan. Both the RGB photo and 3D scan are obtained with different devices, introducing various distortions in their final output. Even though most modern devices correct for lens distortions, the RGB image and 3D scan do not align appropriately. This creates the need for alignment. Unfortunately the output images differ too much for an automatic outline procedure, therefore we chose a manual alignment method.

Since all reconstructions are based on the RGB picture, we chose to align the 3D scan with the RGB image. We manually selected **59** control points in the *cpselect* environment of Matlab with the prediction mode on.
Convert, the conversion from depth map to eigenvalues is done in the same way as discussed in Section iii. The only difference is that we use $S(x, y)$ as measure to evaluate the shape:

$$S(x, y) = \lambda_1 + \lambda_2 \qquad (26)$$

with $S(x, y)$ being normalized between -1 and +1:

$$S(x, y) = \begin{cases} \frac{S(x,y)}{|\min S(x,y)|} & \text{if } S(x, y) < 0 \\ \frac{S(x,y)}{\max S(x,y)} & \text{if } S(x, y) > 0 \end{cases} \qquad (27)$$

This measure does not discriminate between blobs and tubular structures, but is rather a measure for the curvature in terms of convexity and concavity, where **-1** is the most convex shape in the depth map and **+1** the most concave.

Compare, we compare the reconstructions with the ground truth (3D scan), once every pixel is transformed with the surface measure of Equation 26. The comparison is done with the root mean square error (**RMSE**):

$$\text{RMSE}(x, y) = \sqrt{\left(\hat{S}(x, y) - S(x, y) \right)^2} \qquad (28)$$

Where $S(x, y)$ is the ground truth and $\hat{S}(x, y)$ is the predicted value with our reconstruction method. The RMSE is displayed in a visual heat map (FIGURE) to show the quality of the reconstruction based on its location. To evaluate the overall quality of a reconstruction we use the RMSE over all pixels:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n \left(\hat{S}(x, y) - S(x, y) \right)^2}{n}} \qquad (29)$$

ii Implementation

The RMSE is computed over multiple resolutions. In our case, $\sigma = 5$ and $\sigma = 10$ are chosen to evaluate the RMSE value, where $\sigma = 5$ distinguishes the details, while $\sigma = 10$ is dominant for the larger structures. For lower σ-values, the random noise of the stone texture will be too prevailing and with larger σ-values the sculptures are too blurred.

The 3D scan of the relief at the RMO was made with the ATOS III scanner and has an accuracy of 10 microns [1]. However, the photo of the RMO slab has image size of 5 MP with a bit depth of 24. The slab is 480 mm long, which gives a resolution 0.19 mm per pixel (or 135 DPI). The 3D scan was resized with a bicubic interpolation to match the photo size. The resized 3D scan (Figure 12b) was used as ground truth for our evaluation method.

The Gaussian convolution in Equation 14 causes the image to warp at the edges. Therefore, the image is cropped with 100 pixels from each border to select the unwarped data.

The depth maps are divided into 5 groups in Table 2. In the first group the optimal SFL combination was determined. Second, the Artificial reconstruction and the

Elevated SFL depth map (Equation 25) were reviewed, which are both inputs for the Fusion map. Third, the fusion result was evaluated. If there is no best result for both σ levels, then the result for $\sigma = 5$ has priority, since the details are most important.

Furthermore, two control groups are added. The first control group consists of the grey image and a flat plane. The grey image would be the result of applying the SFL algorithm without the EMD and the plane resembles doing nothing. The second control group consist of the ground truth (GT) itself, which must be 0. The GT + 1° rotation test the evaluation method on misalignment. For the GT + Random noise we added random values between 0 and 15 to every pixel of the GT. Table 2 is only a part of the combinations tested, the complete overview can be found in Appendix F.

iii Results

The output of the EMD for the relief at the RMO are shown in Figure 11. The outputs consists of 3 IMF's and a residual.

The RMSE has maximum value of 1 and would be 0 in the perfect case. The lowest error, *SFL Elevated*, is 41.6% lower than the *Grey* image for the details, $\sigma = 5$, and 44.4% lower for $\sigma = 10$. *SFL Elevated* also scores 23.2% lower than *Fusion* for $\sigma = 5$ and 27.3% lower for $\sigma = 10$.

Table 2: The RMSE over all pixels, $N_{\text{pixels}} = 4291872$. A lower RMSE is a better result. The lowest value defined in **bold**.

| Type of Depth map	RMSE$_{\sigma=5}$	RMSE$_{\sigma=10}$
IMF1	0.194	0.205
IMF2	0.270	0.195
IMF3	0.615	0.675
IMF1 + IMF2	0.187	0.188
IMF1 + 2 x IMF2	0.162	0.162
IMF1 + 2 x IMF2 + IMF3	0.152	0.165
Artificial Reconstruction	0.199	0.196
SFL Elevated	**0.149**	**0.149**
Fusion	0.194	0.205
Grey	0.255	0.268
Plane	0.797	0.868
GT	0	0
GT + 1° rotation	0.127	0.134
GT + Random Noise	0.133	0.134

(a) An overview photo of the relief, which shows the shadows. (b) The luminance input of the RMO relief for the EMD
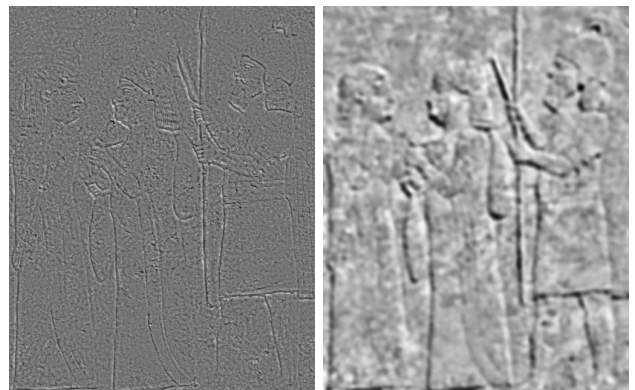
(c) IMF1

(d) IMF2
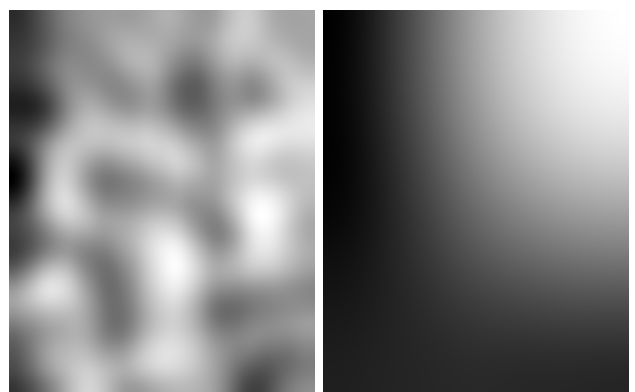
(e) IMF3

(f) The residual

Figure 11: An overview of the EMD results of the RMO relief. Notice that the illumination drift in the residual corresponds with the direction of the observed shadows.
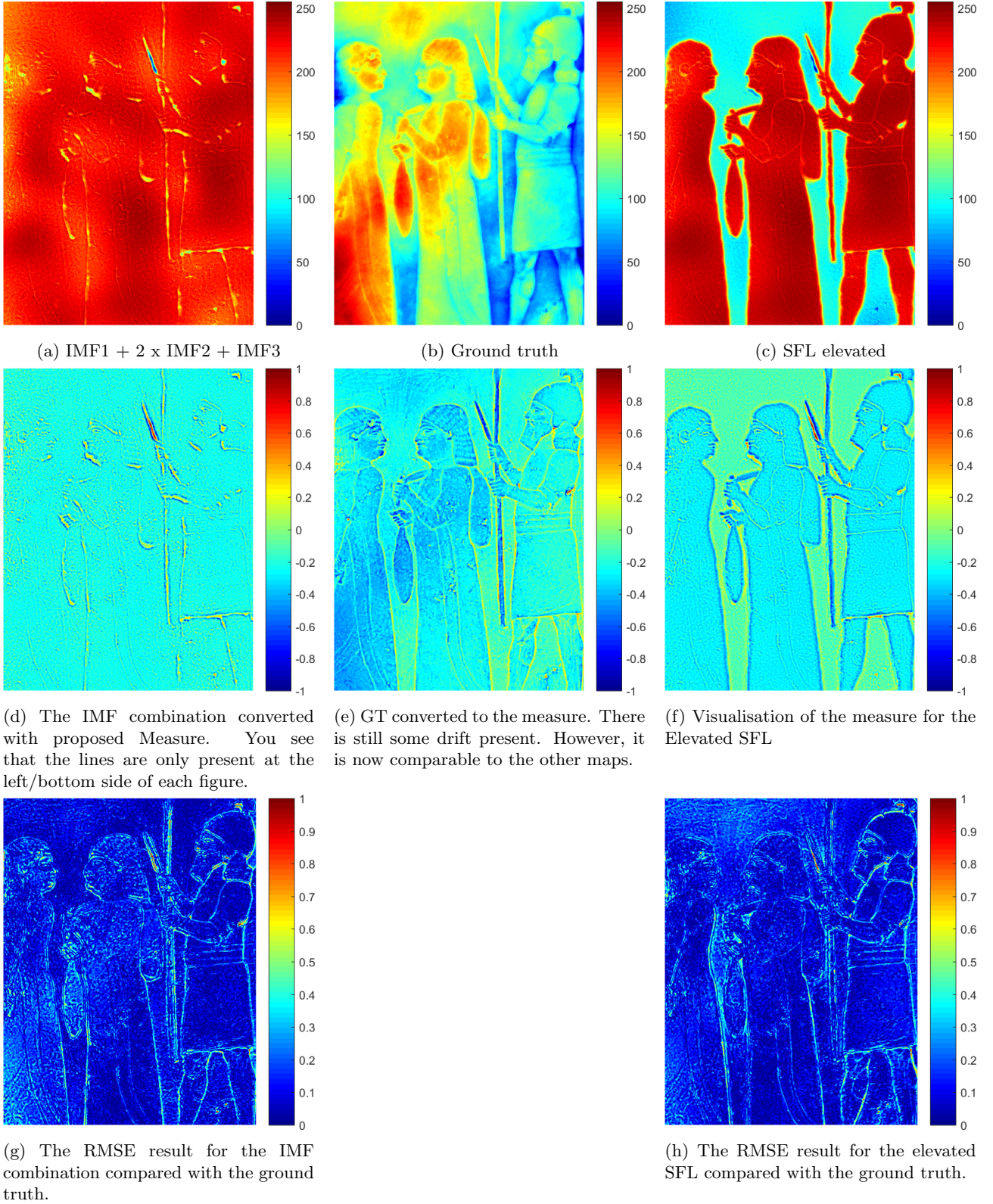
(a) IMF1 + 2 x IMF2 + IMF3

(b) Ground truth

(c) SFL elevated

(d) The IMF combination converted with proposed Measure. You see that the lines are only present at the left/bottom side of each figure.

(e) GT converted to the measure. There is still some drift present. However, it is now comparable to the other maps.

(f) Visualisation of the measure for the Elevated SFL

(g) The RMSE result for the IMF combination compared with the ground truth.

(h) The RMSE result for the elevated SFL compared with the ground truth.

Figure 12: **a** to **c** are the original depth maps, which are not comparable due to the differences in the base. The scale is given in gray scale intensity from 0 - 255. The next row shows the transformed depth maps with our proposed measure, where 1 is the most convex structure in the image and -1 the most concave. The last row shows the RMSE per pixel in the form of a heat map. Where red corresponds with the largest error.

IV Reproduction

The ultimate goal is not only to have a digital reconstruction, but to have a physical reproduction as well. In this research we selected 3 different reproduction techniques:

- CNC milling in limestone.
- Full colour sandstone 3D printing of **Z corporation**.
- Elevated printing technology by **Océ**.

CNC milling, the original reliefs have been fabricated in limestone with a set of carving tools. Nowadays, this process is automated with computer numerical control (CNC) milling. Milling is a machining process of using rotary cutters to remove material [4]. The benefit of using CNC milling is the durability and strength of the material. Also, the material texture will be the closest to the original, since we make use of the same sedimentary rock, limestone. A disadvantage of milling is the resolution. For practical reasons, a ball end mill with a diameter of 1.5 mm is used. However, if the stepsize is within the range of 0.1 mm, the resolution for concave shapes is still limited to diameter of the mill, 1.5 mm. This causes the mill to act as a mechanical filter. Another limitation to CNC milling is the absence of added colour to the stone. The process of ageing causes a heterogeneous stone colour, while the newly excavated stone will be uniformly coloured, as shown in Figure 13b. To recreate the authentic look, the milled object will need post-processing as for example airbrushing.

Full colour sandstone, this 3D printing techniques deposits a binder onto a bed of gypsum powder to create the object layer by layer. Simultaneously, a colour jet injects the CMYK ink to add colour to the layers. The inkjet printhead has a resolution around 300 DPI. In practice this will be approximately 50 DPI, because of the absorption of ink by the gypsum powder [27]. Additionally, the combination of powder and binder also results in a brittle object. The model is coated in a cyanoacrylate solution (superglue) after printing, to strengthen the model and enrich the colours. Another disadvantage is the size limitation. The Zprinter 650 of Zcorps is only capable of producing objects within the bounding box of 254 x 381 x 203 mm. Which means that the reliefs
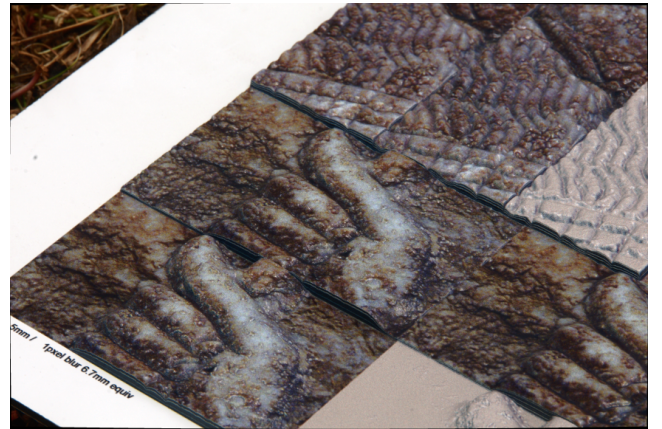

(a) Original input RGB photo

(b) CNC milling

(c) Render of the Sandstone 3D print

(d) Elevated printing technology samples to choose the correct parameters.

Figure 13: The original input photo, **a**, and the corresponding repoductions, **b**, **c** and **d**.

need to be subdivided into smaller parts, introducing partition lines in the final reproduction. Nonetheless, the full colour sandstone 3D printing is the only technology which offers full colour, while maintaining the stone-like texture.

Elevated printing technology [24] [25], although public technical details are limited, the technology can be explained as printing numerous layers of ink onto a surface to create a 3D object. As a result of using ink as deposit material, the resolution is within the range of microns. Therefore, this technology by Océ is able to accurately capture the details in full colour. However, the final reproduction will have a plastic gloss and touch, losing the stone texture of the reliefs. Another benefit of this 3D printing technology is its printing bed of more than $4m^2$.

V Discussion and Conclusion

The challenge was to make reproductions of Assyrian reliefs of the Southwest Palace of Sennacherib in Niniveh from monocular images In the first part of this paper we introduced a method for reconstructing reliefs from monocular images. Next,the method was evaluated with a relief from the RMO. Our reconstruction method had 44.4% lower RMSE for the shape of the relief. The shape was determined by the eigenvalues of the Hessian. In the last part several reproductions were made of the lost reliefs with 3 different production techniques. Each reproduction has their own strength. However, they all have the capability to reproduce the key-feature of Assyrian reliefs: the lively details.

The results in Figure 5vii and 12a show a low elevation for Shape from Luminance. This also expresses the need for the elevation of the SFL reconstructions with an Artificial reconstruction. The lack of elevation is explained by the distance between the camera and the object. From the camera data we deduced that the object distance is between the $1 - 1.5m$. Compare this to a maximum elevation of $5mm$ in the reliefs and it becomes obvious that this information disappears with a frontal photo.

Furthermore, if we visually compare the photo at Niniveh, Figure 5vii and the photo at RMO,12a, the Niniveh photo contains more details than 12a. This is probably caused by the difference in light exposure. Figure 5vii is only illuminated by the sun, while Figure 12a is displayed in room with multiple light sources and a spotlight aimed on the relief from the upper right corner as shown in Figure 11. This is a strong violation of the assumption of a unique light source, placed far away from the scene. The difference becomes more evident, if the residual of the relief at RMO, Figure 11f, is compared with the residual of 5v.

The EMD is expected to compensate for the illumination drift. Nevertheless, the illumination drift affects the quality of the reconstruction. Unfortunately we did not have the resources to research the magnitude of this effect. The compensation of the illumination drift is visible in the number of modes retrieved with the EMD. Where the Niniveh photos have been decomposed into 2 IMF's and a residual, the RMO photo has 3 IMF's and the residual. There is a significant similarity between the Niniveh residual, Figure 5v, and RMO, Figure [RES RMO]. So, the EMD is capable of compensating the illumination bias, but it fails to remove the strong casted shadow.

We should be careful with extending the tuned parameters of the evaluation method to the photos of Niniveh. However, in the absence of better ground truths, we stick to our findings of the relief at the RMO. Until parts of the reliefs at the site are recovered, which could serve as ground truth.

The line drawings are at the base of the artificial reconstruction and are manually drawn. Therefore, the line drawings are subjective to the human perception. Although the perception of art could be a philosophical question, the line drawings sometimes contain small mistakes, Appendix D.

Another problem is the miss-segmentation of the watershed algorithm. The segmentation is used to create closed contours. On the other hand, it produces sometimes unnatural boundaries: weird angles or surprisingly straight lines. This effect occurs mostly around areas where the gaps between the contour lines are too large. These gaps are often found where the stone has been deteriorated, causing a gradual change in height, instead of a sharp line. We did not observe any large miss-segmentations in the reconstructions. Most of the miss-segmentation is prevented by inverting the Euclidean distance. The view miss-segmentations left over, are compensated by the SFL depth map.

According to the results in Table 2, there is no need for Data Fusion. There is an important difference between the relief at RMO and Niniveh, namely the damage of the stones. The relief at RMO does not contain much abrasion to the stone surface. This abrasion in the Niniveh photos is responsible for the high frequent noise in the SFL reconstruction, Figure 5vii. For this reason, data fusion is necessary. However, it needs to be manually tuned by preference.

The evaluation method in this research compares the eigenvalues of the Hessian of the reconstructed depth map with the 3D scanned ground truth. The eigenvalues are a common way to evaluate orientation patterns and

give an intuitive representation of the shape. In the literature it is mentioned that noise is experienced during 3D scanning of certain materials [19]. However, the magnitude of this noise is not determined and often the theoretical accuracy of the 3D scanner is used. In our research the theoretical accuracy is substantially larger than the accuracy of our reconstruction. Furthermore, the GT with severe noise still performs 11% better than our best reconstruction, which means that the evaluation method is robust against noise.

David et al. [8] use the RMSE to evaluate their 3D results, which works well if the reconstruction and ground truth have the same base or reference points. However, small alignment problems or a global drift in the background (as we experienced), will introduce large errors. We tested our method to misalignment by rotating the GT with 1°. The introduced error is still small compared to our best reconstruction. Moreover, if there is error due to misalignment, it will be constant among all reconstructions, since all reconstructions have the same base.

Elkhuizen et al. [7] used a board of experts to visually inspect the reproductions. We deliberately chose to evaluate the digital reconstructions, instead of the physical reproductions, to create an objective evaluation method. This would reduce the costs, but it would mostly isolate possible errors to our reconstruction method.

In our evaluation method we have to align the images to compensate for the non-linear camera lens distortion. In practice, we can not align the photos of Niniveh with a ground truth. Even more, the calibration data of the photos have been lost due to lack of backwards compatibility, which rules out the common ways of removing the lens distortion. For now, we accept that the reconstructions will contain lens distortions.

The most burdensome part of the research was the database. Only a small portion of the database could be used for reconstruction, due to the natural wear and fire. Though, portions of the sculptures could be identified and completed by using examples of other Assyrian reliefs, it is beyond our scope to make an educated guess of what the reliefs used to look like.

Our goal was to reconstruct the 3D shape of Assyrian bas-reliefs from the region of Niniveh from monocular 2D images. The results show that even when the originals are lost, reconstructions could be made from pictures only. Furthermore, we learned that results are vastly dependent on the picture quality: extensive tuning was needed when using 15 years old photographs. However, our method produces highly detailed and realistic reproductions.

VI   Future Research

The immediate result of this research is the usage of the artificial reconstruction method by the RMO to speed up their reconstruction work. The idea of mapping lines to changes in the height of the reliefs saves a lot of time and manual labour. Furthermore, it would great to expand and apply this research to online databases of lost cultural heritage like Rekrei [29] and the Million Image Database [10].

The results in this research could be improved by experimenting with markers for the watershed algorithm. Now the markers are automatically set at the minima of the inverted Euclidean distance, which results in numerous segmented areas caused by local minima. Currently, the small areas are automatically fused with their closed neighbour. However, by manually applying markers, the user can control the "sources" of the flooding and reduce the number segmentation areas. This could potentially save time in creating the user defined height map and improve the accuracy.

