

Recovering and Visualizing Deformation in 3D Aegean Sealings

Bartosz Bogacz¹, Nikolas Papadimitriou², Diamantis Panagiotopoulos² and Hubert Mara¹

¹*Forensic Computational Geometry Laboratory, Heidelberg University, Germany*

²*Corpus der minoischen und mykenischen Siegel, Heidelberg University, Germany*

Keywords: Computational Reconstruction, Feature Extraction, Pattern Recognition, 3D Cultural Heritage.

Abstract: Archaeological research into Aegean sealings and sigils reveals valuable insights into the Aegean socio-political organization and administration. An important question arising is the determination of authorship and origin of seals. The similarity of sealings is a key factor as it can indicate different seals with the same depiction or the same seal imprinted by different persons. Analyses of authorship and workmanship require the comparison of shared patterns and detection of differences between these artifacts. These are typically performed qualitatively by manually discovering and observing shared visual traits. In our work, we quantify and highlight visual differences, by exposing and directly matching shared features. Further, we visualize and measure the deformation of shape necessary to match sigils. The sealings used in our dataset are 3D structured light scans of plasticine and latex molds of originals. We compute four different feature descriptors on the projected surfaces and its curvature. Then, these features are matched with a rigid RANSAC estimation before a non-rigid thin-plate spline (TPS) matching is performed to fine-tune the deformation. We evaluate our approach by synthesizing artificial deformations on real world data and measuring the distance to the re-constructed deformation.

1 INTRODUCTION

A seal is a small portable artifact mostly made of stone but also found in other materials, such as bone/ivory, metal, and various artificial pastes. It displays engraved motifs and is generally perforated so that it can be suspended e.g. on a necklace. Seals have played an important role in Aegean society and, as functional objects, they have served three main purposes: securing, marking, and authorizing. Their study provides important insight into the Aegean socio-political organization and administration¹.

One important question arising in this study is the determination of authorship and origin of seals. Given two or more visually similar but not equal seals, possibly of different provenance, do they originate from the same stamp or the same author? Research on sealing artisan craftwork typically is concerned with qualitative judgments based on manually discovered and

observed traits. Our quantifiable results and visualizations can serve as foundations for evidence-based reasoning. Thus, we address the challenge of analyzing patterns and deformations common to a pair sealings.

The contributions of our work are (i) a pre-processing and feature extraction pipeline for analyzing the similarity and unique features of 3D scans of sealings, (ii) the evaluation of this pipeline with synthetic data for estimating deformations between sealings, and (iii) the visualization of the deformation and differences between sealings for Archaeological applications.

This paper is structured as follows: In Section 2 we present related work on the challenge to generate correspondences between sets of image pairs of related but visually differing content. In Section 3 we introduce our dataset and the necessary pre-processing of the 3D data for further image processing. Then, in Section 4 we describe four feature extraction approaches to generate and weight correspondence hypotheses. In Section 5, based on the proposed correspondences, first a rigid mapping is esti-

¹Further introductory information can be found at <https://www.uni-heidelberg.de/fakultaeten/philosophie/zaw/cms/seals/sealsAbout.html>

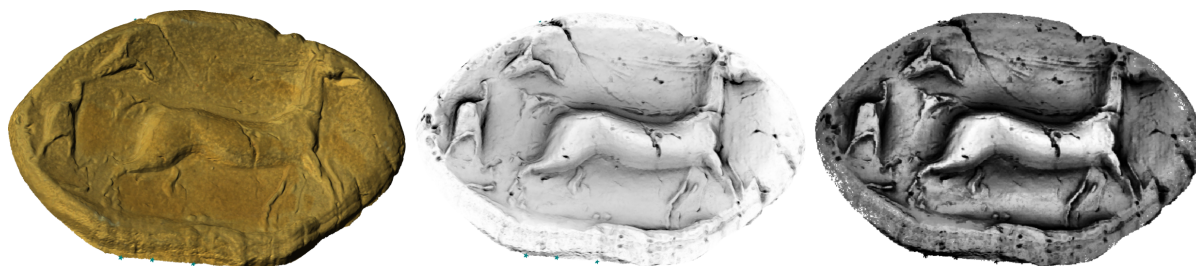


Figure 1: Three stages of preprocessing are shown from left to right. (i) The original scanned 3D data with texture and light. (ii) The surface curvature of the seal 3D data rendered into a raster image. (iii) The seal image after local histogram normalization and padding.

mated and hypotheses are filtered to match the estimated model. On the remaining hypotheses a non-rigid model is fine-tuned. In Section 6, we visualize and describe the application of our results in an Archaeological context. A summary and conclusion is given in Section 7.