Some of the user parameters are scale related. They are currently manually tweaked. However, we could research the relation between the parameters and the absolute size of the image to reduce number of user parameters, σ (Equation 14), β and c (Equation 20) for example are all vastly dependent on the scale.

The line drawings are user dependant and time consuming to make. Attempts to automate the line drawing process did not succeed due to the noise present in the current database. Nevertheless, the Hessian based Frangi filter provides a distinctive signature image. This image, even with the noise, could be used to recognize the object within the image and use other databases with "clean" data to auto-fill the missing information or remove the noise, like the new application Autodraw of Google [23]. As mentioned before, Assyrian bas-reliefs are distributed among several musea. Although the different reliefs display numerous stories, the style is very comparable. This makes it possible to build a clean database.

Once a clean database is obtained, the possibility of applying neural networks becomes available as well. Zheng et al. [32] suggested the use of conditional random fields ad recurrent neural networks for pixel-level labelling tasks, such as semantic segmentation. This could replace the need of the watershed algorithm and user defined height map. Simonyan et al. [28] use a convolution network to create a saliency map. The produced saliency map could be an alternative for our Hessian based filter in Section iii: Data Fusion.

An application outside the field of cultural heritage restoration could be texture mapping for digital models.

For example, a brick wall can be easily modelled in any CAD program, but modelling the texture will be a different job. With our method it would be possible to project textures of real world objects onto digital models at the click of a button.

More, the combination of an ordinary 3D scanner and the SFL algorithm should be investigated as well. The 3D scanner would then replace the role of the artificial reconstruction in our research. In this way the level of details could be improved, the photo sensor would upsample the low-resolution 3D scan.

Acknowledgements

References

[1] Capture 3D. Ge high accuracy, high throughput inspection technologies challenge winner!, 2017.

[2] Jess Ahmon. The application of short-range 3d laser scanning for archaeological replica production: The egyptian tomb of seti i. *The Photogrammetric Record*, 19(106):111–127, 2004.

[3] Serge Beucher and Fernand Meyer. The morphological approach to segmentation: the watershed transformation. *OPTICAL ENGINEERING-NEW YORK-MARCEL DEKKER INCORPORATED-*, 34:433–433, 1992.

[4] R. I.) Brown & Sharpe Manufacturing Company (Providence. *Practical Treatise on Milling and Milling Machines*. Brown & Sharpe Manufacturing Company, 1914.

[5] Paul Collins. *Assyrian Palace Sculptures*. 2008.

[6] James E Cutting. Potency, layout and knowing distances: The integration, relative, potency, and contextual use of different information about depth. *Perception of space and motion*, page 69, 1995.

[7] Willemijn S. Elkhuizen, Tim Zaman, Wim Verhofstad, Pieter P. Jonker, Joris Dik, and Jo M. P. Geraedts. Topographical scanning and reproduction of near-planar surfaces of paintings. In *SPIE 9018*, volume 9018, pages 901809–901809–12, 2014.

[8] David Ferstl, Christian Reinbacher, Rene Ranftl, Matthias Rther, and Horst Bischof. Image guided depth upsampling using anisotropic total generalized variation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 993–1000, 2013.

[9] Luc M. J. Florack, Bart M. ter Haar Romeny, Jan J. Koenderink, and Max A. Viergever. Scale and the differential structure of images. *Image and Vision Computing*, 10(6):376–388, 1992.

[10] The Institute for Digital Archaeology and Dubai Future Foundation. The million image database, 2017.

[11] Alejandro F. Frangi, Wiro J. Niessen, Koen L. Vincken, and Max A. Viergever. Multiscale vessel enhancement filtering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 130–137. Springer, 1998.

[12] Wilson S. Geisler. Mechanisms of visual sensitivity: backgrounds and early dark adaptation. *Vision Research*, 23(12):1423–1432, 1983.

[13] Berthold K. P. Horn. Obtaining shape from shading information. In *Shape from shading*, pages 123–171. MIT press, 1989.

[14] Norden E. Huang and Zhaohua Wu. A review on hilberthuang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46(2), 2008.

[15] James T Kajiya. The rendering equation. In *ACM Siggraph Computer Graphics*, volume 20, pages 143–150. ACM, 1986.

[16] Jan J. Koenderink. The structure of images. *Biological cybernetics*, 50(5):363–370, 1984.

[17] M. Kolomenkin, G. Leifman, I. Shimshoni, and A. Tal. Reconstruction of relief objects from line drawings. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 993–1000, 20-25 June 2011.

[18] B. A. J. Lenseigne, E. C. C. Van Breda, and P. P. Jonker. Adaptive multi-dimensional data decomposition, 2016.

[19] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 131–144. ACM Press/Addison-Wesley Publishing Co., 2000.

[20] Tony Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–156, 1998.

[21] C. Lippolis and A. Rubino. *The Sennacherib Wall Reliefs at Nineveh*. Le lettere, 2011.

[22] Fernand Meyer. Topographic distance and watershed lines. *Signal processing*, 38(1):113–125, 1994.

[23] Dan Motzenbecker and Kyle Phillips. Autodraw, 2017.

[24] Oce. Eiger elevated printing, 2017.

[25] Oce. The genius of the great masters at your fingertips, 2017.

[26] T. Rindfleisch. Photometric method for lunar topography. 1966.

[27] Shapeways. Designing for better colors in full color sandstone, 2017.

[28] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.

[29] Matthew Vincent. Rekrei, 2017.

[30] Qiong Zeng, Ralph R. Martin, Lu Wang, Jonathan A. Quinn, Yuhong Sun, and Changhe Tu. Region-based bas-relief generation from a single image. *Graphical Models*, 76(3):140 – 151, 2014. Computational Visual Media Conference 2013Second Computational Visual Media Conference (CVM).

[31] Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. Shape-from-shading: a survey. *IEEE transactions on pattern analysis and machine intelligence*, 21(8):690–706, 1999.

[32] Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip H. S. Torr. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1529–1537, 2015.

A
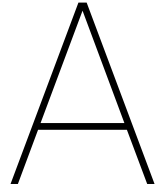
Dataset

A.1. Intro

The pictures are all taken in room V of the Southwest Palace of Sennacherib. A map of the palace is shown in Figure A.1. More information about the original excavation, drawings, photos and notations of the original excavations can be fount in the book Barnett et al. [1].

V34-9
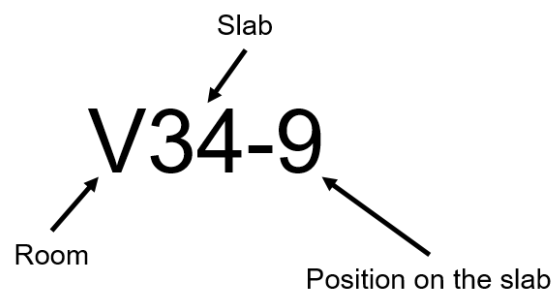
The photo's are numbered accordingly to their room number, slab and position on the slab:

A map of the position of each slab is shown in Figure A.2.
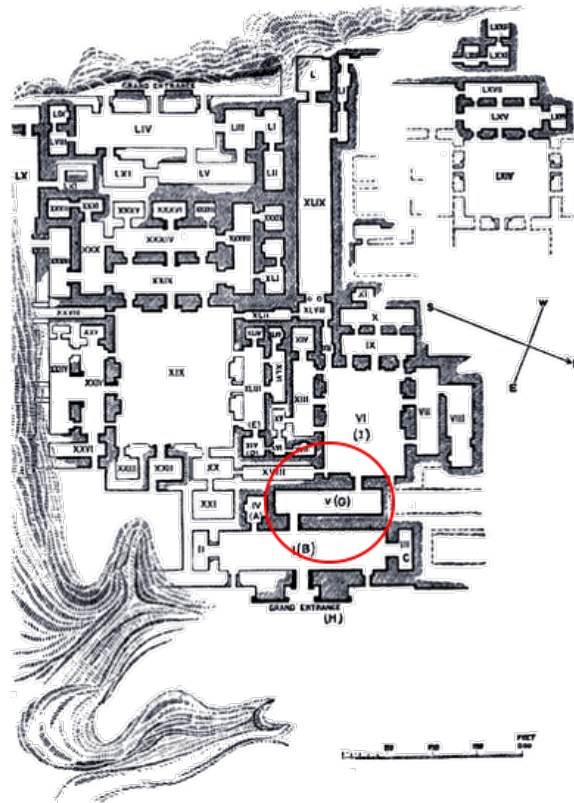
Figure A.1: A map of the SouthWest Palace of Sennacherib. Room V is appointed by the red circle [1].
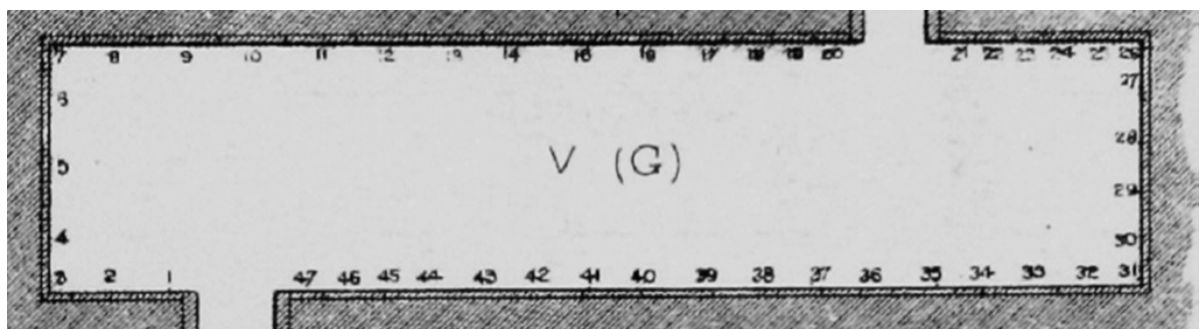
Figure A.2: A map with the numbering of the slabs in Room V [1].

A.2. Pictures

The photo database has been subjectively been divided into 4 categories:

- Best

- Good

- Bad

- Unrecognisable

The sorting was done based on the impression of the details and damage by fire or wear. Photos with strong casted shadows or vegetation were considered unrecognisable as well.

Some examples of each database can be found below. In our research we only focused on the Best and Good photos. In the future we could try to recover parts of the other categories as well.

(a) Photo V34-9: The head of an soldier, with a bow and arrows in his back.

(b) Photo V37-13: An archer and a servant
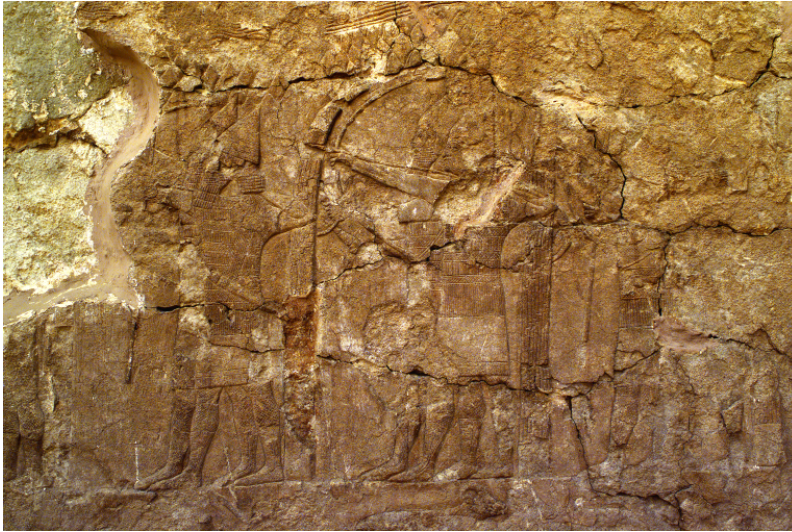
(c) Photo V39-21: Several archers together, grouped in pairs.

(d) Photo V3d-2, Three soldiers, one on the left and a pair on the right, appear to throw objects. Photo is with flash and some burn spots.

Figure A.3: **a** and **b** are of the category: Best. **c** and **d** are from the category Good.

(a) Photo V11-1

(b) Photo V33s-1: Some plants with the diamond shape background. This decoration was typical for the bottom and top of the slabs.

(c) Photo V11-7: parts of Cuneiform script.

(d) Photo V35-6: From the the double pair of legs, it is possible to see that it were 2 soldiers.

Figure A.4: **a** and **b** are of the category: Bad. **c** and **d** are from the category Unrecognisable.

A.3. Miscellaneous

The Italian researchers have performed some 3D measurements. Unfortunately, the 3D is not accessible anymore, since the software is no longer compatible with the current computer systems. The data is already 15 years old. Only DWG files (Figure A.6) are left of the slabs, which can provide us the outer dimensions. With *GetUnits.m* it is possible to select points in the DWG file and the corresponding photo to obtain the correct units of length for each photo.

(a) Left image

(b) Right image
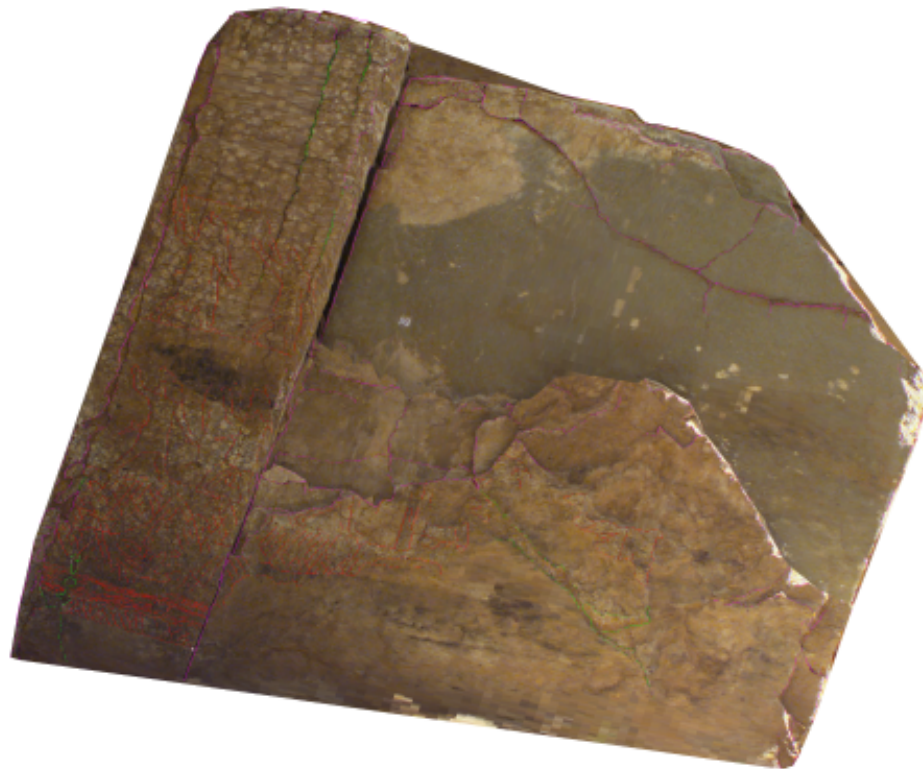
(c) 3D reconstruction of V3 and V4

Figure A.5: An image pair and the corresponding reconstruction.
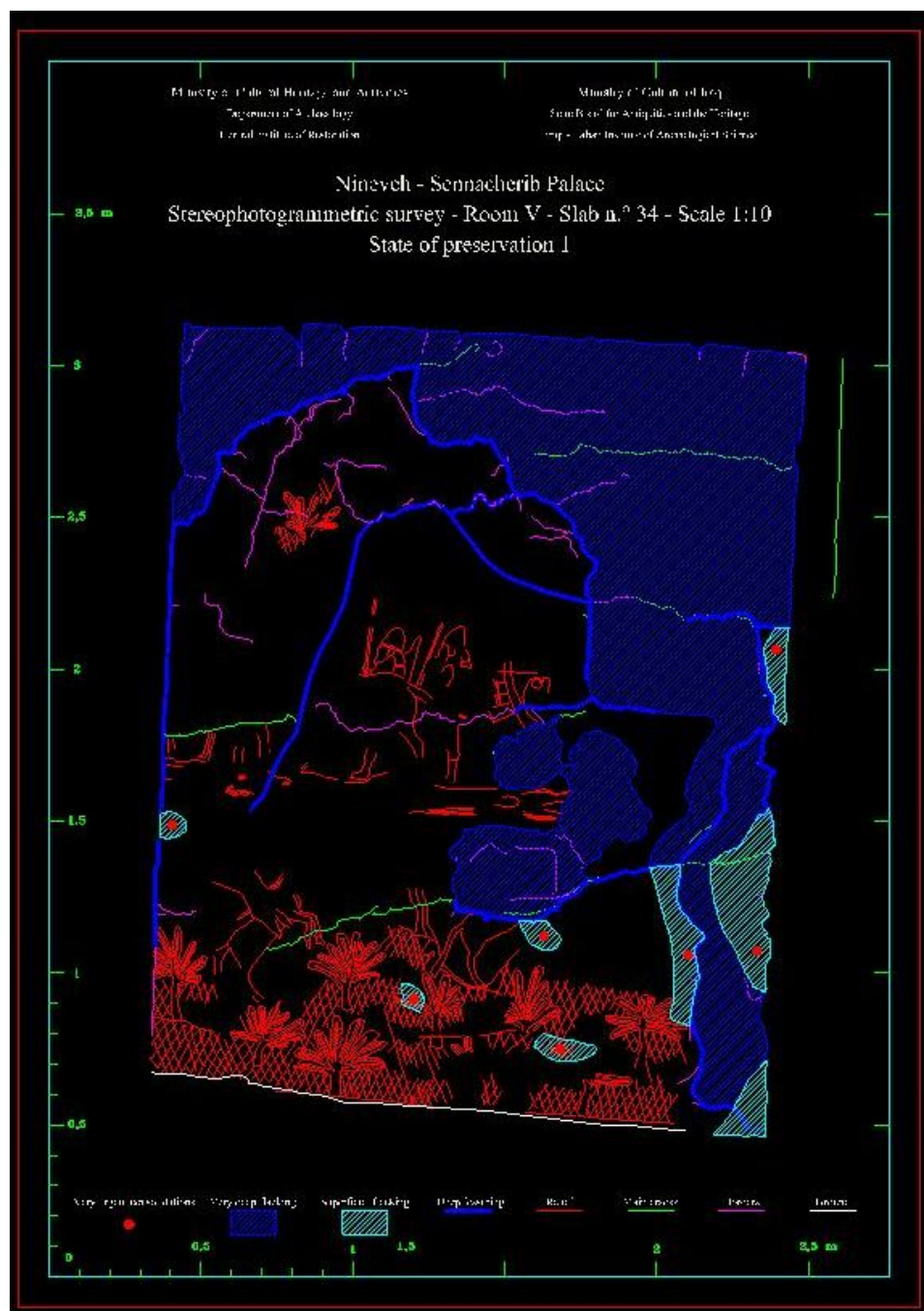
Figure A.6: DWG file of slab 34 with the outer dimensions and some rough sketches on it.

A.4. 3D scan

The 3D scan was made with the ATOSS III Triple scanner of GOM [5].

(a) Original relief

(b) 3D scan of the relief in greyscale

Figure A.7: The 3D scanned object.
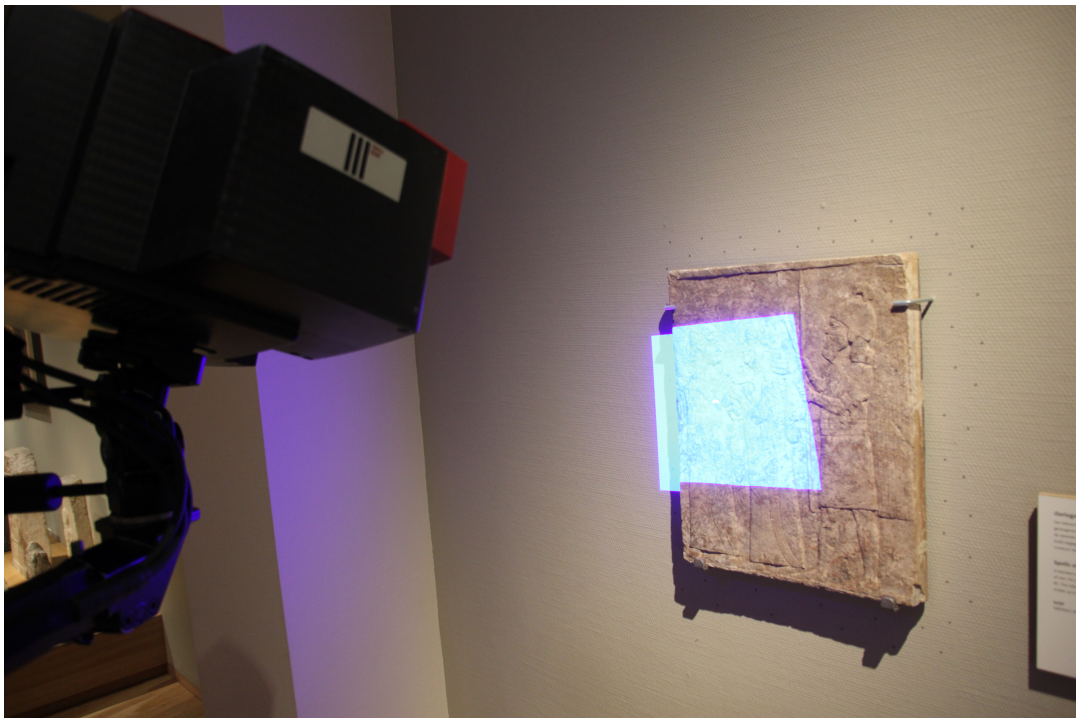
Figure A.8: During the process of 3D scanning. Around the relief, small markers have been applied on the wall as reference.

B

Line Detection

As mentioned in the paper, the line drawings are the most user-intensive part of this research. This was already recognized from the start of this research. Therefore, the focus in the beginning of this project was on automatically producing the line drawings. Unfortunately the results were not sufficient and the line drawings had to be drawn by the archaeologists of the RMO. However, the computer generated results are still interesting for future results. The different results and techniques are discussed here below. All methods are applied on the converted grey images of the original photos.

B.1. HSV Histograms

One of the simpler ways of line segmentation is based on colour. In our research, the RGB colour space, Figure B.1a, was converted into the HSV colour space, Figure B.1b. The HSV colour space was chosen for the segmentation, since it provides invariance to the luminosity of the colour. HSV is decomposed in the Hue, Saturation and Value.

The images used for colour analysis are V34-9 and V37-4, which are both part of the "Best" category discussed in the Appendix Dataset, Figure B.2. Next, the seperate colour signals (Hue, Saturation and Value) were plotted in histograms, to see whether there was a distinctive difference between lines, cracks and rest (background), Figure B.4 and Figure B.5.

By comparing the histograms of V34-9 and V37-4, we conclude that the colour distribution is comparable for both images. Furthermore, the peaks for the lines and cracks are distinguishable from the rest (background) in both the Saturation as Value channel. Between the cracks and lines there is no clear distinction. More, there is no clear threshold to separate the lines, cracks and rest from each other, since their distributions overlap with each other.

The HSV values in Matlab are in the range of 0 to 1. Every bin has a width of 0.02 in the histogram with 50 bins in total. For the colour segmentation in Figure B.3, we set the thresholds for Saturation > 0.88 and Value < 0.22, respectively bin 44 and 11. Although, the original sculptures are recognisable in the colour segmentation results, Figure B.3, a lot of details are lost. So, the HSV thresholds alone are not sufficient enough for the line detection, but could be used in a feature vector to improve the reliability.

B.2. Basic Matlab filters

Matlab contains a set of basic line detection algorithms. The best results were obtained with the *Prewitt* and *Canny* filters, Figure B.6. Although, these results are not sufficient at all. Another problem is the location of the detected lines. The Matlab filters are edge detectors, but in practice we need to detect the ridges. With the edge detectors of Matlab, the boundaries of these ridges are found, instead of the centreline.

B.3. Frangi filter

The Frangi filter has been designed to find ridges, [3]. The results for V34-9 are shown in Figure B.8.

B.4. Empirical Mode Decomposition

While we use the EMD in our research as part of Shape from Luminance, the EMD can be used as line detection as well. Recall that the first IMF1 contains the fine details of the image. These details correspond with the lines in the image and appear darker in the IMF. By applying a manual threshold to the IMF, we obtained the line drawing in Figure B.7. The first IMF also contains the stone texture, which becomes visible as noise in the final result. Some noise filtering is done by removing areas smaller than 350 pixels.

B.5. Support Vector Machine

The Support Vector Machine (SVM) is linear classifier, which uses supervised learning. We trained the SVM classifier with a small dataset of Figure B.2a. The data was labelled as Line or Not-a-Line. The data itself was computed by building a Surffeatures with a feature lengt of 64 around every point. After training the SVM, we applied the SVM on the whole image of V34-9, which produced the result in Figure B.9. The SVM result suffers from the same problem as the EMD method: noise of the stone texture.

B.6. Conclusion

In conclusion, the SVM, Figure B.9 and the Frangi Filter, B.8, are the most promising methods. However, the results are not sufficient to make an artificial reconstruction. Therefore, we decided to use the manual drawn line drawings.

Thresholding in combination with the EMD has been discarded, since the SVM results slightly contained more details, as for example the arrow tails in the backpack of V34-9. Also, the threshold for the IMF can differ substantially among different pictures.

Note that the results of the Frangi filter are in grey scale, B.8, so no threshold was applied. All other methods used a threshold, which could be adapted to change the results. However, we believe that the shown results represent each method good enough to make a fair comparison between each method.

Another interesting property to investigate is the specific direction of the different structures. In his project [9], Chris Mata was able to distinguish camouflaged objects from the background using Gabor filters. With Gabor filters it is possible to filter an image for certain frequencies in a specific direction. His hypothesis, was that man-made structures are mostly orientated in one direction, while natural structures will show appear in multiple directions. He tested his hypothesis by filtering the images for multiple directions and then accumulated the score over all orientations. The man-made structures had the lowest accumulated score in the image. If we look at the noise in the SVM and Frangi result, it is clear that the man-made lines appear to be smooth as well, while the textural noise do not have a dominant orientation.

For directional filtering, the Histogram of Gradients (HOG) can be used as well. For more information, one should read the research of Dalal and Triggs [2] about HOG for human detection.

(a) RGB colour space   (b) HSV colour space
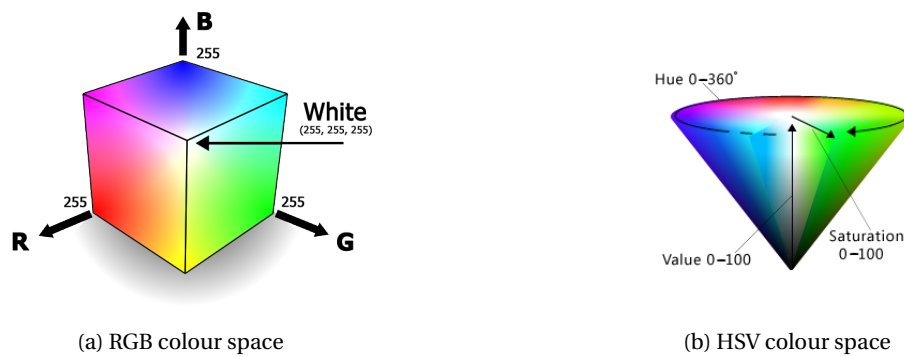
Figure B.1: The different colour spaces. The lumination is in HSV scale represented by the Value on the vertical axis.

(a) Photo V34-9

(b) Photo V37-4

Figure B.2: The trainings images for edge detection with colour separation. The red lines correspond with the man-made carvings and the green lines are large cracks, which could be misinterpret as lines.
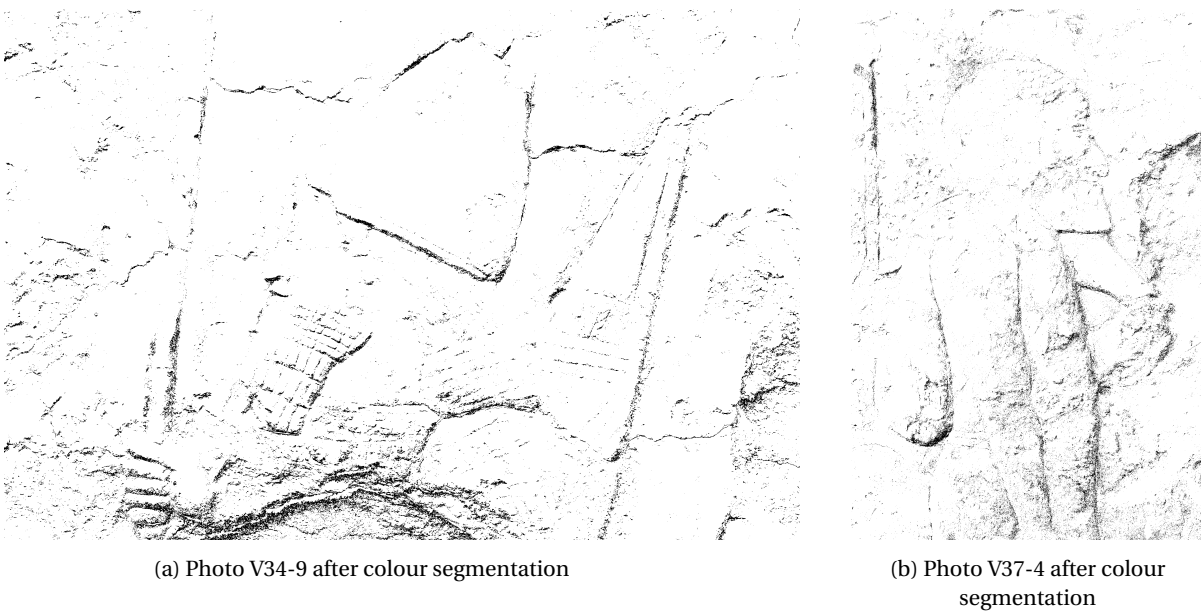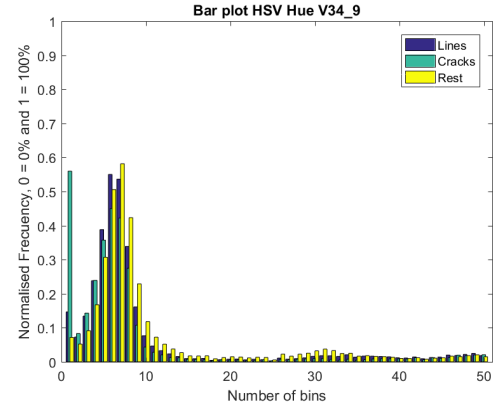

(a) Photo V34-9 after colour segmentation

(b) Photo V37-4 after colour segmentation

Figure B.3: The segmentation results with the thresholds: Saturation > 0.88 and Value < 0.22
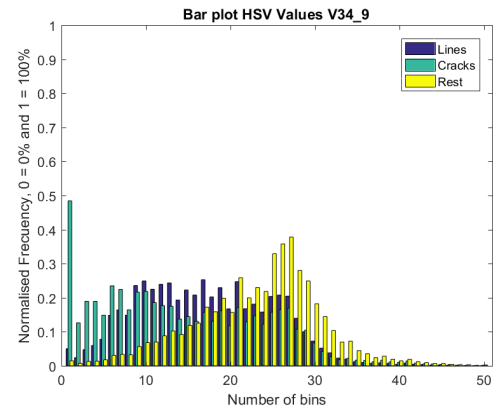
(a) Hue channel of V34-9

(b) Histogram of the Hue in V34-9

(c) Saturation channel of V34-9
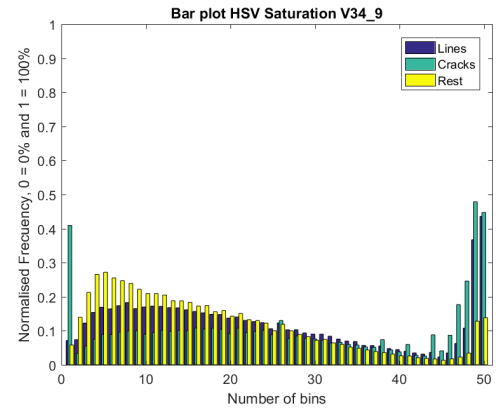
(d) Histogram of the Saturation in V34-9

(e) Value channel of V34-9

(f) Histogram of the Value in V34-9

Figure B.4: An overview of the effect of using Saygili's fusion method on several shapes. The left-sided figures show the shapes before fusion and the figure on the right-side are after fusion. The fusion is done with the same shape and same weights. All the shapes are without units, however the height is distinguishable by colour, with red being the highest value and dark blue the lowest.

(a) Hue channel of V37-4
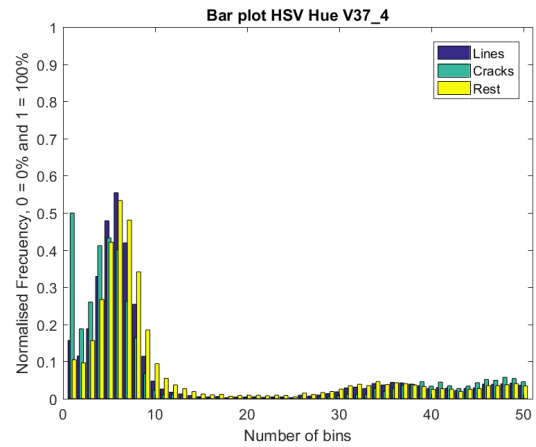
(b) Histogram of the Hue in V37-4
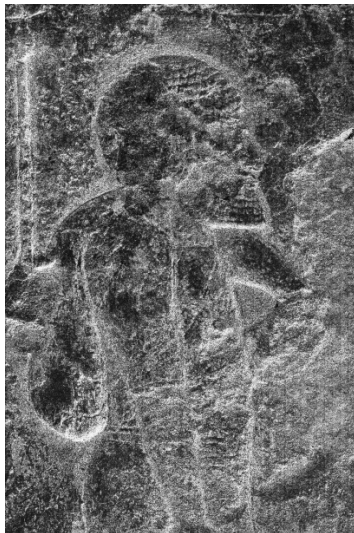
(c) Saturation channel of V37-4
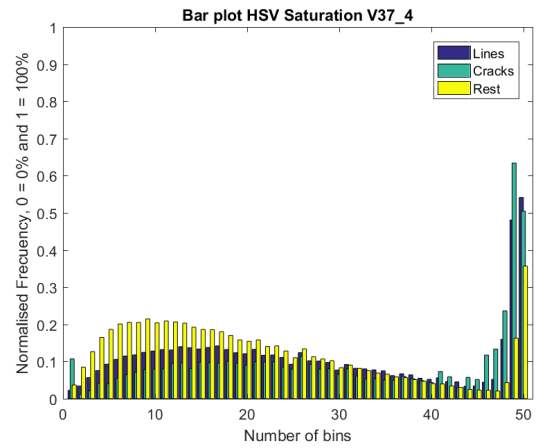
(d) Histogram of the Saturation in V37-4

(e) Value channel of V37-4
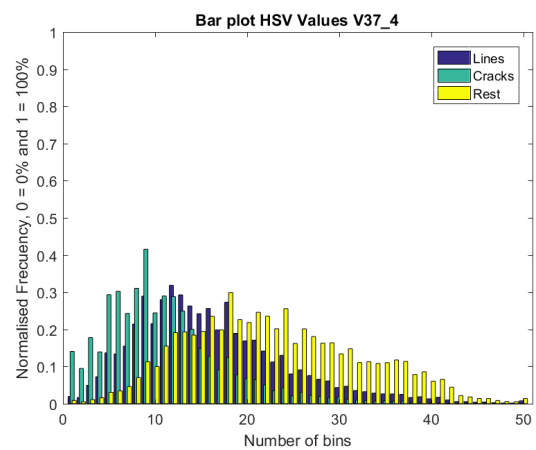
(f) Histogram of the Value in V37-4

Figure B.5: An overview of the effect of using Saygili's fusion method on several shapes. The left-sided figures show the shapes before fusion and the figure on the right-side are after fusion. The fusion is done with the same shape and same weights. All the shapes are without units, however the height is distinguishable by colour, with red being the highest value and dark blue the lowest.
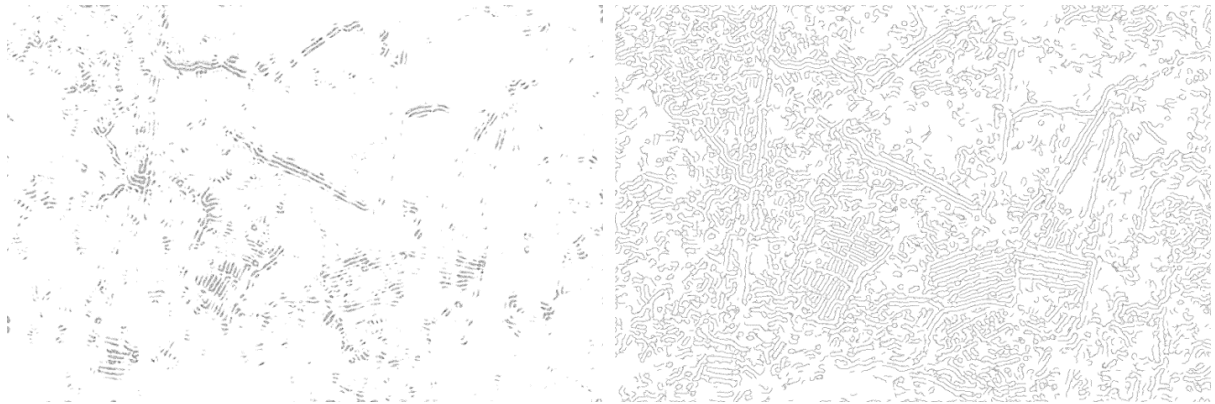
(a) Photo V34-9 filtered with the Prewitt operator.

(b) Photo V34-9 with the Canny edge detector.

Figure B.6: The two best filter results of the basic Matlab functions. The images have been pre-filtered with a Gaussian blur of $\sigma = 10$.
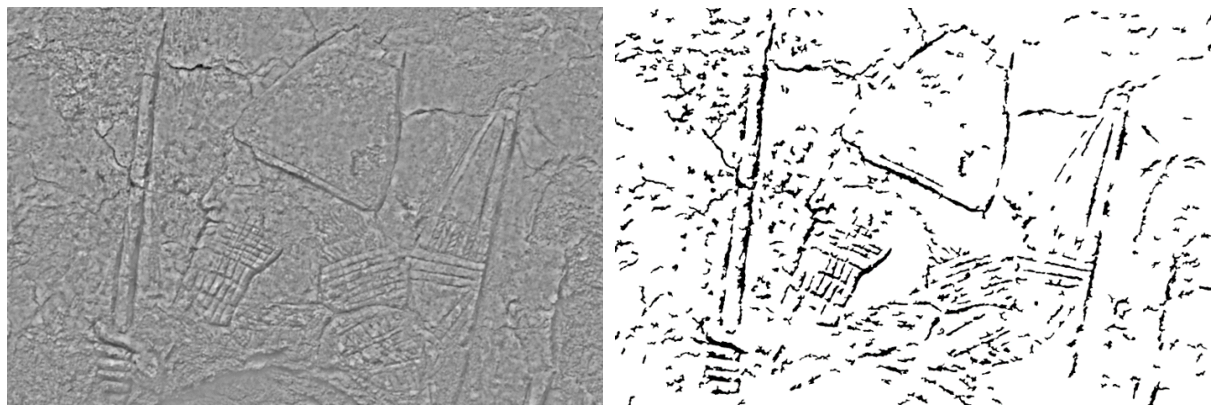
(a) IMF 1 of photo V34-9 filtered with the EMD.

(b) The IMF is in grayscale between 0 and 255. The edges are darker, so the pixels were selected which are below the value of 20. Next all areas, which are smaller than 350 pixels, were removed.

Figure B.7: Manual threshold with the first IMF of the EMD on photo V34-9.

Figure B.8: The Vesselness measure of the Frangi filter. Note that this image is in grey scale, while the other line images are in Black and White. The other methods use some threshold.

Figure B.9: Photo V34-9 filtered with SVM, which is trained with surf features. The feature length was set for 64. After filtering blobs with a size smaller than 350 pixels were removed.

C

Shape from Luminance

C.1. V37-13

(a) The original RGB image
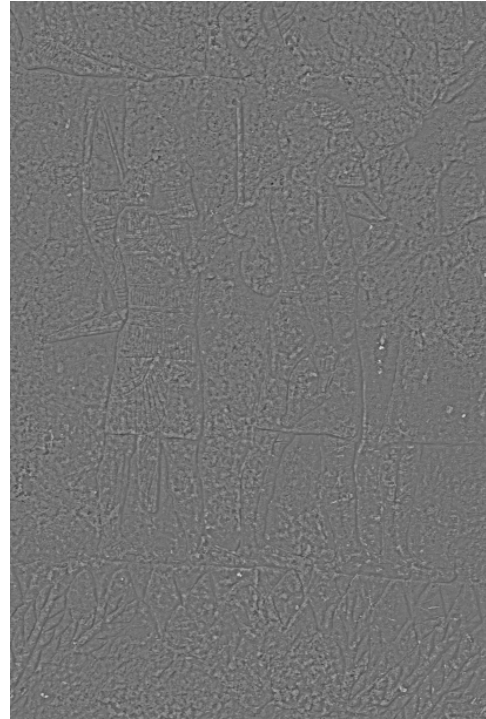
(b) The first IMF
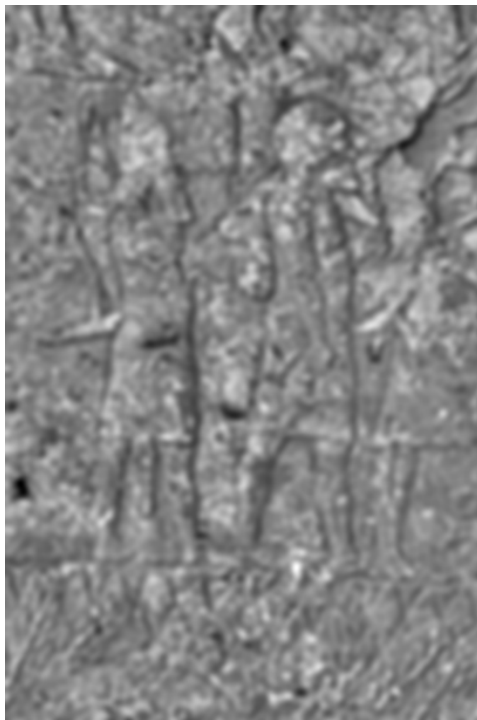
(c) The second IMF
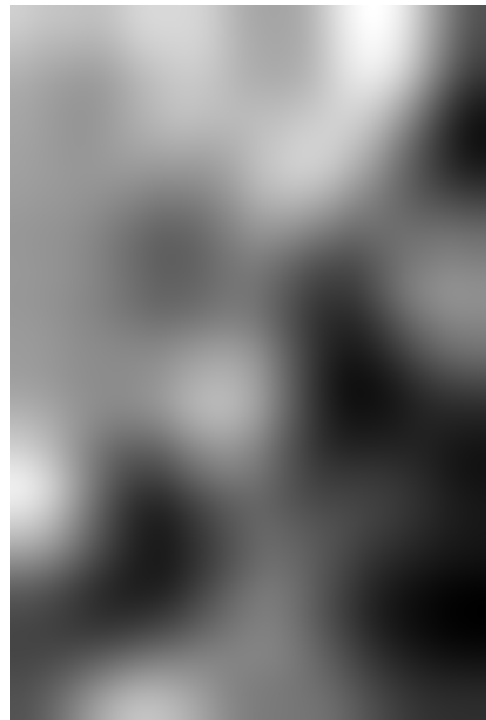
(d) The residual of the luminance

Figure C.1: SFL results of V37-13

C.2. V44-3

For the final reconstruction of V44-3, we used a cropped version. As you can see, a part of the ground is visible in the IMF. We removed this to prevent strange artefacts in the reconstruction.