2 RELATED WORK

Finding image correspondences and image registration is a challenging task usually approached by research in medical settings (Holden, 2008), Geosciences (Mongus and Žalik, 2012; Chen and Li, 2012; Tennakoon et al., 2013), and stereo matching (Tola et al., 2010). In these cases, the images being compared depict the same scene and vary only in small detail or pose. The large-scale content and identifying features are common to both images. However, in our setting, both sealing images differ in the depicted content and correspondence metrics must consider high-level semantic similarity.

In (Ham et al., 2016) an approach is presented for correspondence matching and image deformation based on a set of region proposals called *Proposal Flow*. The authors proceed in two steps to match two images. First, a set of multi-scale region proposals are generated and matched. The matching considers the the spatial support of the region pair, i.e. its geometric closeness, and its visual similarity, computed with Spatial-Pyramid Matching (SPM) (Lazebnik et al., 2006) or a convolutional neural network (CNN) (Krizhevsky et al., 2012). The spatial support is computed and weighted with neighboring matches in mind, i.e. support is low if neighbors disagree. This scheme enforces local smoothness of matches. Secondly, from the set of matches, the authors compute a dense flow field for each pixel partaking in a correspondence. The source image is then transformed according to the flow field to match the target image.

In (Rocco et al., 2017) the authors enhance the classical computer vision correspondence pipeline of extracting descriptors and estimating a transform by making it fully differentiable and end-to-end trainable. The feature extraction is handled by cropping the classification layer from a pre-trained CNN. Then, the feature maps of two images processed by these networks are correlated with a scalar product and passed to regression CNN that estimates rigid transformation parameters. This approach is repeated with a second regression CNN that estimates non-rigid parameters for a thin-plate spline (TPS) based transform. The main advantage of this approach is the ability for end-to-end training of the CNNs and the single step estimation of the rigid and non-rigid transformation parameters, i.e. no optimization is performed since the neural network predicts the parameters in one forward pass.

A typical approach for robust estimation is RANSAC. However, due to the adaptable complexity of TPS models a minimal subset of points cannot be defined, i.e. a minimal deformation can be found for any pair of point-sets. However, in (Tran et al., 2012) the authors show that one can reject outliers reliably by fitting a hyperplane in the feature space of correspondences. A separation exists between the distribution of inliers and the distribution of outliers.

In our approach, due to the nature of our data, i.e. historical artifacts, no training data is available. While following the classical approach of extracting descriptors, computing correspondences and estimating a transformation we also adopt the two-stage approach of Rocco et al. by first estimating a rigid transformation and then fine-tuning with TPS.

3 DATASET

The CMS project (Corpus der minoischen und mykenischen Siegel), with its archive of physical

sigils and sealings currently residing in Heidelberg, Germany, systematically documents and publishes all known ancient Aegean seals and sealings. The CMS consists of an physical archive, published volumes, and an open-access digital database². The database contains photos and tracings of seals and sealings with associated manually added meta-data in line with current Aegean glyptic research.

Within an interdisciplinary research project we are gathering hundreds of sealings of a collection holding approx. 12.000. We make use of the physical archive and the manufactured impressions of sealings in plasticine, silicon, and gypsum. The impressions of the sealings are their negative imprint. That way, the motif and detail can be discerned more clearly. The surface of the plasticine negatives, the sealing impressions, are acquired with 800dpi resolution using a structured-light 3D scanner. In this work we use the word sealings and sigils interchangeably as our approach is applicable to any artifacts of similar shape.

3.1 Pre-processing

The resulting 3D data is processed in GigaMesh³ by computing the surface curvature of the impressions. The computation proceeds using either Multi-Scale Integral Invariants (MSII) (Mara and Krömker, 2017; Mara, 2016) or by Ambient Occlusion (AO) (Miller, 1994). While MSII provides better details for small-scale surface features, for aligning sealing impression we use AO which provides smoother and medium-scale surface curvature.