(a) The original RGB image

(b) The first IMF

(c) The second IMF

(d) The residual of the luminance
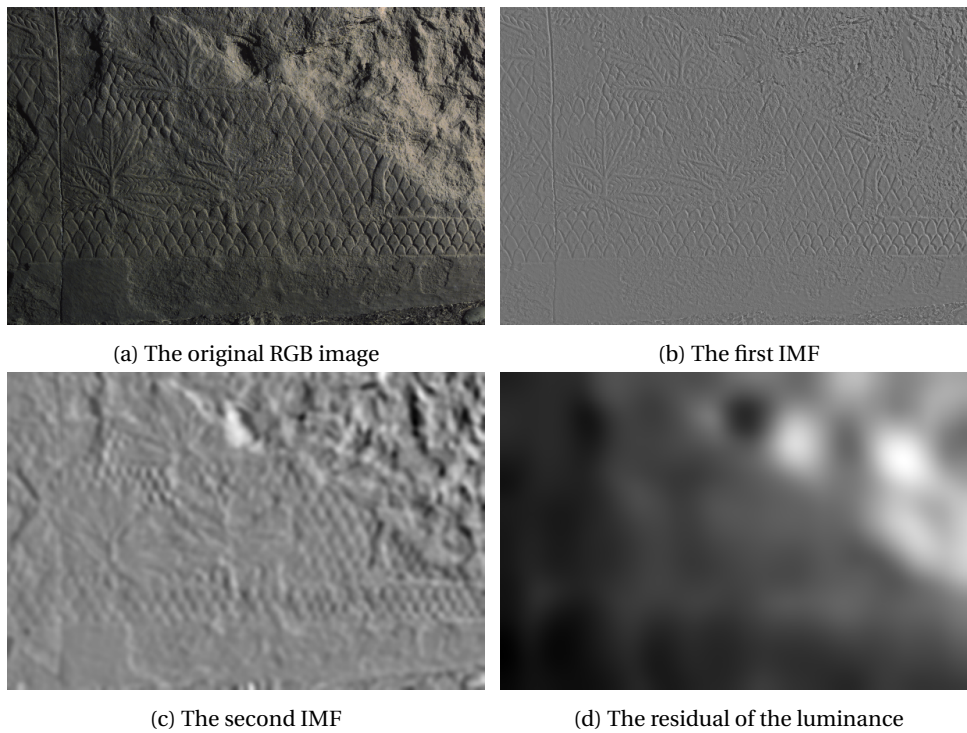
Figure C.2: SFL results of V44-3

C.3. Illumination Sensitivity

The illumination sensitivity can be adapted by changing the value for n. Recall Equation C.1 [4]:

$$D_{SFL}(x, y) = \frac{L_f^n(x, y)}{L_f^n(x, y) + \sigma^n} \tag{C.1}$$

with σ being the correction for illumination sensitivity:

$$\sigma^n = \exp\left(\frac{1}{N} \sum_{x,y} \log(L_f(x, y))\right) \tag{C.2}$$

According to the paper of Geisler [4], the optimum of n would be around 0.73. Closer investigation showed that playing with this value could be beneficial. We have taken V34-9 as example and show the SFL reconstruction for multiple values of n in Figure C.3 and C.4. We observe that the results improve with in increase of n. This is remarkable, since the parameter n has not been tuned before in combination with the EMD. Our results clearly show that the SFL results improve by tuning n. However, the case of the reliefs is special, since the luminance is more or less constant across the scene. Future research should determine a possible strategy of tuning the illumination sensitivity according to the scene.

In our research, we chose $n = 2$. This prevents the saturation effect at the lower grey values as in $n = 3$ (Figure C.4h), but emphasizes the details much more than $n = 0.73$ (Figure C.4b).

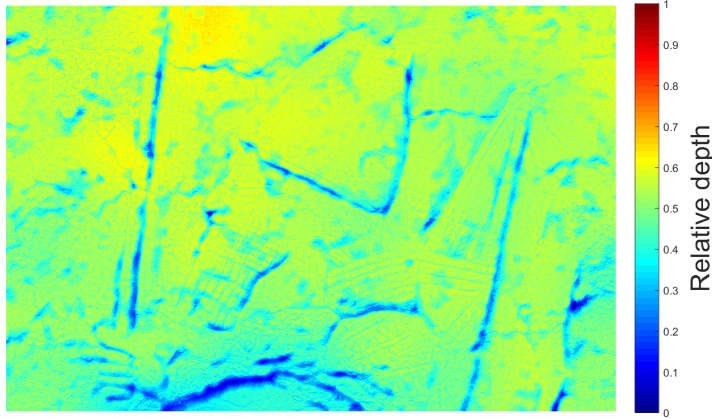The SFL reconstruction is made with 1xIMF1 + 2xIMF2 + 1xIMF3.

(a) SFL depth map for $n = 0.5$ (b) Illumination sensitivity for $n = 0.5$
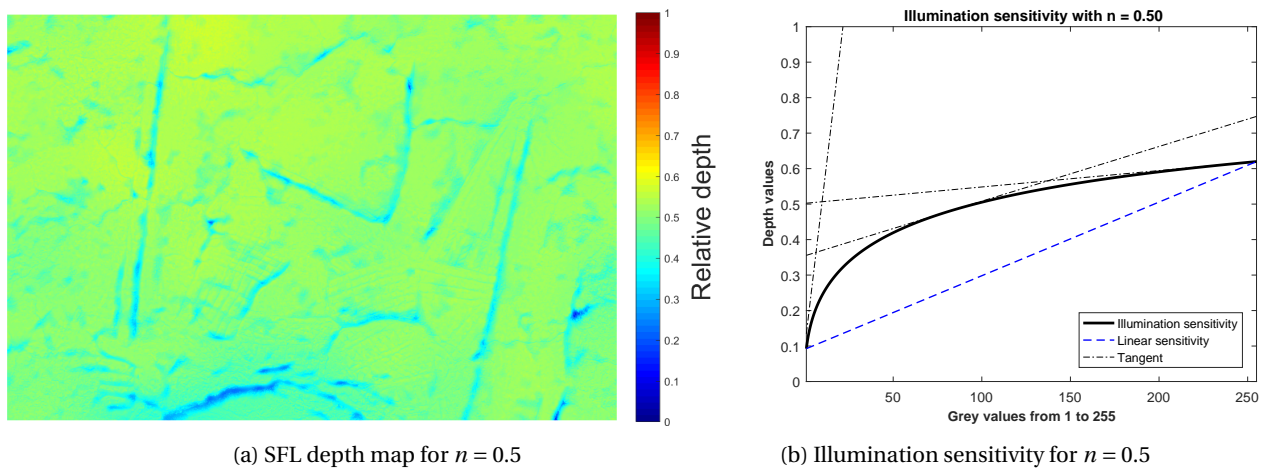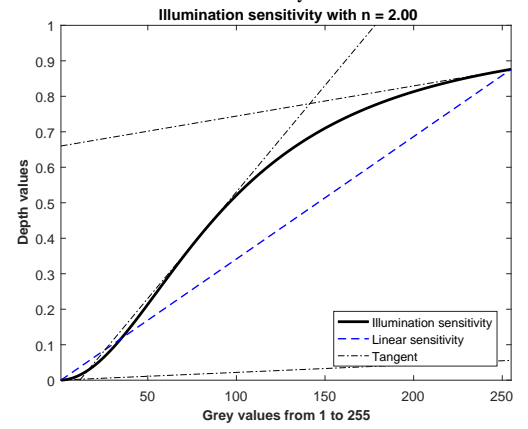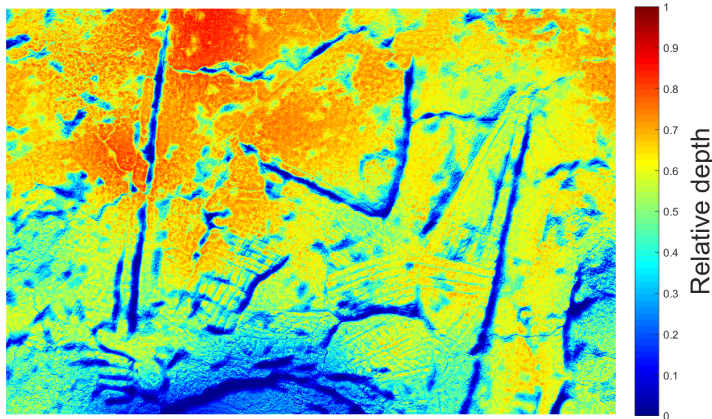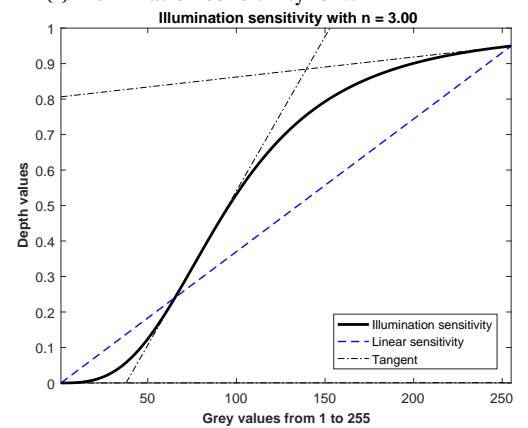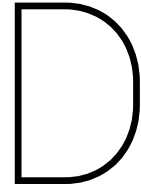
Figure C.3: $n = 0.5$
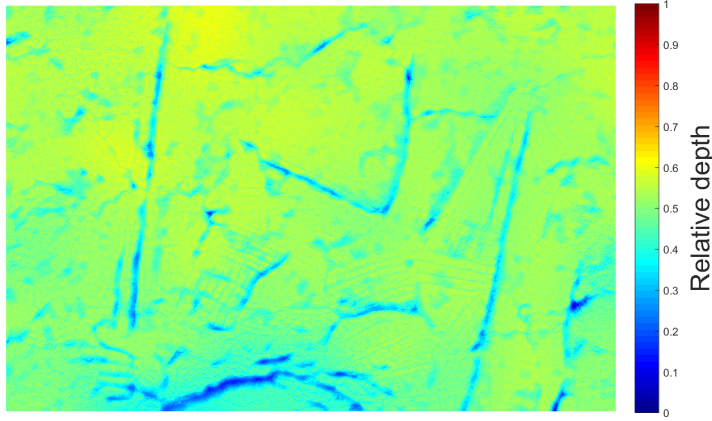
(a) SFL depth map for $n = 0.73$
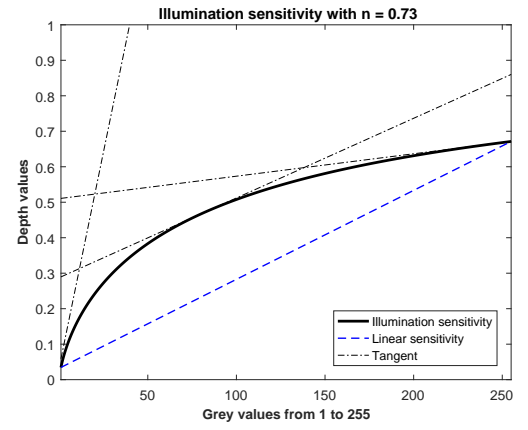
(b) Illumination sensitivity for $n = 0.73$

(c) SFL depth map for $n = 1$
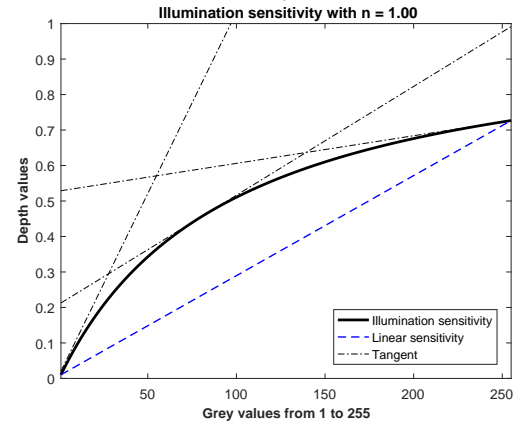
(d) Illumination sensitivity for $n = 1$
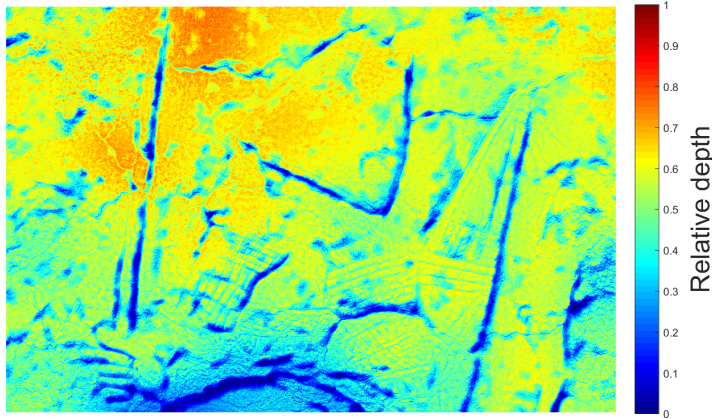
(e) SFL depth map for $n = 2$

(f) Illumination sensitivity for $n = 2$

(g) SFL depth map for $n = 3$

(h) Illumination sensitivity for $n = 3$

Figure C.4: Overview for $n = 0.73 - 3$.

D

Artificial Reconstruction

D.1. Line drawings

The line drawings used for reconstructions in this research are shown in Figure D.3 and D.4. The line drawings are human drawn and therefore sensitive to human interpretation. In Figure D.1 and D.2, two of those mistakes are displayed.

Nonetheless, the manual line drawings are better than the automated line drawings and useful for the artificial reconstruction.

D.2. Watershed

In Figure D.5a the euclidean distance per pixel for photo V34-9. The watershed result is shown in Figure D.6a. The original sculpture is barely visible. To solve this problem the Euclidean distance is converted into a grey scale between 0 and 1 and then inverted (Figure D.5b). This provides us with a more useful watershed result (Figure D.6b).

The watershed algorithm places the sources (starting point of the algorithm) at the local minima. In Figure D.5a these minima are located at the lines: the distance is there 0. However, we would like to have the minima located at the centre of each section. This is done with the inversion. It is also possible to use user-defined markers by assigning local minima. We have not tested this method. Still, we believe that it could improve future results.
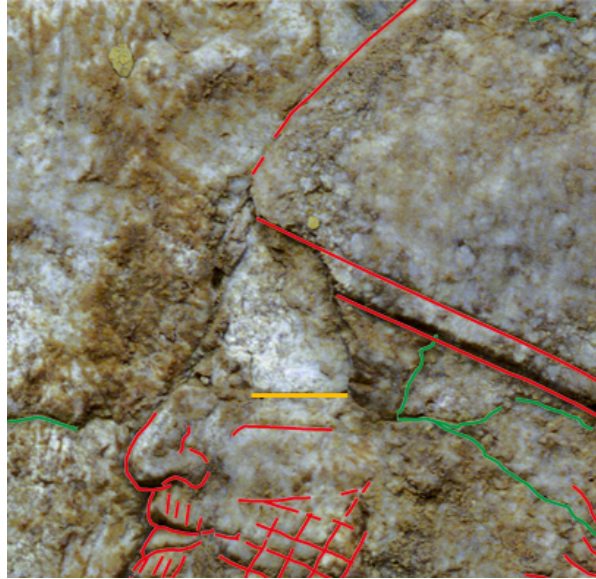
Figure D.1: A miss-drawing in V34-9, The orange line gives the missing line.

(a) A photo of the RMO slab

(b) The line drawing of the RMO relief.

(c) Line drawing of V34-9, provided by the RMO. Red are the lines and green the cracks.

Figure D.2: A mistake in the RMO line drawing. The pointing finger does not appear on the 3D scan and is a mistake by human interpretation.
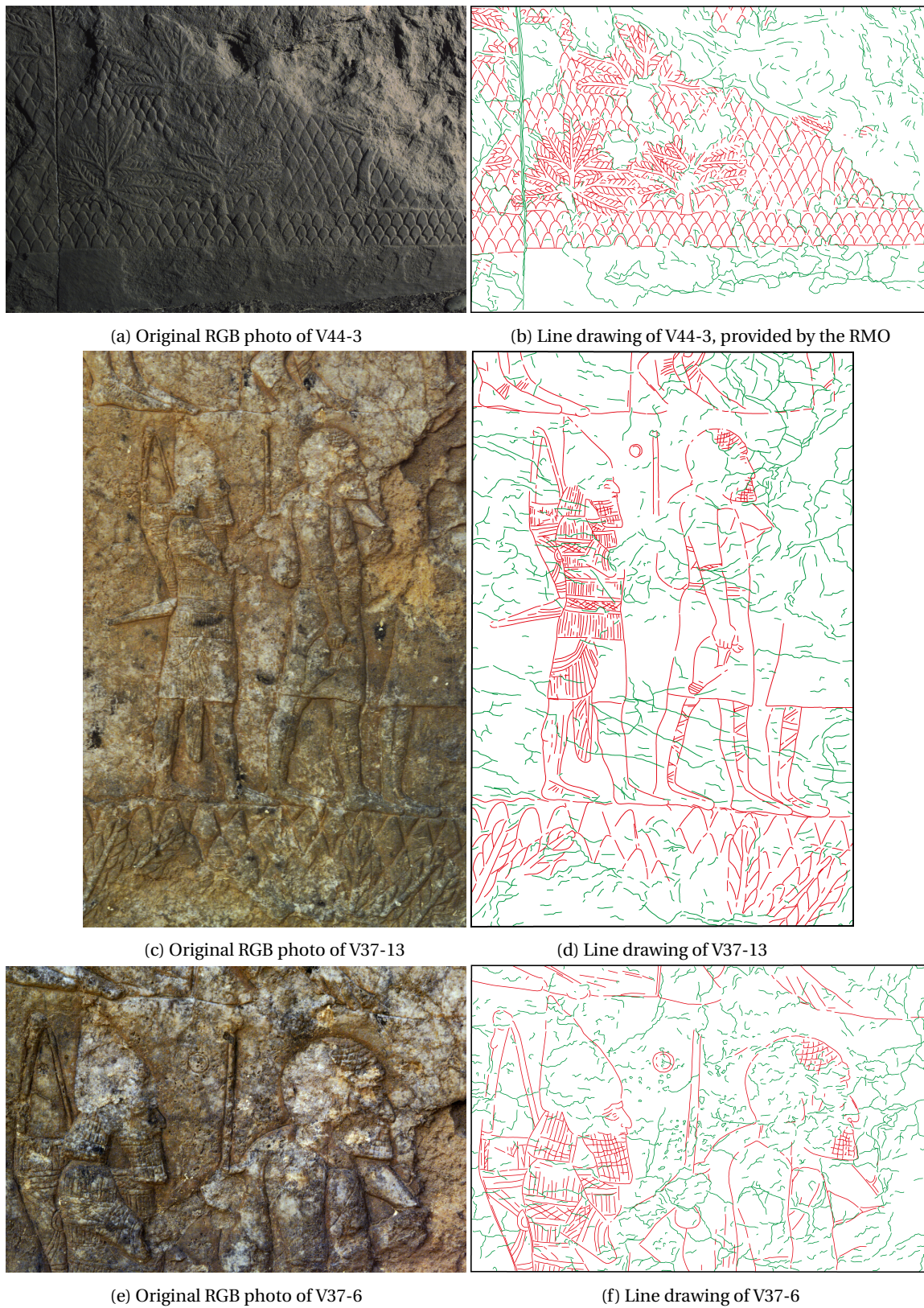
(a) Original RGB photo of V44-3

(b) Line drawing of V44-3, provided by the RMO

(c) Original RGB photo of V37-13

(d) Line drawing of V37-13

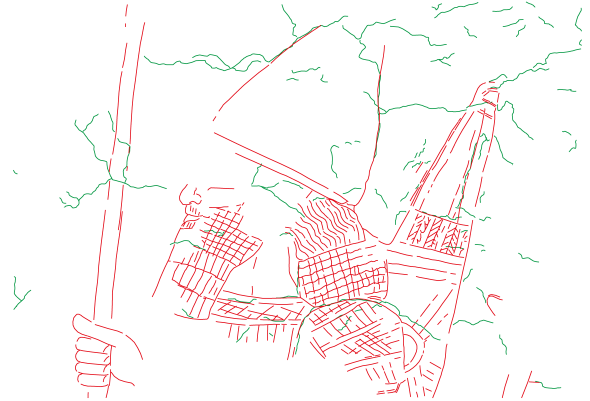(e) Original RGB photo of V37-6

(f) Line drawing of V37-6

Figure D.3: An overview of the line drawings used in this research of images V44-3, V37-13 and V37-6.