The surface data augmented with surface curvature is projected into a raster image of 400×600 pixels extents. We then apply local pixel histogram scaling with a disk of 50 pixels, thus further emphasizing medium-scale surface curvature. Finally, the images are padded to provide the necessary space for subsequent deformation operations. The process from a 3D sealing with texture data to a pre-processed sealing raster image is shown in Figure 1.

4 CORRESPONDENCE

The underlying assumption of our work is that pairs of images under study, in our case impressions of sealings, share visually similar and semantically equivalent regions. If we then identify and align these region pairs, the deformation that has been applied to

one of the sealings will become apparent. This identification requires a discriminative distance metric between image patches. We evaluate four different distance metrics of increasing complexity.

In this work we define the left sealing image as the sealing being deformed to match the right sealing image. For reasons of brevity, in the following sections, we only give definitions of the feature descriptors for left sealing image $\{D,Y,Z,V,N\}_{f_i}$ and omit definitions for the features of the right sealing image $\{D,Y,Z,V,N\}_{g_j}$ as they are identical up to interchanged variables.

4.1 Direct Matching

In the baseline approach sub-regions are compared directly by pixel value using the Euclidean distance. For the pair of sealing images under study I, J we generate two sets A, B of key-points $a_i, b_j \in \mathbb{R}^2$ arranged in a grid with 60×60 grid points overlaid on the sealing images. At each key-point we extract sub-regions of the sealing images, the patches $p_i, q_j \in \mathbb{R}^{\alpha \times \beta}$ with height α and width β , centered around the key-points. By flattening these patches, we compute respective feature vectors ${}_D f_i, {}_D g_j \in \mathbb{R}^{\alpha\beta}$. Then the distance between the image patches ${}_D d_{ij}$ is given by the Euclidean distance between the image patches.

$${}_D d_{ij} = \|{}_D f_i - {}_D g_j\| \quad (1)$$

Since our image data depicts local surface curvature, the direct comparison of pixel values represents a direct comparison of curvature values of the original surfaces.

4.2 DAISY Descriptor

The DAISY image descriptor (Tola et al., 2010) is a reformulation of the SIFT (Lowe, 2004) and GLOH (Mikolajczyk and Schmid, 2005) descriptors efficiently computable for each pixel in an image. This is achieved by computing multi-scale histograms of oriented gradients only once per image region sharing these among neighboring pixels. Similar to the direct comparison image patches, we extract DAISY descriptors ${}_Y f_i, {}_Y g_j \in \mathbb{R}^{\delta\gamma}$ with δ orientations and γ rings, centered at the key-points a_i, b_j .

$${}_Y d_{ij} = \|{}_Y f_i - {}_Y g_j\| \quad (2)$$

The distance between the feature descriptors is computed using the Euclidean distance.

4.3 BOVW Descriptor

We aggregate locally bounded sets of DAISY features into feature vectors to better capture and de-

²<https://www.uni-heidelberg.de/fakultaeten/philosophie/zaw/cms/databases/databasesfull.html>

³<https://gigamesh.eu>

scribe repeating higher-order visual patterns. Such an approach has been first introduced as Bag-of-Visual-Words (BOVW) in (Fei-Fei and Perona, 2005).

We denote the union of both sets of DAISY descriptors ${}_Y f_i$ and ${}_Y g_j$ as $E = \{{}_Y f_i\} \cup \{{}_Y g_j\}$. These descriptors are then quantized into the sets z_{f_i} and z_{g_j} . The quantization dictionary is computed with k-means (MacQueen, 1967) which minimizes the following term to find a set of visual words $v_k \in V$. The size of the dictionary, i.e. the size of the set $|V|$ is chosen as κ .

$$\operatorname{argmin}_V \sum_{v \in V} \sum_{e \in E} \|e - v\| \quad (3)$$

Then, for each DAISY descriptor ${}_Y f_i$ and ${}_Y g_j$ we find the closest visual word v_k and record them in the scalars $z_{f_i}, z_{g_j} \in \{1 \dots \kappa\}$.

$$z_{f_i} = \operatorname{argmin}_k \|{}_Y f_i - v_k\| \quad (4)$$

The radius of such a patch of visual words centered at a key-point is given by θ .

$${}_v f_i = \begin{pmatrix} |\{z_{f_i} = 1 \wedge \|a_i - a\| < \theta \mid \forall a \in A\}| \\ \vdots \\ |\{z_{f_i} = \kappa \wedge \|a_i - a\| < \theta \mid \forall a \in A\}| \end{pmatrix} \quad (5)$$

The BOVW feature vectors ${}_v f_i, {}_v g_j \in \mathbb{N}^\kappa$ count the occurrences of visual words z_{f_i}, z_{g_j} close to a key-point a_i, b_j .

$${}_v d_{ij} = \|{}_v f_i - {}_v g_j\| \quad (6)$$

Comparison is performed using the Euclidean distance between the counts of visual words.

4.4 DenseNet Descriptor

To capture abstract visual concepts and meaningful objects in the sealing images, e.g. heads, arms, or legs, we employ CNNs. A CNN captures increasingly abstract concepts and patterns that are used to describe and classify an image.

An approach to use CNNs as feature extractors without a labeled dataset for the target domain is introduced in (Razavian et al., 2014). By pre-training the CNN with a common large natural image dataset with ground-truth labels, common visual patterns are learned. The authors make use of this high-level pattern description as means for embedding images into a feature-space by removing the last fully-connected classification layer.

Given all but the last fully-connected layer of a network, which we denote as the function W :

$\mathbb{R}^{H \times W} \rightarrow \mathbb{R}^{1000}$. The dimensionality of the target domain \mathbb{R}^{1000} is the count of classes used in the training dataset. The features ${}_N f_i \in \mathbb{R}^{1000}$ of the image patch p_i, q_j are computed as follows:

$${}_N f_i = W(p_i) \quad (7)$$

Then, the features vectors of two image patches are compared using the Euclidean distance.

$${}_N d_{ij} = \|{}_N f_i - {}_N g_j\| \quad (8)$$

The network W uses the DenseNet architecture (Huang et al., 2017) trained on the ImageNet (Deng et al., 2009) dataset.

4.5 Evaluation

For our purposes the suitability of a image descriptor is given by two properties: (i) the correctness of semantic equivalence, e.g. a point on the horse head on the left sealing is assigned to a point on the horse head on the right sealing, and (ii) by the uniqueness of the assignment, i.e. a point of the left image is assigned to dense unimodal concentration on the right.

We visually evaluate the performance of the four presented image patch comparison methods by sampling points-of-interest from the left sealing and visualizing the descriptor distances on the right sealing as shown in Figure 2.

We experimented with different values for image patch sizes, DAISY parameters and counts of visual words. The following values were chosen based on the quantitative evaluation performed in Section 5.3. We used a image patch size of $\alpha = 128, \beta = 128$ pixels for all descriptors but the DenseNet based approach, which used the architecture native patch size of 224×224 pixels. For both of the DAISY based descriptors we used $\delta = 8$ orientations and $\gamma = 4$ rings with a radius of 100 pixels for the pure DAISY descriptor and 30 pixels for BOVW descriptor. Finally, we used clustered the DAISY descriptors into $\kappa = 512$ distinct visual words for the BOVW description.

Based on Figure 2, we see that the direct patch comparison (a) performs well when large-scale features are compared, i.e. the back of a horse, but fails at uniquely identifying features with fine detail, such as the head of a horse. Since pixel values are compared directly, even small misalignments cause significant differences between two image patches. The DAISY feature descriptor (b) and the BOVW feature descriptor (c), are robust against small offsets and perform significantly better with small and large feature structures. Finally, the CNN (d) feature descriptor does neither capture semantics of the image well nor is the assignment concentrated in one place. We assume