(a) Original photo of V34-9

(b) Line drawing of V34-9, provided by the RMO. Red are the lines and green the cracks.

Figure D.4: The RGB photo (Left) is used by the RMO to create the line drawing on the right.
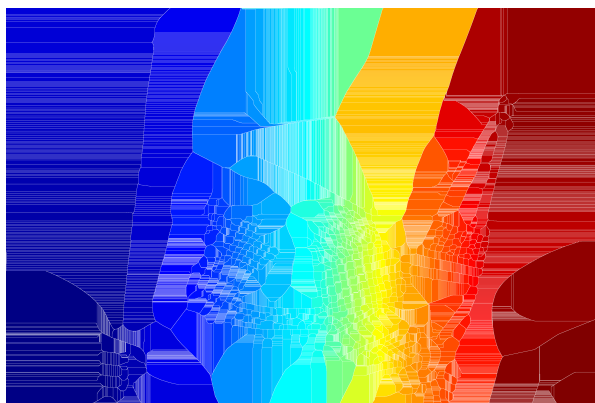

(a) The Euclidean distance of every pixels to the closest line in V34-9. Black is the lowest value and white the largest.
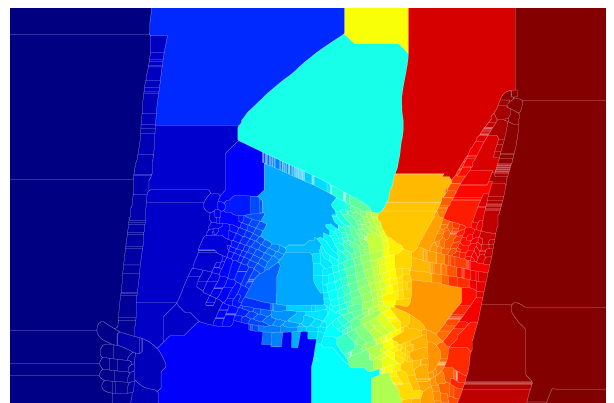
(b) The inverse of the Euclidean distance

Figure D.5: Left the results of the Euclidean distance per pixel. Right the inverse of the Euclidean distance.
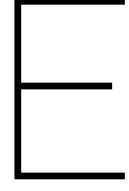

(a) Watershed result of the Euclidean distance map.

(b) Watershed result of the inversed Euclidean distance

Figure D.6: Left the watershed result for the Euclidean distance. Right the result for the inversed Euclidean distance. The different colours serve as label for each section.

E

Data Fusion

Originally we planned to use the data fusion method of Saygili, [6] as fusion method, since it already had been successfully applied to the fusion of several 3D maps. Unfortunately it turned out not to work in our case. The difference has been found in the usage of a continuous 3D scene versus discontinuous. Saygili's method was previously only applied in scenes, where multiple objects were standing in front of a background, such as a wall. This creates "jumps" in the depth map between the objects and also the background, whereas our depth maps are close-up of one object, resulting in a continuous shape. To understand the problem, we will first explain Saygili's method and then show the problem. Furthermore, we will elaborate more on our proposed solution: the Frangi Filter with the vesselness Measure.

E.1. Saygili

Saygili [6] has developed method which fuses depth maps based on their local confidence. The fusion algorithm has no limitation on the number of input sources, as long as each source has its own confidence map. This makes this approach suitable for fusing the results of SFL and the artificial reconstruction. The local adaptability is important to preserve the features of each reconstruction: the fine details from SFL and the smoothness of the artificial method.

The algorithm contains three stages. In the first stage the consensus set $H(x, y, d)$ is built, where x and y are the pixel coordinates and d is their corresponding disparity, Equation E.1.In the first stage the consensus set $H(x, y, d)$ is built, where (x, y) are the coordinates and d is the disparity:

$$H(x, y, d) = \sum_{\forall i} \sum_{(x_n, y_n) \in N} S_i(x_n, y_n) \mu_i(d_n, d) \tag{E.1}$$

with:

$$\mu(d_n, d) = \begin{cases} 1, & d_n = d \\ 0, & \text{otherwise} \end{cases} \tag{E.2}$$

The confidence S_i of each disparity is summed up over all inputs in a window N around each pixel to construct the consensus set H in Equation E.1.

In the second stage the disparity, d^*, is selected with the highest cumulative confidence, equation:

$$d^* = \operatorname*{argmax}_d(H(x, y, d)) \tag{E.3}$$

The maximized disparity is then used to select the location, n^*, of highest individual confidence:

$$n^* = \operatorname*{argmax}_n(S_i(x_n, y_n) \mu_i(d_n, d^*)) \tag{E.4}$$

Next, a modified disparity, C_i^* is created by substituting each measured disparity with disparity C_i of the

winner's location.

$$C_i^*(x, y, d) = C_i(x_{n^*}, y_{n^*}, d) \tag{E.5}$$

In the third stage is the fusion process itself. First the local weights are determined:

$$w_i(x, y) = \frac{S_i(x, y)}{\sum_{\forall i} S_i(x, y)} \tag{E.6}$$

Last, the final fused disparity is calculated by using the weighted sum of the modified measurements:

$$C_f(x, y, d) = \sum_{\forall i} w_i(x, y) C_i^*(x, y, d) \tag{E.7}$$

The size of window N is the only local parameter in this framework. In practice the best results are obtained for a 3x3 or 5x5 window. Increasing the window size does not only distort the local depth information, but it also increases the computation time [6].

E.1.1. Monocular Adaptation
Saygili originally designed his method for the fusion of the Kinect sensor[1], which presents the depth data in terms of disparity. Disparity is a natural number and represents the difference between two views in the number of pixels. With the Shape from Luminance algorithm the depth values will be represented in a relative scale between 0 and 1. In Equation E.1 the consensus set is built on the disparity. The probability of two neighbouring pixels having the same value is quite likely in the case of stereo depth maps. In the case of SFL this would be improbable [12]. To convert the SFL result to the same measure as the stereo depth maps, a resolution has to be set to divide the relative output of 0 - 1 into bins. These bins make it possible to built a consensus set.

E.2. Saygili Fusion results
In Figure E.2 the results of Saygili's fusion are shown. We fused the shapes on the left-side with the same shape and gave an equal weight to every pixel. One would expect that if you fuse the same shape with itself, you would obtain the same shape: the contrary seems to be true.

The problem lies within the defined window of N to construct the consensus set H. Imagine having a circular peak, Figure E.2c, and we define an arbitrary point with a 3x3 window around this point, E.1a. Due to the change in radius, there will be more "lower" values present than "higher" values with respect to the centre point. Imagine that all pixels have the same weight, the "lower" values will have the highest accumulated confidence and therefore be assigned as the correct depth, resulting in the distorted image of Figure E.1b. More, the window "walks" through the picture from left to right, starting from the upper left corner. This explains the bleeding effect in Figure E.1b from the upper left to the lower right corner. If multiple values have the same maximum accumulated confidence, Matlab will just select the first value.

The distortion increases with more complex shapes as shown in Figure E.2. The square shape, Figure E.2a, only shows some small distortions at the corners, Figure E.2b. If we would take a window around a corner point, roughly 25% would be elevated and 75% background. Which means that if the confidence is the same for all the pictures, the background would win, since it has the highest combined confidence. This effect only becomes worse with more rounder and continuous shapes.

Saygili already warned about possible distortions and therefore recommended a window size of 3x3 of 5x5. However, in our case both windows lead to an unacceptable distortion. One could try to correct the weights based on the local derivative or some other shape factor, though this would be a research on its own. The "quick" way to prevent this problem, was to use a single pixel based fusion method.
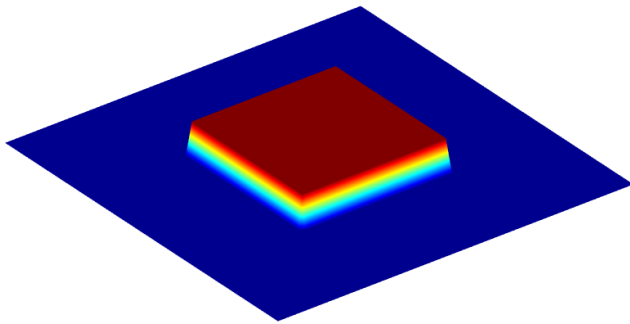

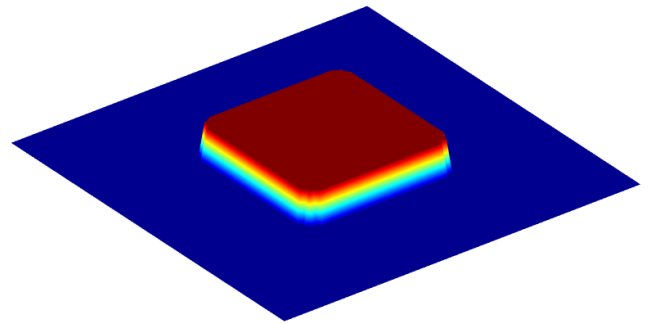[1] developer.microsoft.com/nl-nl/windows/kinect/hardware
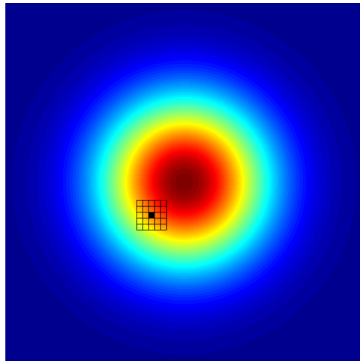
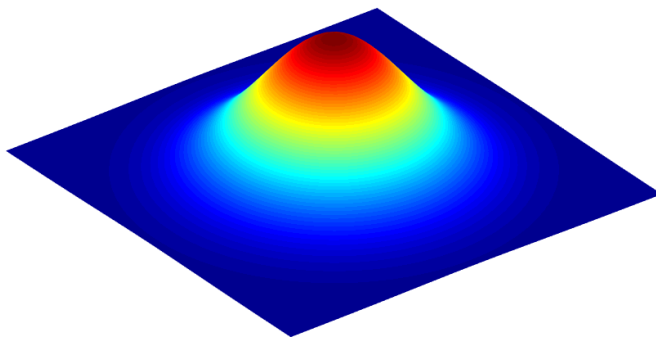(a) A 5x5 window for the Consensus set around a point on the Gaussian kernel.
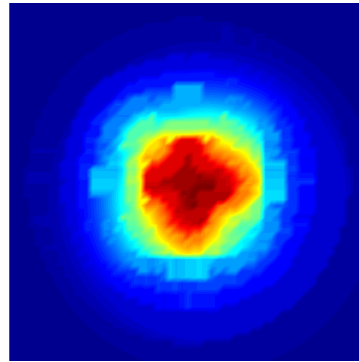
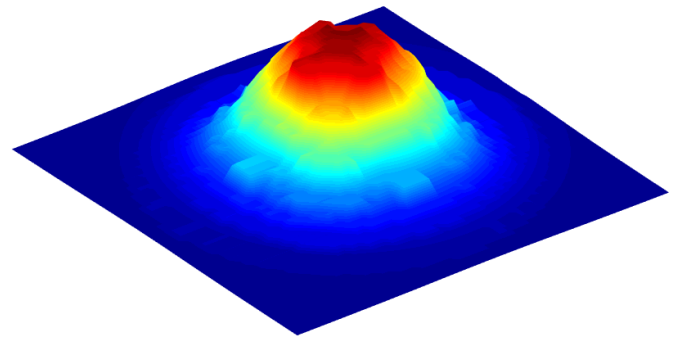(b) A topview of of the distorted Gaussian Kernel.
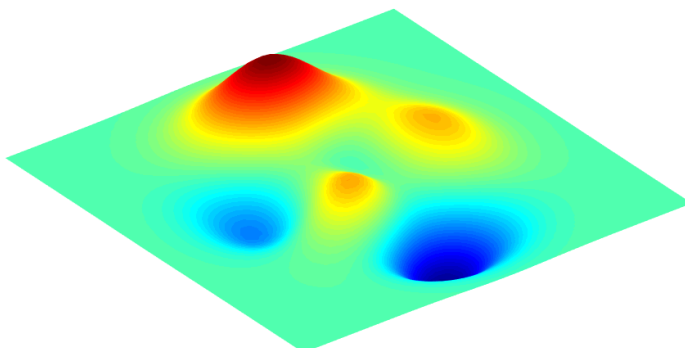
(a) An spatial elevated depth map of a square.

(b) The spatial elevated square after the fusion with Saygili's method. You see some artefacts appearing at the corners.
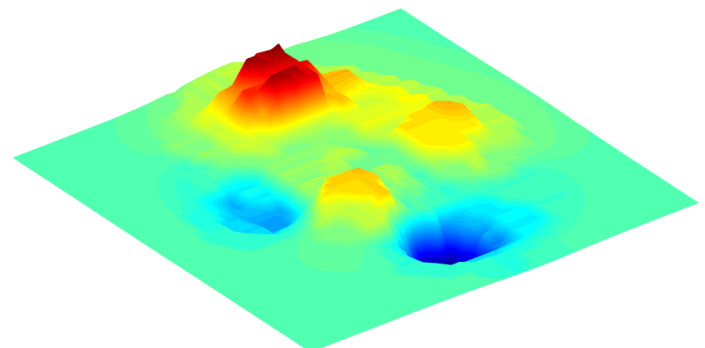
(c) A depth map of a continuous Gaussian kernel

(d) The same Gaussian kernel, but after fusion with Saygili. The kernel is not that smooth any more

(e) A more complex depth map generated with the Matlab function: *peaks*

(f) The distorted depth map of peaks after fusion

Figure E.2: An overview of the effect of using Saygili's fusion method on several shapes. The left-sided figures show the shapes before fusion and the figure on the right-side are after fusion. The fusion is done with the same shape and same weights. All the shapes are without units, however the height is distinguishable by colour, with red being the highest value and dark blue the lowest.

E.3. Vesselness measure

In our paper we introduced the "Vesselness" measure by Frangi [3], which combines the eigenvalues into a single measure:

$$\mathcal{V}(x,y,\sigma) = \begin{cases} 0 & \text{if } \lambda_2 > 0 \\ \exp\left(-\frac{\mathcal{R}_\beta{}^2}{2\beta^2}\right)(1-\exp\left(-\frac{S^2}{2c^2}\right)) \end{cases} \tag{E.8}$$

With the blobness measure:

$$\mathcal{R}_\beta = \frac{\lambda_1}{\lambda_2} \tag{E.9}$$

and structureness as explained in the paper.

$$S = \sqrt{\sum_{j\leq D} \lambda_j^2} \tag{E.10}$$

The parameters β and c could respectively be used to adapt the sensitivity to blobness and the magnitude of the eigenvalues of the Hessian. In Figure E.3, the "Vesselness" score for several values of the two parameters are shown.

We see that if β is increased, the Vesselness measure becomes more sensitive(Figure E.3d) to blob-like structures (E.1). An decrease of β (Figure E.3d) leads to a more selective measure for tubular structures.

If parameter c is decreased, the Vesselness score inscreases (Figure E.3c). Vice versa, if c increases, the Vesselness score decreases (Figure E.3e). However, in our research, the vesselness measure is normalised. Therefore the parameter c is less interesting to tune.

Table E.1: The different possibilities for the eigenvalues and the corresponding orientation patterns

| λ_1	λ_2	orientation pattern
Low	Low	noisy, no preferred direction
Low	High-	tubular structure (bright)
Low	High+	tubular structure (dark)
High-	High-	blob-like structure (bright)
High+	High+	blob-like structure (dark)

E.4. Results Frangi Filter

As mentioned before, the Frangi filter is a multiscale approach. The range of σ can be adjusted. The results of these adjustments are shown in Figure E.4. The lower values of σ contain the fine details of the image, while for larger values the structures become more coarse. Another interestings aspect, is that with an increase in the σ range, the dominant lines become wider (faded) as well.
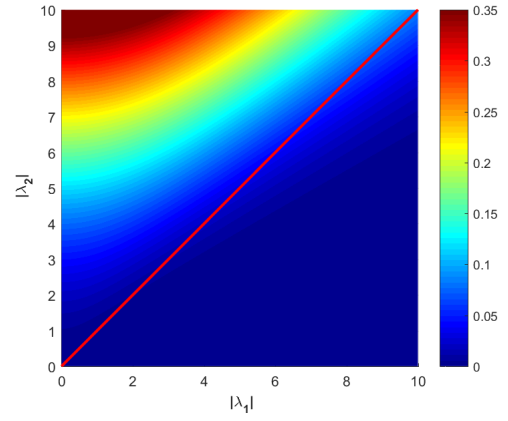
In our case, the lower bound of the σ range was set to 0, since we wanted to keep the stone structure. The upper bound was set to 10, so that the lines do not fade too much.
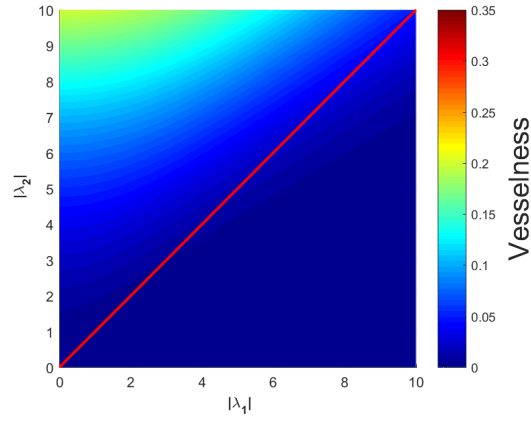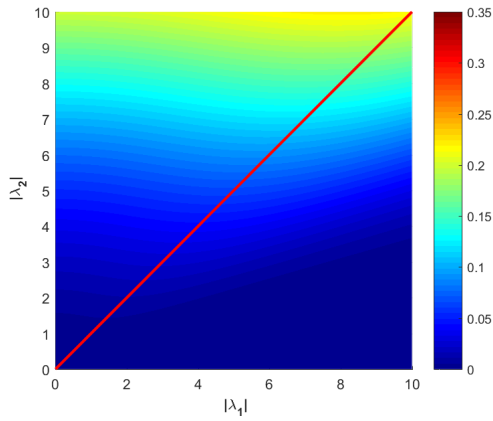
(a) Settings choses in the paper, $\beta = 0.5$ and $c = 15$. Note that the scale of the Vesselness measure is different than the paper.
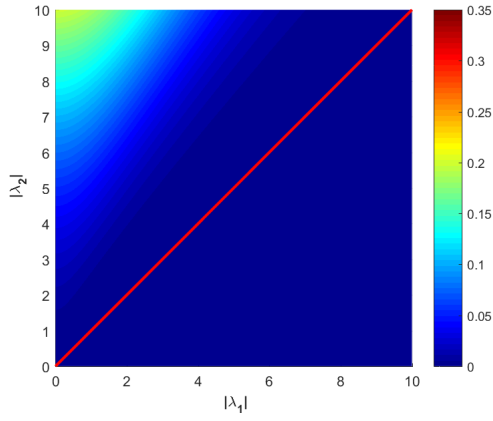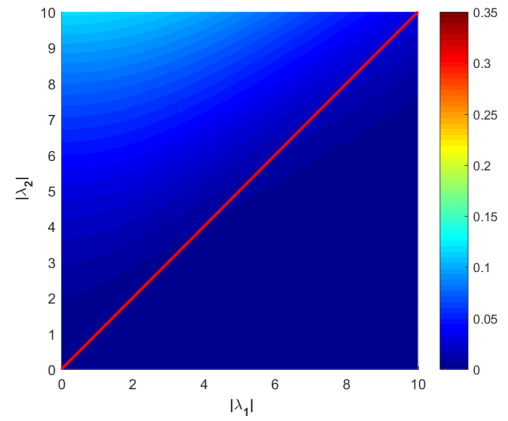

(b) $\beta = 0.25$ and $c = 15$
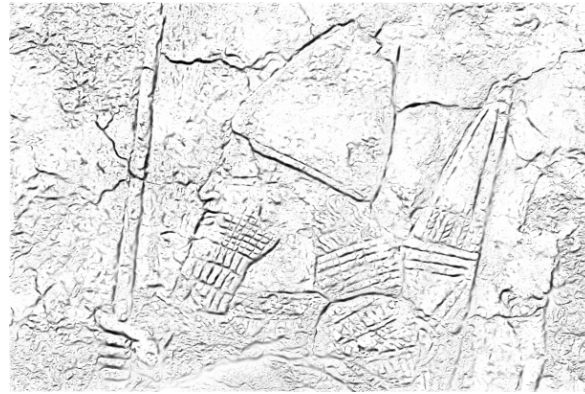

(c) $\beta = 0.5$ and $c = 10$

(d) $\beta = 1$ and $c = 15$

(e) $\beta = 0.5$ and $c = 20$

Figure E.3: Heatmaps of the Vesselness measure. On the left side β is tweaked. On the right side c is adjusted. The values below the red line are not considered: $|\lambda_1| \leq |\lambda_2|$

(a) Settings choses in the paper, $\sigma = 1 - 10$

(b) $\sigma = 1 - 5$
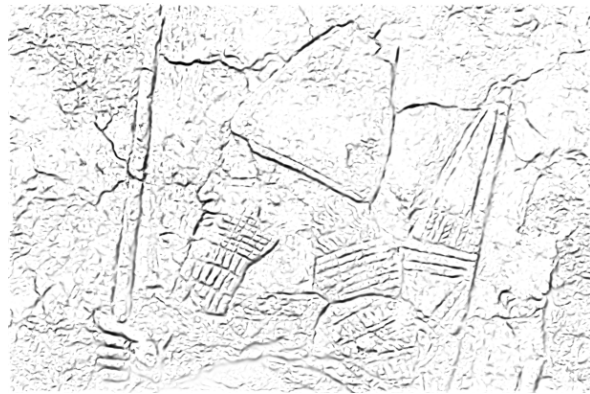
(c) $\sigma = 5 - 10$
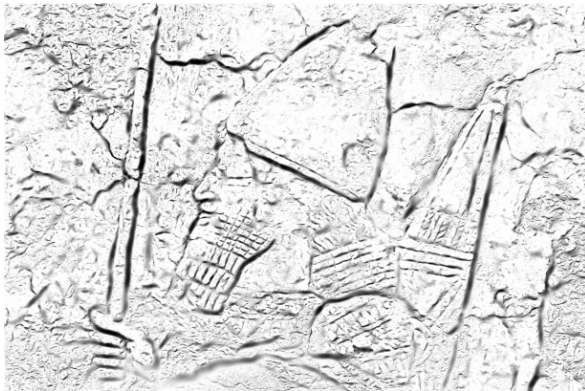
(d) $\sigma = 1 - 20$

(e) $\sigma = 10 - 20$

Figure E.4: Overview of the Vesselness for different ranges of σ. The range is in grayscale, where black is the highest value and white the lowest.