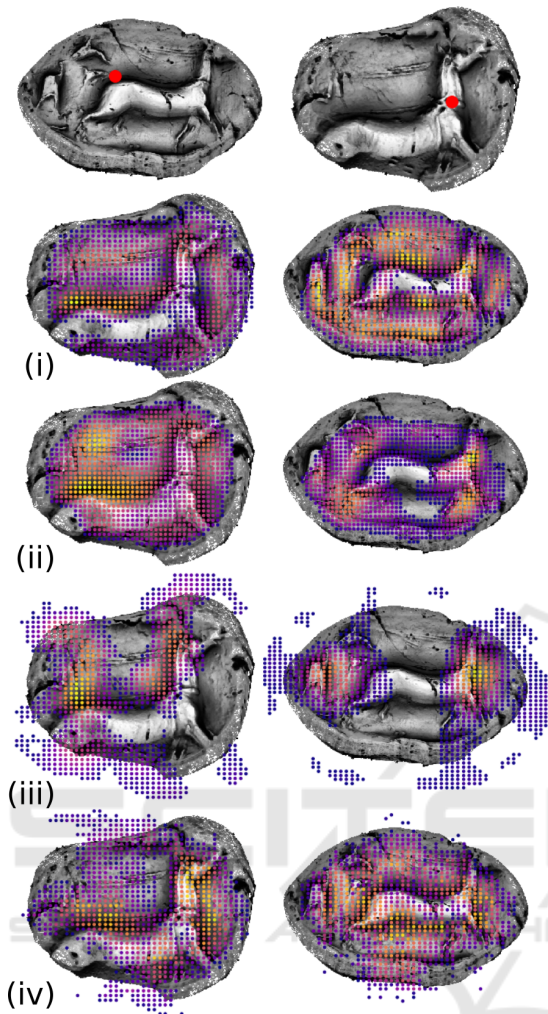


Figure 2: Point query responses of the four presented descriptors. Given a point on a left sealing image, points on the right sealing image are shown that are closest in feature space. Brighter colors indicate less distance, i.e. more similarity. From top to bottom: Original image with query point in red, (i) Euclidean distance of patches, (ii) DAISY feature descriptor, (iii) Bag-of-Visual-Words, and (iv) DenseNet pre-trained on ImageNet.

that our dataset, patches of a filtered and normalized surface curvature of a mesh, are significantly outside the distribution of the training dataset and the learned visual patterns are therefore not transferable.

5 ALIGNMENT

After computing the visual distance between image patches, or circular regions in the case of BOVW, at grid key-points a_i and b_j , we have to find unique correspondences. We compute the matches using a greedy approach by determining for each left feature

vector f_i the visually closest right feature vector g_j and vice versa. If either disagrees, such a match is discarded. Then, assuming a generic distance metric of features descriptors d_{ij} , the set of matches is denoted by $m_{ij} \in \{0, 1\}$ and is computed as follows.

$$m_{ij} = \begin{cases} 1 & \text{if } \operatorname{argmin}_i d_{ij} = \operatorname{argmin}_j d_{ij} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

For each match in m_{ij} the left visual descriptor is closest to its right counterpart and vice versa.

5.1 Rigid Pre-alignment

We align the sealing images in two stages. First, a rigid alignment model is estimated and only correspondences matching that model within a small error ϵ are kept. In a second stage, a non-rigid fine-tuning fits the remaining correspondences to minimize the previously allowed error.

We estimate the rigid transformation model using the RANSAC algorithm (Fischler and Bolles, 1980). Let $R \in \mathbb{R}^{2 \times 2}$ denote a rotation matrix in Euclidean space and let $v \in \mathbb{R}^2$ denote a translation. Then, the RANSAC algorithm minimizes the squared error between the correspondences m_{ij} of the key-point sets a_i and b_j .

$$\min_{(R,t)} \sum_i^M \sum_j^N m_{ij} \| (Ra_i + v) - b_j \|^2 \quad (10)$$

Given the rigid transformation model (R, t) estimated with RANSAC, we discard correspondences m_{ij} that do not fit the estimated model outside the aforementioned threshold ϵ .

$$\bar{m}_{ij} = \begin{cases} m_{ij} & \text{if } \| (Ra_i + v) - b_j \| < \epsilon \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

The remaining correspondences \bar{m}_{ij} are used for the subsequent non-rigid fine-tuning. With $\bar{a}_i \in \bar{A}$ and $\bar{b}_j \in \bar{B}$ we denote grid points which are only part of the set of remaining correspondences, a subset $\bar{A} \subset A$ and $\bar{B} \subset B$ of the original points.

5.2 Non-rigid Fine-tuning

The non-rigid transformation is based on a direct estimation of a thin-plate spline model closely following the approach presented in (Chui and Rangarajan, 2003). However, since we already computed correspondences \bar{m}_{ij} , we skip the expectation maximization procedure and skip outlier detection to directly estimate the model.

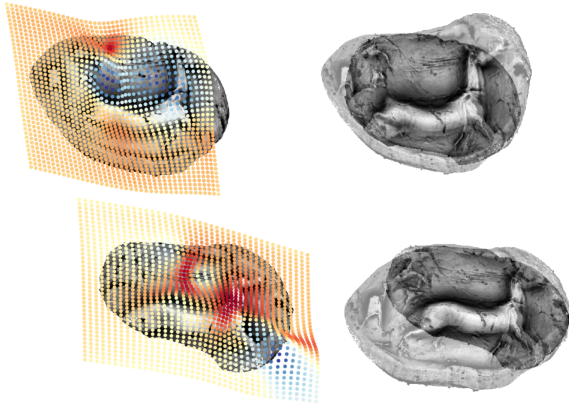


Figure 3: Predicted deformation visualized with a colored grid on the deformed left sealing image. Hotter colors on the deformation grid indicate a higher amount of compression and cooler colors indicate a higher amount of stretching. Then, deformed left and original right sealing images overlaid with 50% opacity.

The TPS model minimizes the following equation to find a transfer function $t_{d,w} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ to map matching source key-points a_i to target key-points b_j . The operator L follows the definition in the work of Chui et al., it is a double integral of the square of the second order derivatives of the mapping function, and is used as regularization. The parameter λ controls the strengths of the regularization and enforces the chosen smoothness.

$$\min_{t_{d,w}} \sum_i \sum_j \|\bar{a}_i - s_{d,w}(\bar{b}_j)\| + \lambda \|Lt_{d,w}\| \quad (12)$$

We follow the procedure outlined in Chui et al. work to minimize the term and compute the parameters d and w . The parameter $d \in \mathbb{R}^{3 \times 3}$ denotes a rigid transform matrix in homogeneous coordinates, the parameter $w \in \mathbb{R}^{M \times 3}$ denotes the non-rigid spline transformation parameters. The TPS kernel Φ contains information of the spatial relationships of the control points, the left grid points of the remaining correspondences.

$$\Phi = \|\bar{a}_i - \bar{a}_j\|^2 \log \|\bar{a}_i - \bar{a}_j\| \quad (13)$$

The mapping function for some homogenous point $x \in \mathbb{R}^3$ is written as follows:

$$t_{d,w}(x) = x * d + \Phi * w \quad (14)$$

Given the computed parameters d and w , we now can transform pixel positions of the left sealing image to semantically meaningful positions in the right sealing image.

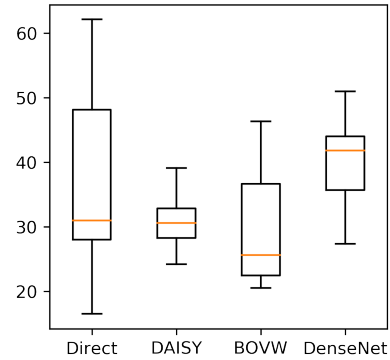


Figure 4: Mean distance between the vertices of the synthesized deformation grid and the respective vertices of the predicted deformation grid.

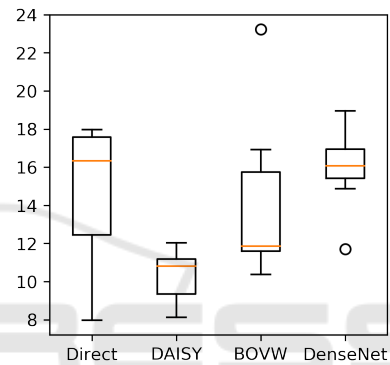


Figure 5: Mean pixel-wise difference between synthesized deformed image, i.e. the target image in the evaluation test, and the estimated deformation applied to the source image.

5.3 Evaluation

The nature of the dataset and research question posed in this work does not admit to a quantifiable ground-truth of expected alignments. Nevertheless, to be able to draw conclusions from our research, we evaluate the precision of our deformation estimation approach by synthesizing artificial deformations.