F
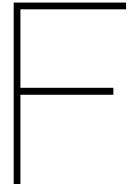
Evaluation

An overview of the tested reconstructions can be found in Table F.1. The names of the depth maps are according to the parameters used to construct that depth map. When only a N or A is given in the given, it is either a normalisation or amplitude.

Some of the results are displayed in Figure F.1, F.2 and F.3 as well.

Figure F.1: **a** to **c** are the original depth maps, which are not comparable due to the differences in the base. The scale is given in gray scale intensity from 0 - 255. The next row shows the transformed depth maps with our proposed measure, where 1 is the most convex structure in the image and -1 the most concave. The last row shows the RMSE per pixel in the form of a heat map. Where red corresponds with the largest error.

Figure F.2: **a** to **c** are the original depth maps, which are not comparable due to the differences in the base. The scale is given in gray scale intensity from 0 - 255. The next row shows the transformed depth maps with our proposed measure, where 1 is the most convex structure in the image and -1 the most concave. The last row shows the RMSE per pixel in the form of a heat map. Where red corresponds with the largest error.

(a) Depth map of gray

(b) Depth map of SFLRES

(c) Transformed Measure

(d) Transformed Measure

(e) Transformed Measure
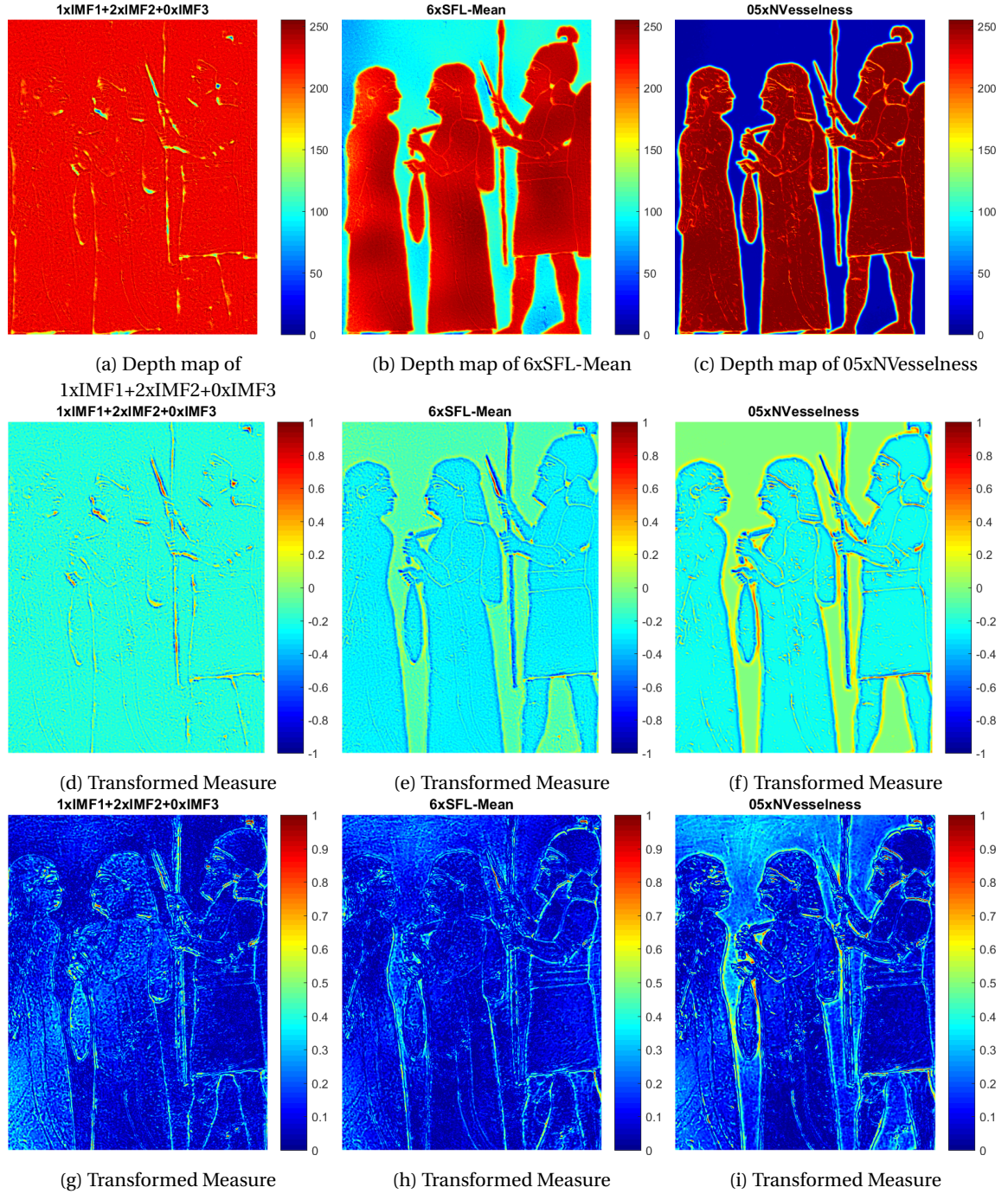
(f) Transformed Measure

Figure F.3: **a** to **b** are the original depth maps, which are not comparable due to the differences in the base. The scale is given in gray scale intensity from 0 - 255. The next row shows the transformed depth maps with our proposed measure, where 1 is the most convex structure in the image and -1 the most concave. The last row shows the RMSE per pixel in the form of a heat map. Where red corresponds with the largest error.
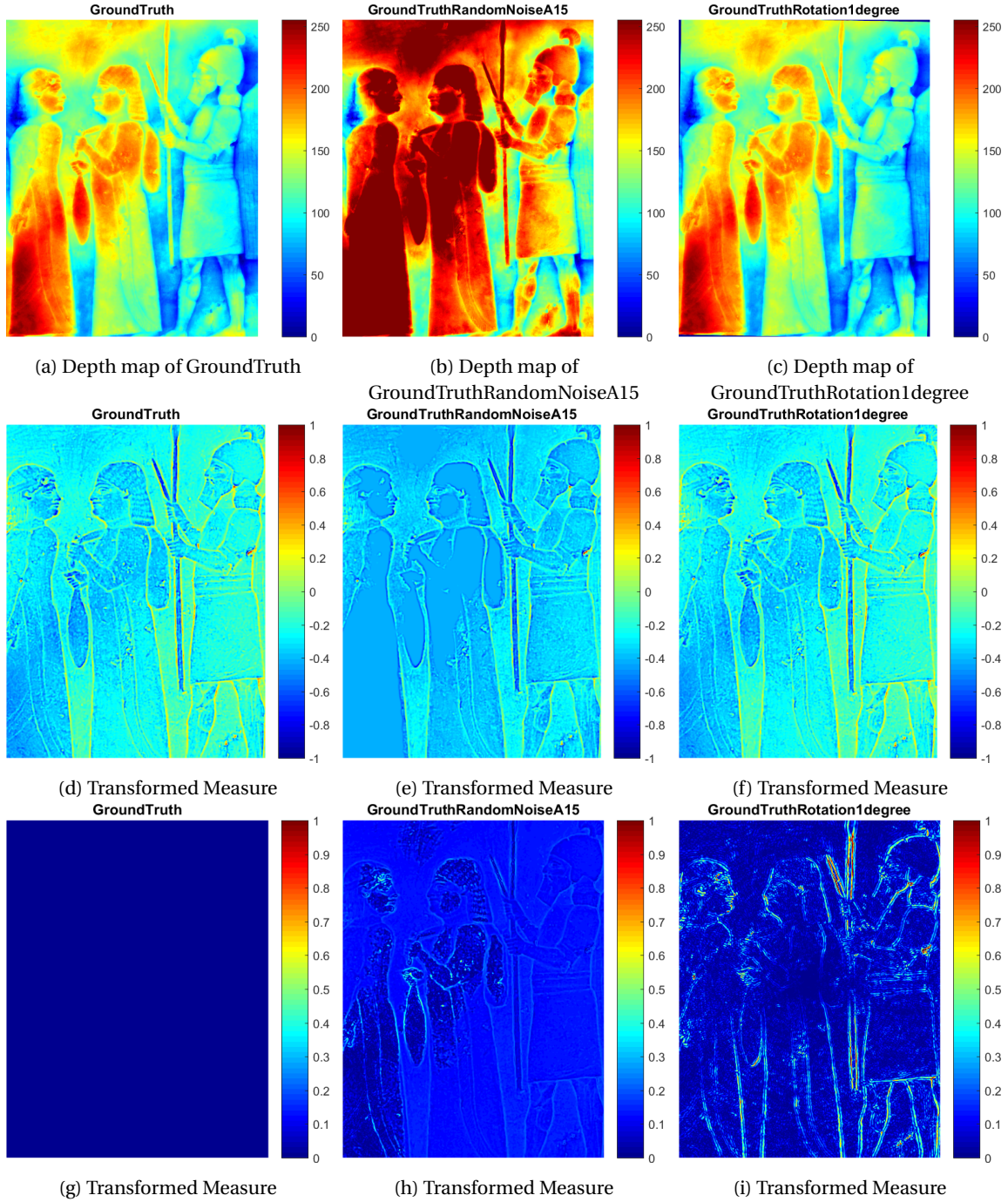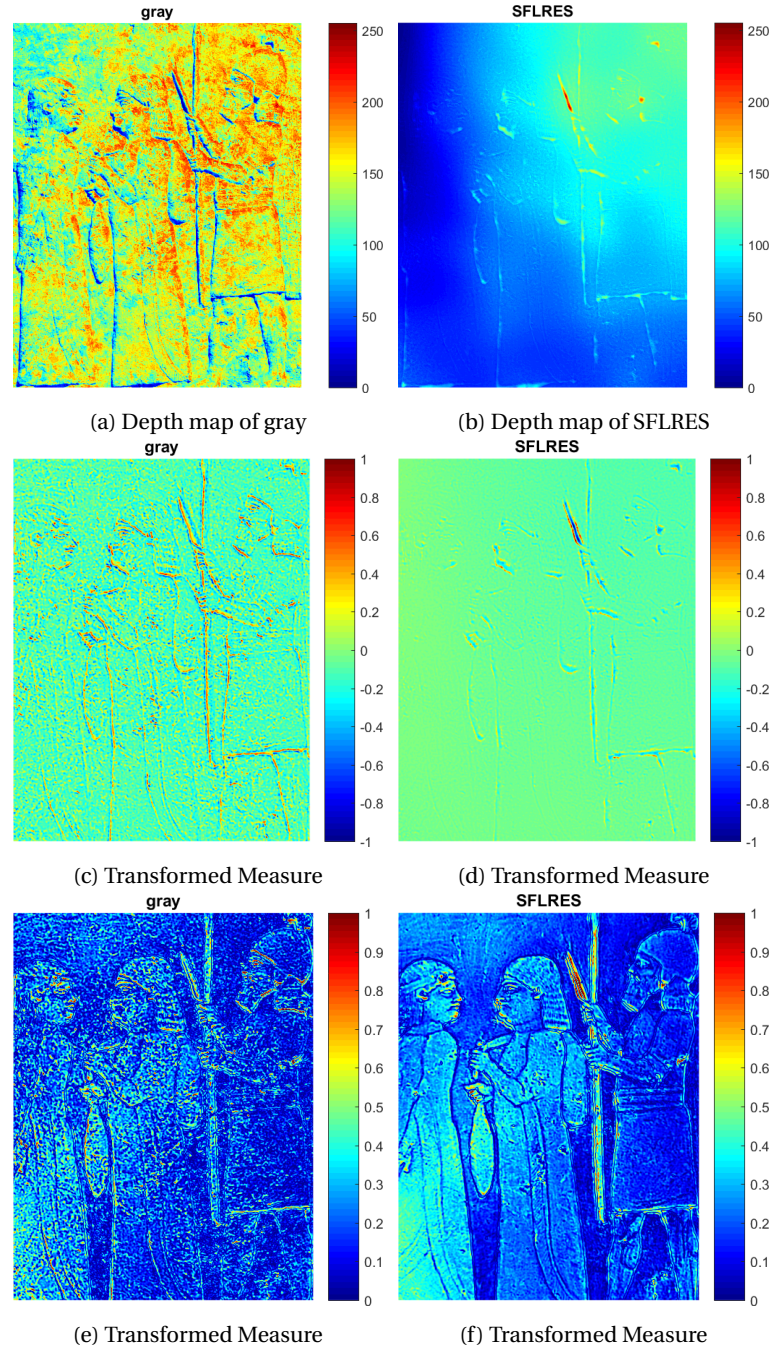
Table F.1: The RMSE over all pixels, $N_{\text{pixels}} = 4291872$. A lower RMSE is a better result. The lowest value defined in **bold**.

Type of Depth map	RMSE$_{\sigma=5}$	RMSE$_{\sigma=10}$
SFL		
1xIMF3	0.615	0.675
1xIMF2	0.270	0.195
1xIMF2+1xIMF3	0.309	0.207
1xIMF2+2xIMF3	0.393	0.278
2xIMF2+1xIMF3	0.161	0.180
1xIMF1	0.194	0.205
1xIMF1+2xIMF3	0.202	0.323
1xIMF1+1xIMF2	0.187	0.188
1xIMF1+1xIMF2+1xIMF3	0.162	0.198
1xIMF1+1xIMF2+2xIMF3	0.186	0.246
1xIMF1+2xIMF2	0.162	**0.162**
1xIMF1+2xIMF2+1xIMF3	**0.152**	0.165
1xIMF1+2xIMF2+2xIMF3	0.156	0.199
2xIMF1+1xIMF3	0.207	0.322
2xIMF1+1xIMF2	0.191	0.213
2xIMF1+1xIMF2+1xIMF3	0.213	0.203
2xIMF1+1xIMF2+2xIMF3	0.235	0.242
2xIMF1+2xIMF2+1xIMF3	0.179	0.176
2xIMF1+2xIMF2+2xIMF3	0.228	0.226
1xIMF1+2xIMF2+1xIMF3 + RES	0.763	0.8264
Artificial Reconstruction		
A1x3W1x10A2x2W2x6	0.202	0.197
A1x3W1x1A2x2W2x2	0.212	0.201
A1x3W1x2A2x2W2x6	0.225	0.203
A1x3W1x5A2x2W2x6	0.212	0.201
A1x5W1x10A2x2W2x6	**0.199**	0.196
A1x5W1x10A2x3W2x3	0.200	**0.195**
A1x5W1x10A2x4W2x5	0.203	0.197
A1x5W1x5A2x2W2x6	0.211	0.201
SFL Elevated		
4xSFL-Mean	0.165	0.166
5xSFL-Mean	0.155	0.157
6xSFL-Mean	**0.149**	**0.149**
Fusion		
05xNVesselness	**0.202**	**0.187**
05xNVesselnessComplement	0.213	0.192
06xNVesselnessComplement	0.215	0.194
Control Group		
Grey	0.255	0.268
Plane	0.797	0.868
Ground Truth		
GT	0	0
GT + 1° rotation	0.127	0.134
GT + Random Noise	0.133	0.134

G

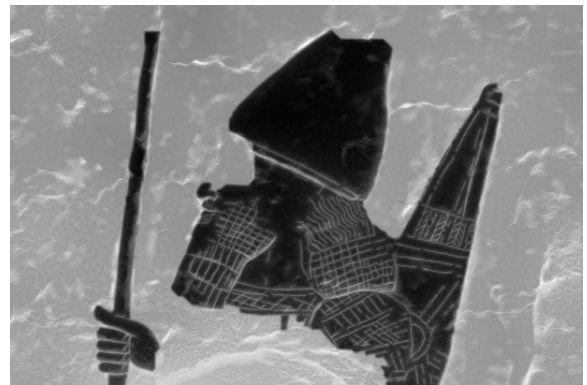# Results

In this research, we reconstructed 3 different reliefs:

- V34-9

- V37-13

- V44-3

The results are shown on the next pages. At the time of writing this report,the 3D print were not ready yet. However, the 3D renders give a good expression how the 3D shape would look like. V34-9 has been CNC milled and 3D printed with sandstone as well to compare the different reproduction techniques.

G.1. V34-9

(a) The original input photo.

(b) The Océ style height map.
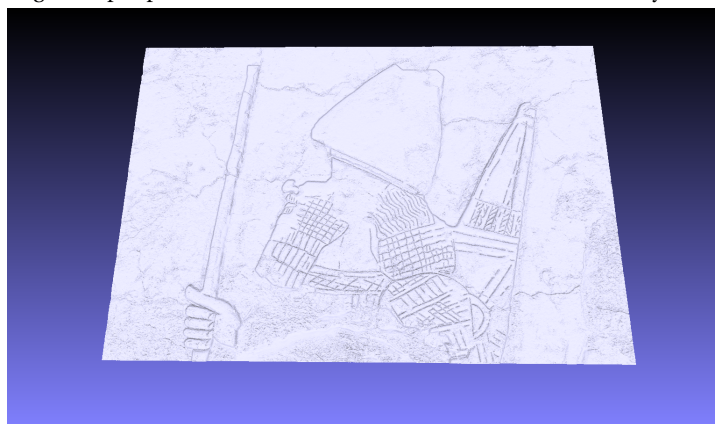
(c) A 3D render of the reconstruction

Figure G.1: Results for V34-9

G.2. V37-13

(a) The original input photo.

(b) The Océ style height map.

(c) A 3D render of the reconstruction

Figure G.2: Results for V37-13

G.3. V44-3

(a) The original input photo.