Deformation synthesis is based on the same TPS modeling as our estimation procedure. We manually create a grid of 12 points on the left sealing image which are randomly perturbed. We chose a low count of points, since natural deformation of sealings, e.g. due to heat or mis-handling, are large-scale.

Let c_i denote a manually chosen grid point. Then \tilde{c}_i is the respective point perturbed by a uniform distribution in $[-50, 50]$. Using these sets of points, we estimate a synthetic transform (d, w) . The pixel positions of the right sealing image are computed by sampling the right sealing image on the position indicated by the transfer function.

We evaluate our approach by comparing the align-

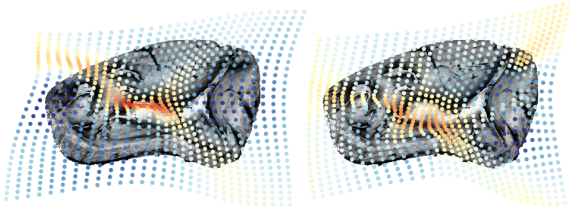


Figure 6: Left: Synthesized expected deformation grid. Right: Estimated deformation grid based on the presented approach using Bag-of-Visual-Words (BOVW) feature extraction. Hotter colors indicate compression, cooler colors indicate stretching.

ment predicted by our approach to the synthesized ground-truth alignment. We set the non-rigid deformation regularization to $\lambda = 1000$ and the allowed rigid model error to $\epsilon = 30$. Other values, such as $\lambda \in [1, 100, 10000]$ and $\epsilon \in [10, 100]$ resulted in worse performance. Given the initial grid of key-points a_i of the left sealing and the predicted transformation parameters (\hat{d}, \hat{w}) , we compute the deformed grid \hat{a}_i .

$$\hat{a}_i = t_{\hat{d}, \hat{w}}(a_i) \quad (15)$$

Then, the error ${}_{\text{grid}}E$ to the expected transformation \tilde{a} , the artificially synthesized transformation, is computed by the sum of euclidean distances between the respective grid points.

$${}_{\text{grid}}E = \frac{1}{|A|} \sum_i \sum_j \|\hat{a}_i - \tilde{a}_i\| \quad (16)$$

We compare the prediction performance of the presented feature extractors and alignment stages, rigid and non-rigid, by comparing their prediction error to the expected transformation. A lower prediction error indicates a better prediction performance.

Additionally, we consider the sum of pixel level differences ${}_{\text{pixel}}E$ between the deformed left sealing image I and the original right sealing image J . The set X , with $|X|$ the count of elements in the set, gives the pixel positions common to both images, i.e. when both images are overlaid on top of each other as shown in Figure 3.

$${}_{\text{pixel}}E = \frac{1}{|X|} \sum_{x \in X} |I(t_{\hat{d}, \hat{w}}(x)) - J(x)| \quad (17)$$

We consider the error to be predictive of alignment performance. The pixel values of the sealing images denote the local curvature of the original 3D scanned surface, they do not represent texture and lighting. Therefore, no undue biases are introduced by such a comparison. Figure 4 and Figure 5 compares the prediction performances over 10 repeated synthesized deformations of the presented descriptors.

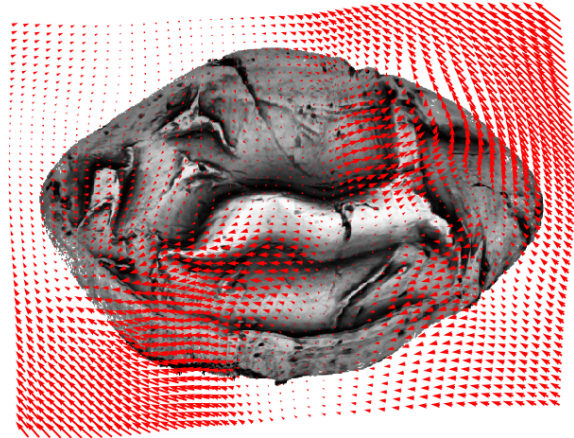


Figure 7: Vector field plot of the differences between the synthesized expected deformation grid and the estimated deformation grid. Red arrows are positioned on estimated grid points and point toward expected grid points. Most of the prediction error is concentrated outside of the sealing image, where no correspondences are present for a deformation estimation.