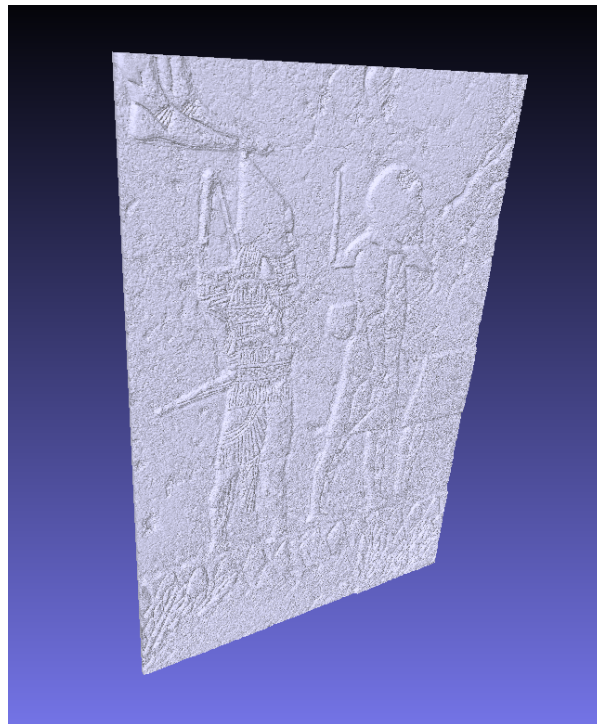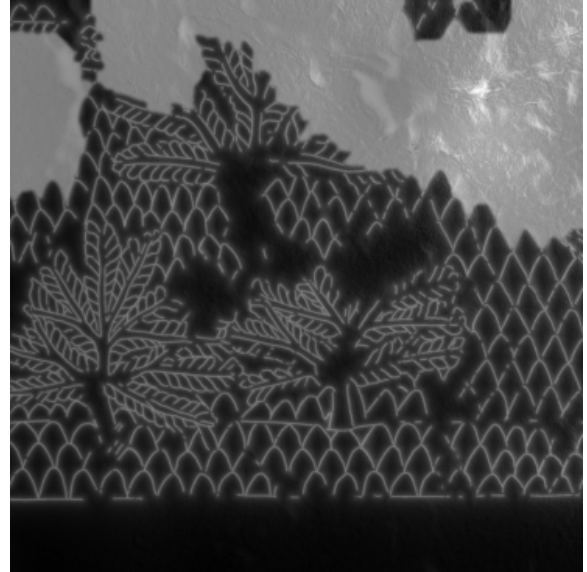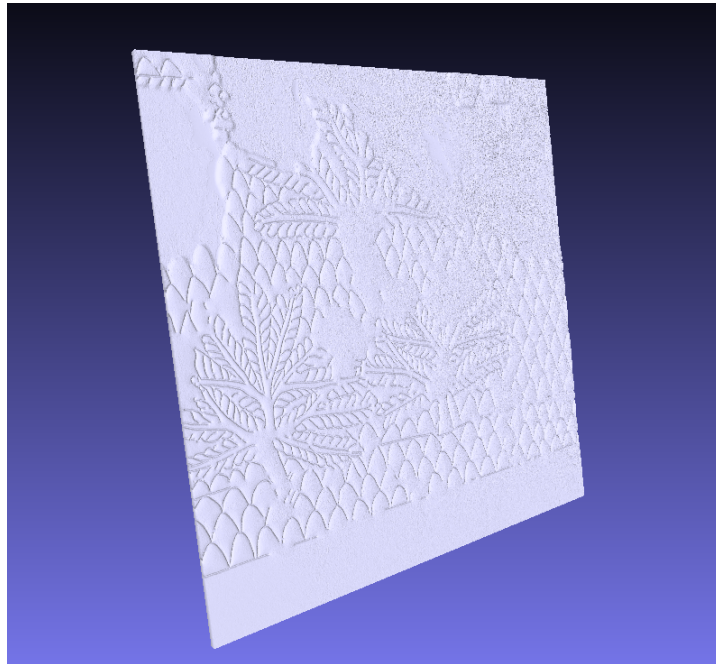(b) The Océ style height map.

(c) A 3D render of the reconstruction

Figure G.3: Results for V44-3

G.4. RMO

The RMO has the ambition to make reproduction of 7 slabs for their exhibition[1], resulting in a total surface of $23m^2$. The photo database is not sufficient to make the full reproduction, because of the natural wear and fire damage of the reliefs. The RMO chose to use several databases, as the old line drawings, pictures of the original excavation and photos of conserved reliefs in other musea [1]. This data was combined by experts of the RMO into a detailed line drawing. These line drawings are converted by a slightly adapted version of the artificial reconstruction in this research. The new method can handle multiple levels in one map and applies an artificial stone texture.
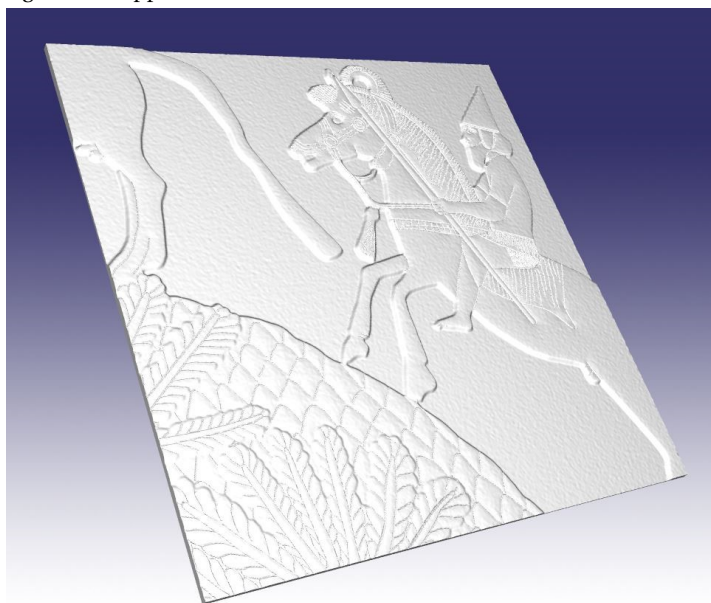
---

[1]`www.rmo.nl/english/exhibitions/nineveh`

(a) The detailed line drawing in combination with the user defined height map. Red is the lowest level, purple middle and green the upper level.

(b) The Océ style height map.

(c) A 3D render of the reconstruction

Figure G.4: Sneak peak of the reconstructions for the exhibition of the RMO

G.5. Tokyo University of Fine Arts

A collaboration with Tokyo University of Fine Arts[2] (Geidai) has been started during this research. Both universities (Delft and Tokyo) use technology to analyse art. However, where in Delft we try to replicate art with digital manufacturing techniques, Tokyo University still uses traditional techniques to recreate objects.

In this collaboration the Geidai will improve our reconstructions (Figure G.5) with their expertise about traditional sculpture techniques. The results are expected in September 2017.
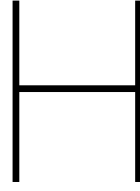
---

[2]`www.geidai.ac.jp/english/`

(a) V34-9, the head of a warrior.

(b) V51-3, plants, and fish in a river

Figure G.5: The two close-ups, which will be further reconstructed by Geidai. We send them our monocular reconstructions

H

Reproduction

This appendix is an addition to the Section Reproduction in the paper of this research.

H.1. Milling

The 3D CNC mill requires an STL file as input. Therefore we needed to convert the height maps into STL files. First we converted height map to a solid with *surf2solid.m* Appendix J. This function transforms the matrix into 2 columns with faces and vertices, which are the basis for a STL file. As input for this conversion, a grid of X, Y and Z (height map) coordinates is given in millimetres. Next, the faced and vertices are stored as a STL file with *stlwrite.m*

The last step is the removal of holes which are too small for the mill. We used a mill of 1.5 mm. Any hole smaller than this, will be 1.5 mm. The filtering of small holes has been done by the company and it part of the software package for CNC milling.

In Figure H.1, the milling set-up and its result are shown. Iranian lime stone has been chosen as milling material, which is comparable to the original lime stone used for the reliefs. The stone has aged yet, since this lime stone was "freshly" excavated and milled. This is the major cause of the difference in stone colour. However, the stone properties and the texture are very comparable to the original, which will create the opportunity of touching an Assyrian relief.
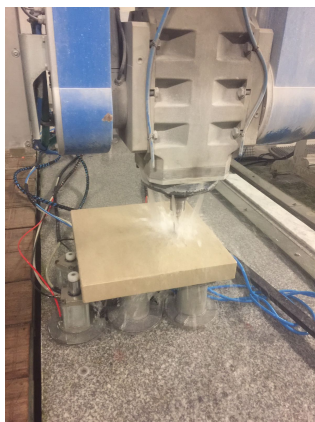
(a) Milling set-up

(b) CNC mill result of V34-9

Figure H.1: The limestone being CNC milled and the result.

H.2. Sandstone

The 3D printer of Z-corporation uses WRL files for full colour 3D printing. STL files do not contain colour information. The type of printer we uses is the Zprinter 650. For printing in gypsum powder there are some design rules to follow [11]. The minimum thickness of the relief is 2 mm. Therefore we create a thin slab of 2 mm with *surf2solid.m* of the relief. However, this slab will not be strong enough, due to its size. We reinforced the design by fusing the slab with a honeycomb. The honeycomb does not only strengthen the structure, it also reduces the production costs, since the design can be hollow. The honeycomb was designed in Solidworks 2016 and exported as STL file. The height of the honeycomb was 10 mm with a wall thickness of 2mm. This was strong enough to support the relief according to the faculty people of the KABK.

Next, Autodesk Meshmixer was used to fuse the thin slab with the honeycomb. The last step was to convert the STL file into a colour WRL file. In Autodesk 3DS Max the original RGB photo of V34-9 is projected on top of the STL file to create a colour file in WRL format.

After 3D printing the object needs to be excavated from the powder with a brush and a airbrush. It is just like in archaeology. The material is still very brittle at this moment and the colours are dull. The model is coated with a cyanoacrylate solution (superglue) to strengthen the whole model and enrich the colours. Figure H.3 shows the effect of this treatment, the left puppet is untreated.

At the time of writing this report the 3D print was not ready yet. Therefore, the final result is not shown in this report.
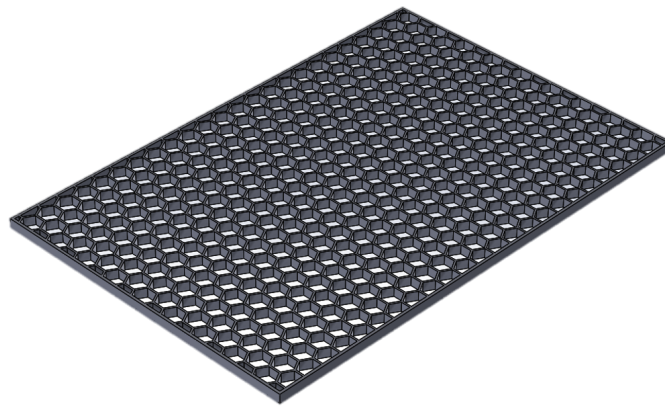

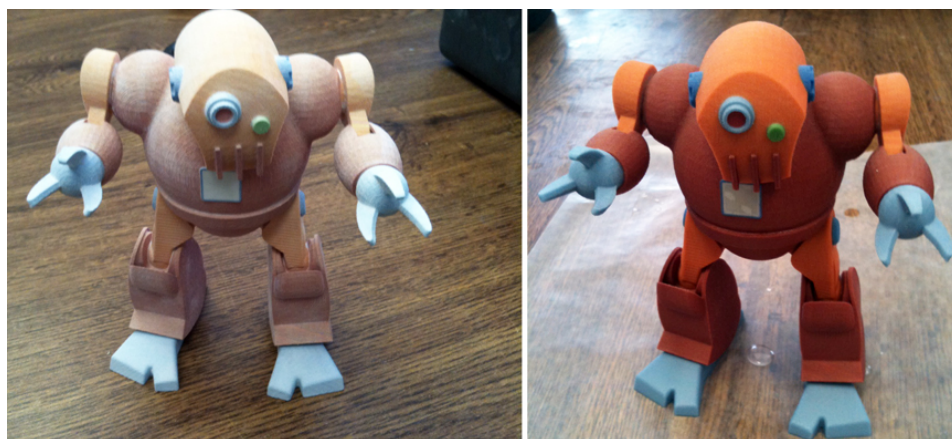Figure H.2: The honeycomb design in Solidworks

Figure H.3: Left before the cyanoacrylate coating and right after the treatment.

H.3. Elevated printing technology

The elevated printing technology requires a 16 bits greyscale height map, where black is defined as the top level and white as the bottom. The minimum adviced resolution for this technology is 300 DPI. Unfortunately, due to 6MP photos, our reconstructions are more in the range of 150 DPI. Before reproduction, the photo texture and height map are both resized to the resolution of 300 DPI with bicubic upsampling.

The elevated printing technology has a maximum range of 5 mm in z-direction. The difference between the lowest value (a crack) and the elevated areas can be up to 10 mm in our reconstructions. Therefore, the reconstructions need to be clipped within 5 mm.

This clipping can be done in multiple ways:

- Normalize between 0 and 5 mm.

- remove the background and only print the elevated areas in the correct scale.

- something in between the two methods above.

A test sample was printed to test each method (Figure H.4). The results look promising and the details of the Assyrian reliefs are incredibly well preserved, Figure H.5. We asked multiple people to touch and inspect the reliefs. People favoured the method of normalizing between 0 and 5 mm. This method is applied to the other reconstructions as well.

In contrast to the sandstone 3D print, the elevated printing technology does not require any honeycomb structure to increase the structural strength. The 3D object is directly printed on a strong sandwich material of aluminium and plastic. For practical reasons, the height map is pre-filtered with a Gaussian blur. The blur radius is 0.5 pixel. This filters the small noise, which can cause strange artefacts during printing.
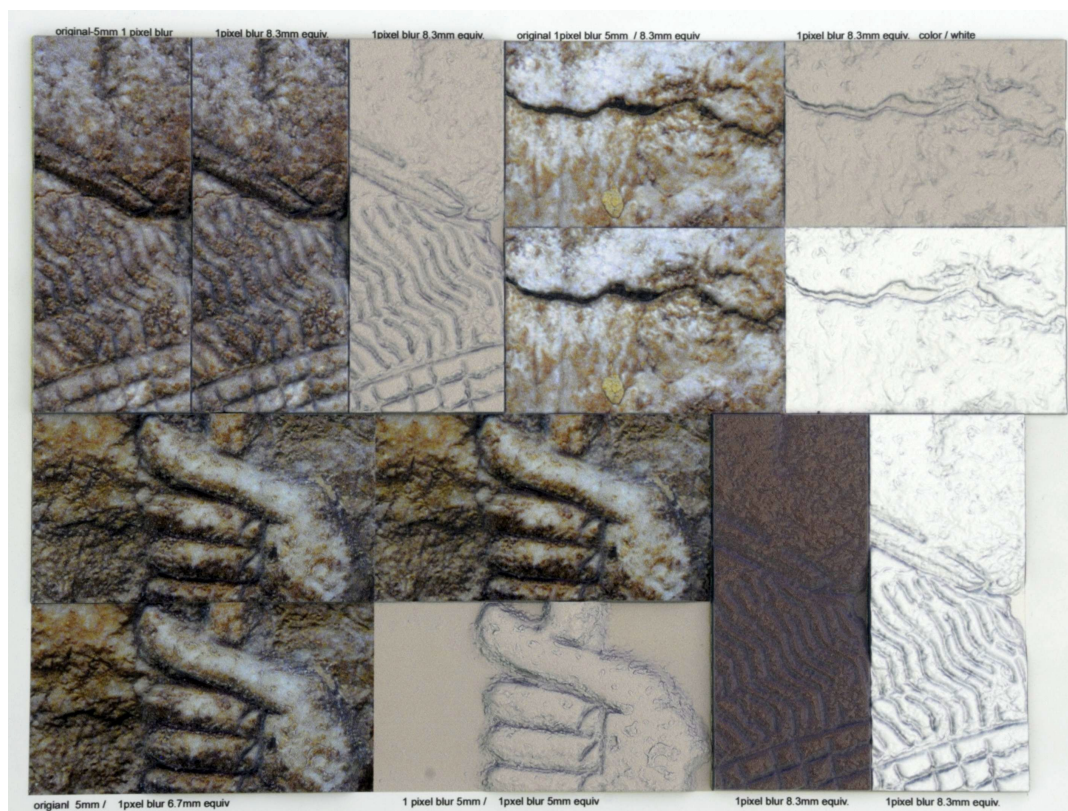
Figure H.4: Overview of the sample print. The colourless samples are chosen to purely show the shape, without colour manipulation. Dark spots seem to be deeper than they really are.
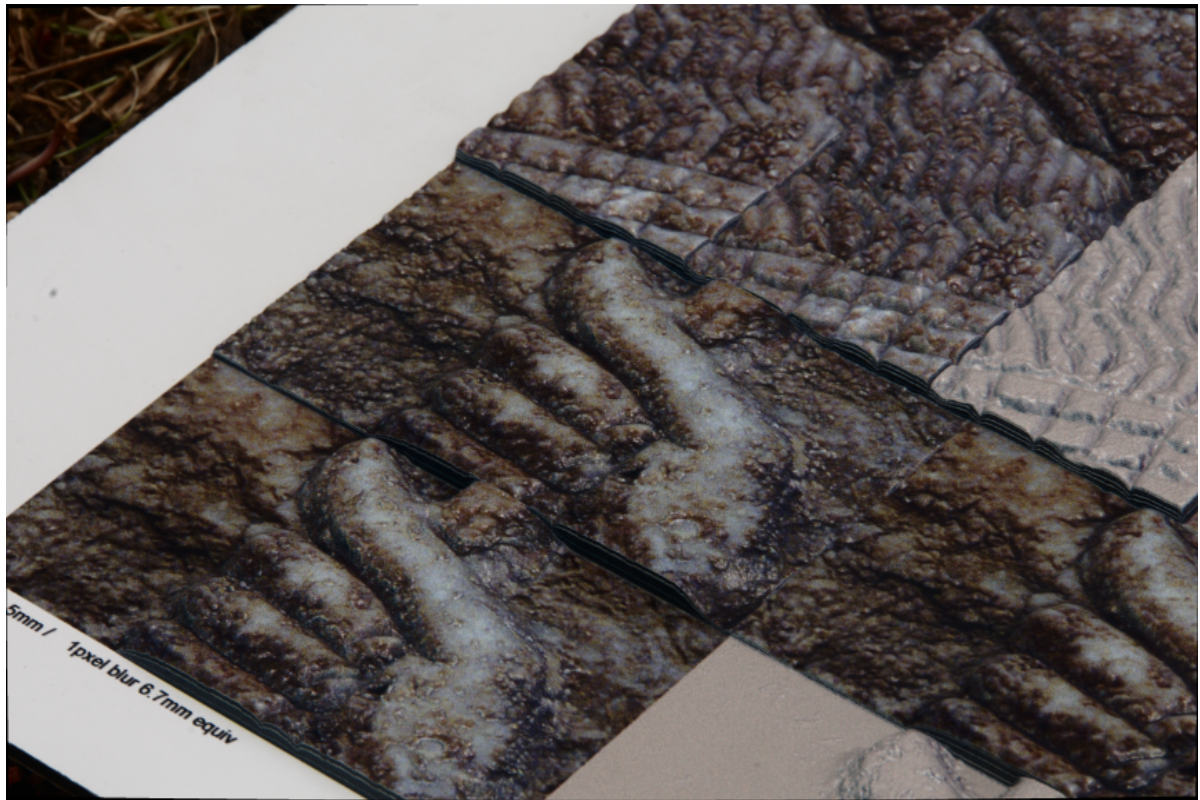
(a) A close up of the sample print. The hand shown is part of V34-9.
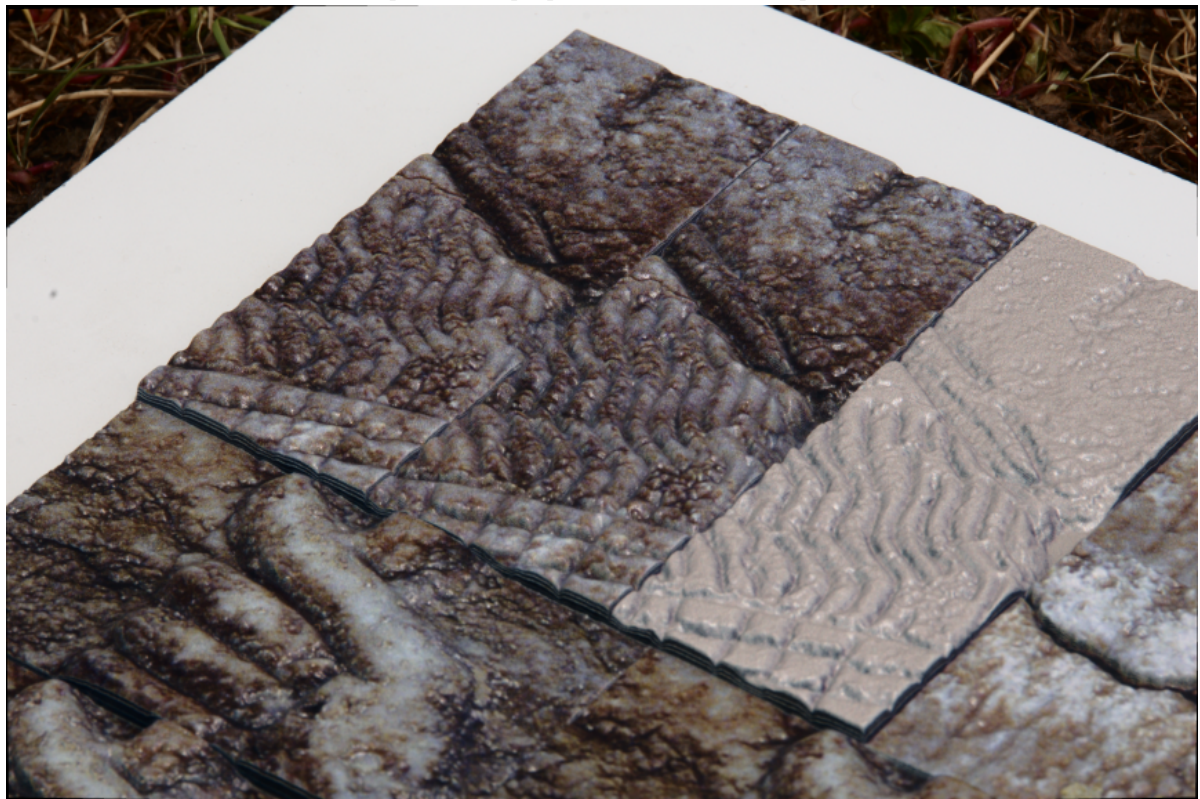
(b) More details (hair) of V34-9.

Figure H.5: SFL results of V37-13

H.4. Contact details

	STnatuursteen	Royal Academy of Art the Hague	Oce
	Kobaltstraat 27	Prinsessegracht 4	St. Urbanusweg 43
	Den Haag	Den Haag	Venlo
Tel:	070 323 82 92	070 315 47 77	-
Email:	info@stnatuursteen.nl	post@kabk.nl	info@projecteiger.com
web:	www.stnatuursteen.nl	www.kabk.nl	www.projecteiger.com

Multi Resolution Fusion

Sometimes there are multiple frontal pictures, with and without flash, or a difference in distance between the camera and relief. Multiple photos are used to improve the quality. Multi Resolution Fusion makes it possible to fuse multiple depth maps.

The different RGB photos are first converted into depth maps and then the depth maps combined, such that the original luminance signal in the RGB images is not distorted.

The current implantation of the monocular reconstruction method only supports dense 3D reconstructions, which means that every pixel should have a depth value. This causes problems if we want to combine multiple close-ups, Figure I.1, to reconstruct the whole relief, since the number of pixels (information) per distance unit varies between the photos. Every picture has a resolution of 6 MP, however the distance between the camera and the object fluctuates. This results in numerous resolutions once all the pictures are converted to the correct physical dimensions. Besides the amount of human effort it takes, you could try to manually make a collage of the different photos. However, it will be difficult to compensate for the non-linear distortion of the camera lens. Therefore we chose to automate this process.

Inspired by Brown and Lowe [7] with their solution for automatic panoramic image stitching, we use invariant features to align and stitch multiple photos. Without going into details about the computation of invariant features, you can describe a feature as a vector containing information about its corresponding location. On a more global scale, if we would describe the feature to distinct a specific car, we could say that this car is: red, has multiple exhausts, its height is small compared to its length, and has a yellow emblem with a horse in it. Guess which car it is... exactly a Ferrari!

In reality the vector would be more abstract, filled with numbers which describe the properties of the corresponding pixel and its neighbourhood. The challenge it to choose features which are unique for different properties and robust against image transformations (invariant). In our case we used the SURF features provided by the Matlab function: *detectSURFFeatures* and *extractFeatures*.

The work flow for image alignment is as follow:

1. Detect the SURF features, points of interest, in both images

2. Extract the features

3. Match the features between the images

4. Estimate the transformation matrix with the matched features

The SURF features are able to align the images used in this research, see Figure I.1c. The last step is to combine the aligned images. With the estimated transformation matrix from the original photos, we can transform the corresponding depth maps, obtained from monocular reconstruction method described earlier.

Since, we are restricted to dense matrices, the low resolution depth map has to be upsampled to match the same resolution as the detailed map. Finally, the overlapping data in the low resolution depth map, is replaced by the high resolution data, to prevent mixing with the noise of the low resolution data.

Unfortunately, some of the close-ups can not be directly aligned with the whole slab. This can be solved by aligning a more distant photo first. In Figure I.2, V37-13 is aligned with slab V37-8. While previously, V37-13 was already aligned with V37-5 (Figure I.1c). With this intermediate step, it is possible to align a close-up as V37-5 (Figure I.1a) indirectly with the whole slab.

Some of the photos in the database are illuminated with a flash. It is possible to align flash images with each other, Figure I.3. However, combining Flash and Non-flash images is not possible. This problem should still be solved. An idea was to use the IMF's for alignment, since their appearance is the same for flash as for the non-flash. Unfortunately, the number of SURF-features in the IMF's are too low and weak for alignment. Another option could be to use Mutual Information [13] [10]. Mutual Information is extensively used in medical image registration and holds promising results for multimodality (images from different sensors) as well [8]. Matlab 2017A contains an alignment toolbox, which provides the multimodality image registration as well. Unfortunately the alignment still failed. In the future we could focus more on the alignment of Flash and Non-flash images and investigate the use of mutual information with either the IMF's or Grey images more extensively.

(a) V37-5 High resolution   (b) V37-13 Low resolution   (c) Aligned

Figure I.1: Pictures **a** and **b** are two close ups of the same relief. By combining these photos, we can increase the number of details in the reconstruction as shown in **c**. In practice only the depth maps are aligned and the original photos are used to estimate the transformation matrix. In this example we fused the original photos for demonstration purposes.

(a) V37-13 High resolution   (b) V37-8 Low resolution   (c) Aligned
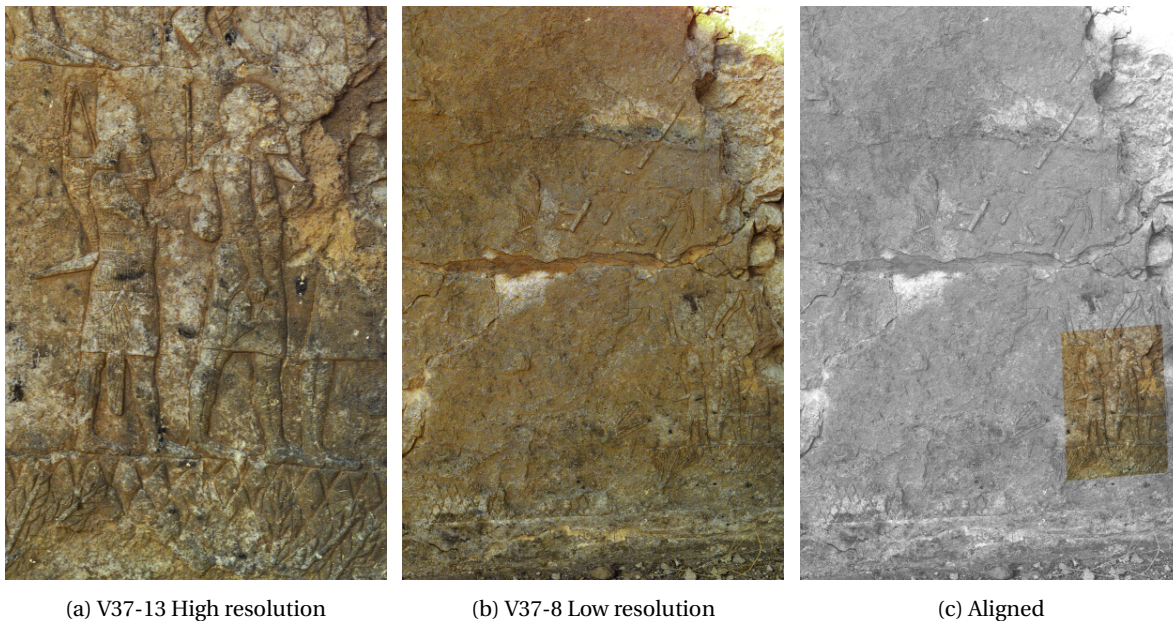
Figure I.2: Picture **a** is a close-up which we saw previously as well. **b** is the overview of the whole slab. By combining these photos, we add the details to the slab as shown in **c**.

(a) V37-15 High resolution   (b) V37-18 Low resolution
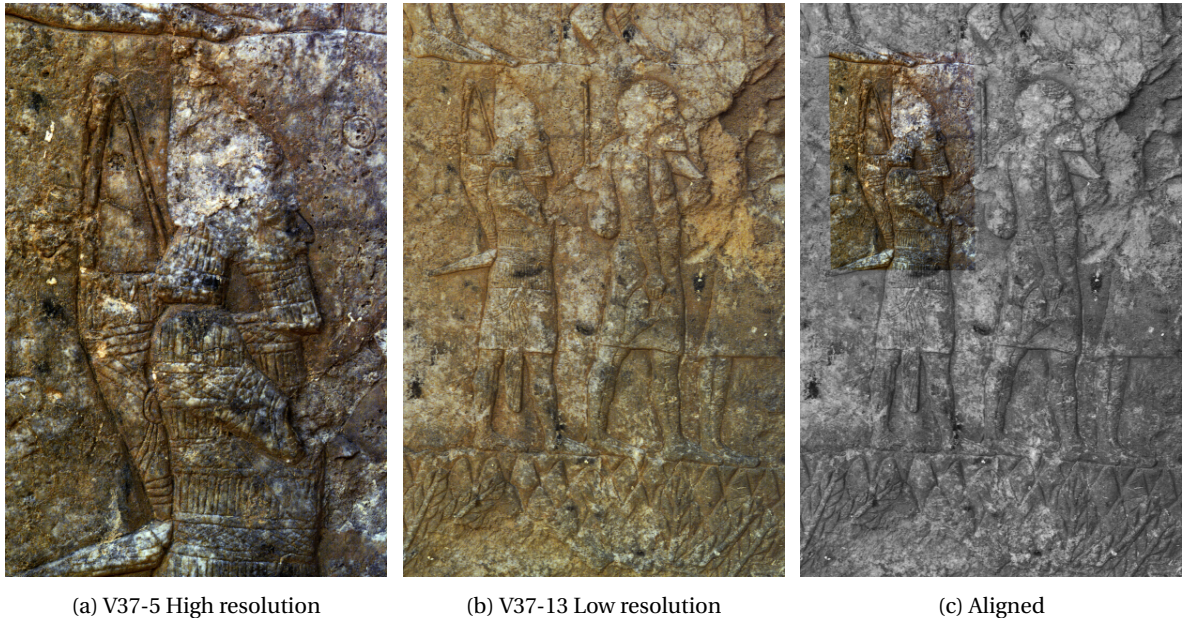

(c) Aligned

Figure I.3: Pictures **a** and **b** are two close ups of the same relief. By combining these photos, we can increase the number of details in the reconstruction as shown in **c**. In this alignment we used the pictures with flash.

J

Software Architecture

J.1. Flowchart

Schematic overview of the software for creating a reconstruction, Figure J.1.
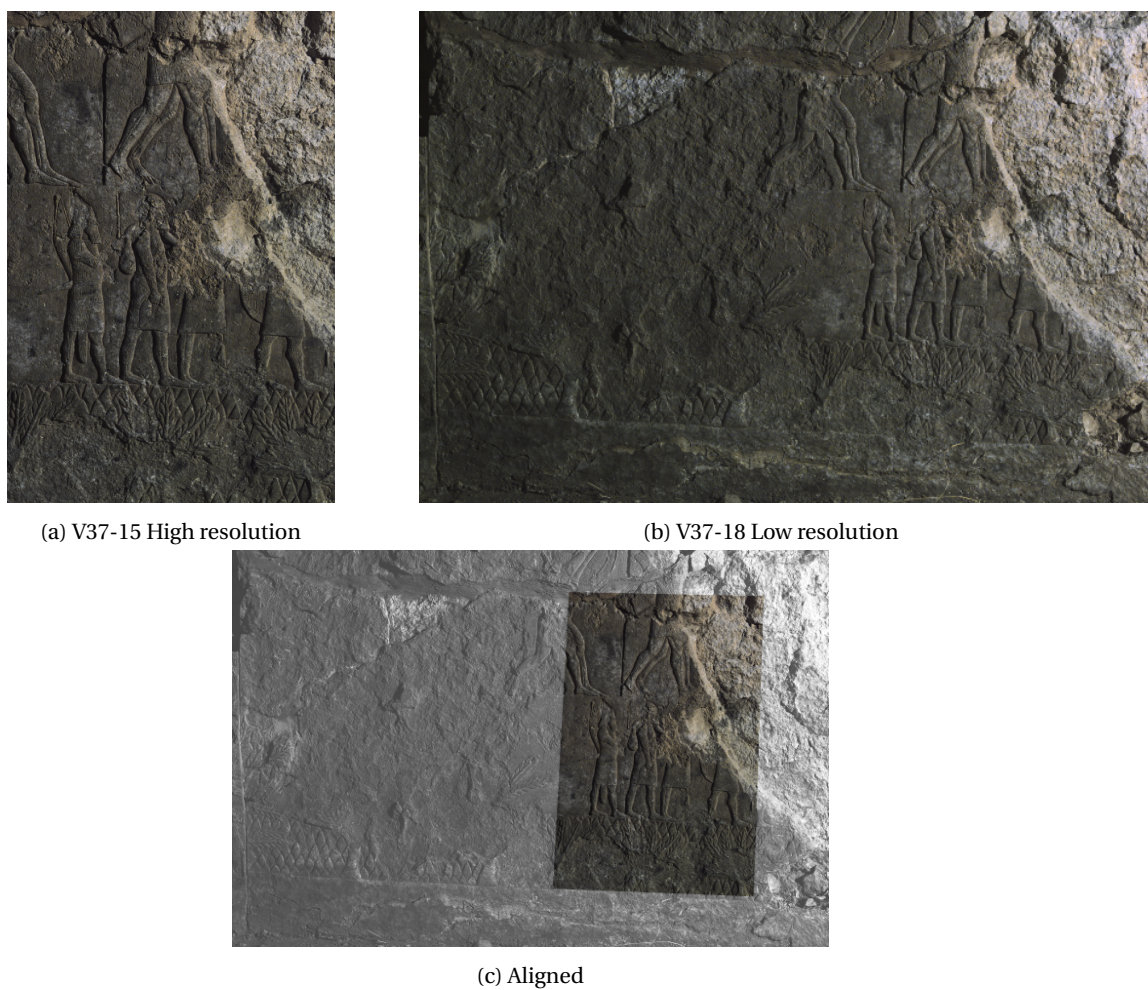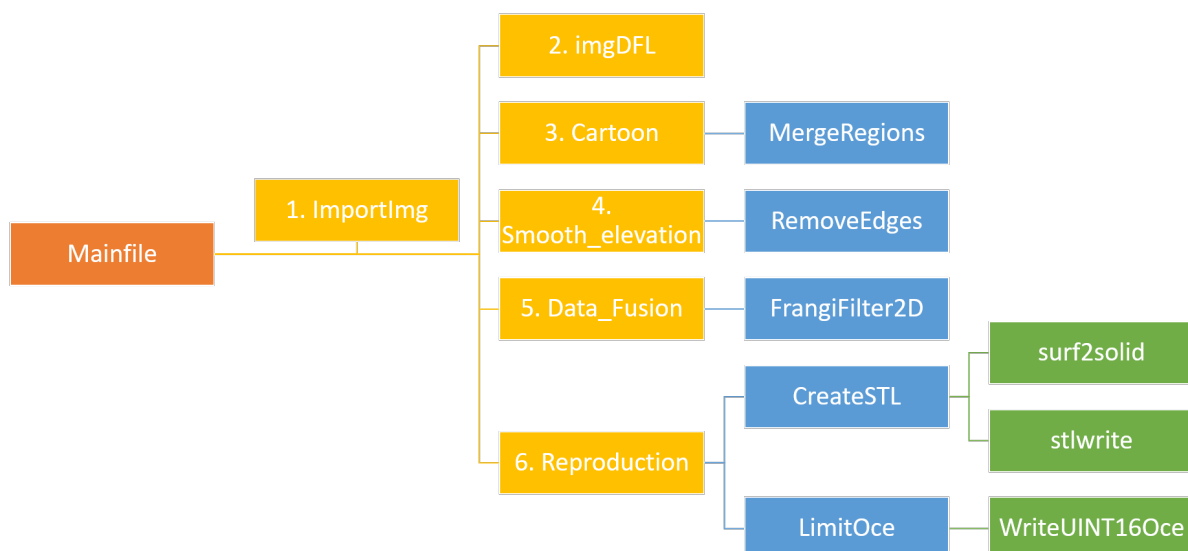
Figure J.1: An overview of the different Matlab functions

Mainfile:

- Imports the correct folders with all the functions.

- Calls the functions in the right order

- Variables to tune:

    – FileName, which photo to call
    – MM_Y, vertical length of the photo in mm.
    – Folder.RGB, location of all the photos and IMF's.
    – Folder.Line, location of the line drawings.
    – OptionArray1, which IMF's to use and their weights.

ImportImg:

- Imports all the correct images.

- Resizes all the images to the same size.

- Variables to tune:

    – X

imgDFL:

- Combines the IMF's and transforms these into a depth map.

- Variables to tune:

    – n, sensitivity level for the depth values.

Cartoon:

- Splits the line drawing into Lines and Cracks.

- The Euclidean distance per pixel to the closest line is calculated.

- The Euclidean distance is converted to greyscale and inverted.

- The inverted Euclidean distance is segmented with the Watershed algorithm of Matlab.

- The original watershed result is merged by **MergeRegions**.

- User should store

- Variables to tune:

    – distThreshold, minimum distance of the centroids for fusing certain section.

Cartoon.m is only called if there is no user-defined height map yet.

Smooth_elevation:

- Imports the user-defined height map.

- Calculate the Euclidean distance only for the contour lines.

- Convert the Euclidean distance in the elevated map with the step function.

- Obtain the details from the line drawing.

- Calculate the Euclidean distance for each pixel to the line drawing.

- Create a mask, only details are allowed on elevated areas.

- Create the detailed carvings.

- Substract the carvings from the elevated map.

- Variables to tune:

  – A1, amplitude of the elevated map.
  – W1, width of the height transitions.
  – A2, amplitude of the carvings.
  – W2, width of the carvings.
  – B, sharpness of the centre of the ridge.

Data_Fusion:

- Elevates the Shape from Luminance depth map with the artificial reconstruction.

- Calculate the Vesselness with **FrangiFilter2D**.

- Compute the confidence map for both the Artificial reconstruction as SFL.

- Variables to tune:

  – W_{mix}, the minimum confidence for SFL.
  – σ, the range of the sigma scale.
  – β, sensitivity parameter for blobs.
  – c, sensitivity parameter for the magnitude of the eigenvalues of the Hessian.

Reproduction:
For the reproduction there are two options: either store it as STL file (3D printing and CNC milling) or export it as a height map (Océ). **CreateSTL:**

- Create a grid of X and Y in the correct resolution. STL needs X, Y and Z value for every point.

- convert the surface plot into a solid with **surf2solid**.

- Store the solid as STL file with **stlwrite**.

- parameters to tune

    - D, assign the desired depth map for conversion.
    - Xa, resolution in x-direction.
    - Ya, resolution in y-direction.
    - Tslab, thickness of the solid slab.

LimitOce:

- Clips the depth values between a maximum range of 5 mm.

- Stores the height map as 16bits file according to Oc'e's standards with **WriteUINT16Oce**.

- parameters to tune

    - X

Acknowledgements

I would like to express my gratitude to Boris Lenseigne for his supervision during this project and professor Joris Dik and professor Pieter Jonker for facilitating this project.

Furthermore, I would like to thank Lucas Petit and Anna de Wit of Rijksmuseum van Oudheden for the context of this assignment.

Last but not least, I would like to thank my family and friends for supporting me throughout this project.

> "Strange things have been spoken, why does your heart speak strangely? The dream was marvellous but the terror was great; we must treasure the dream whatever the terror."
>
> The Epic of Gilgamesj, translation by N. K. Sandars

> "All photographs are accurate. None of them is the truth."
>
> Richard Avedon

]

Bibliography

[1] R. D. Barnett, E. Bleibtreu, G. E. Turner, and D. Collon. *Sculptures from the Southwest Palace of Sennacherib at Nineveh*. Trustees of the British Museum, 1998. URL https://books.google.nl/books?id=pflJAQAAIAAJ.

[2] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[3] Alejandro F. Frangi, Wiro J. Niessen, Koen L. Vincken, and Max A. Viergever. Multiscale vessel enhancement filtering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 130–137. Springer, 1998.

[4] Wilson S. Geisler. Mechanisms of visual sensitivity: backgrounds and early dark adaptation. *Vision Research*, 23(12):1423–1432, 1983.

[5] GOM. Atos triple scan - revolutionary scanning technique, 2017. URL http://www.gom.com/metrology-systems/atos/atos-triple-scan.html.

[6] Laurens van der Maaten Gorkem Saygili and Emili A. Hendriks. *Adaptive stereo similarity fusion using confidence measures Computer Vision and Image Understanding*. PhD thesis, 2015.

[7] Johannes Kopf, Michael F Cohen, Dani Lischinski, and Matt Uyttendaele. Joint bilateral upsampling. In *ACM Transactions on Graphics (TOG)*, volume 26, page 96. ACM, 2007.

[8] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on medical imaging*, 16 (2):187–198, 1997.

[9] Chris Mata. Hidden in plain sight?, 2017. URL https://www.youtube.com/watch?v=-NZakhhB_Do.

[10] Josien P. W. Pluim, J. B. Antoine Maintz, and Max A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE transactions on medical imaging*, 22(8):986–1004, 2003.

[11] Shape. Full color sandstone material information, 2017. URL https://www.shapeways.com/materials/full-color-sandstone.

[12] Vikko Smit. Literature review. TU Delft, December 2015.

[13] Paul Viola and William M. Wells. Alignment by maximization of mutual information. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 16–23. IEEE, 1995.