5.4 Visualization

In addition to the evaluation metrics, we visualize the computed transformations for visual inspection. In particular, we are interested in the non-rigid deformation of the transformation. We transform a regular grid of points with the predicted (d, w) and colorize the grid points by the difference of distances e_i to their neighbors, weighted by the Gaussian distribution, before and after transformation.

$$e_i = \sum_j \sum_k (\|\hat{a}_j - \hat{a}_i\| - \|a_j - a_i\|) e^{-\frac{\|a_j - a_i\|}{\sigma}} \quad (18)$$

The value σ controls the fuzziness of the coloring of the deformations and is set to $\sigma = 0.01$. The resulting visualization is shown in Figure 6. We also visualize the difference between the predicted grid points and the expected grid points using a quiver plot. At each grid point of the predicted grid an arrow points in the direction of the respective expected grid point, as shown in Figure 7.

6 RESULTS

Our work provides two directly applicable results for historical research: (i) the grid visualizations of deformation, as shown in Figure 3, necessary to align two seal images in either direction, (ii) overlays of deformed and target seal images, as shown in Figure 8 and in Figure 9, highlighting concrete differences.

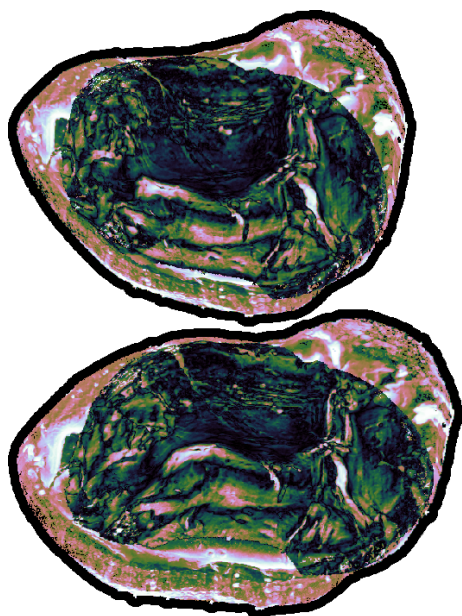


Figure 8: Visualization of pixel-wise differences for both deformation directions. For both configurations of left sealing and right sealing, the deformed left sealing image has been overlaid on top of the right original sealing image. Brighter colors indicate higher pixel-wise differences.

Our overlay visualizations provide a smooth and natural, i.e. differentiable and regularized deformation, embedding of a deformed seal image over the target image. Medium-scale differences can be easily spotted and further investigated. The overlay provides a view uncluttered from non-informative small differences due to scanning noise and damage.

If the characteristic influences of damage or manufacturing handicraft to the shape of the seals and the resulting deformations are known, e.g. by examples or by real-world experiments, the deformations estimated in our approach serves as quantitative evidence substantiating expert hypotheses.

7 SUMMARY & OUTLOOK

Our work provides an approach for the need of quantifiable research results and evidence gathering in the analysis of authorship and origin of Aegean seals and sealings. In particular, we are interested in commonalities and differences in shape and expressions of visual features. However, due to different manufacturing technique or damage to the material, the artifacts may be deformed. We approach this challenge by 3D scanning the seals and sealings and pre-processing the data to arrive at a high-contrast 2D images of the surface curvature. Then, we evalu-

ate four visual feature descriptors, direct comparison of image patches, Bag-of-Visual-Words (BoVW), DAISY descriptors and pre-trained DenseNet, with respect to their ability to find semantically meaningful correspondences. Given the set of correspondences, we first estimate a rigid transform with RANSAC and proceed with fine-tuning residual alignment error with Thin-plate splines (TPS). We evaluate the accuracy of the alignment by generating synthetic deformations of our dataset and comparing the expected deformation to the estimated deformation. Thus, we avoid the need for ground-truth that is often not available for historical artifacts. Finally, we visualize our results with a grid of the estimated deformation and colored by its curvature and an overlay highlighting differences of aligned images.

In future work, we are interested in physically manufacturing ground-truth of characteristic deformations by creating copies of chosen seals and sealings and then intentionally damaging and deforming those. This would allow us to analyze and then automatically classify the type of damage inflicted, by estimating the induced deformations and comparing these to the manufactured ones. Further, the presented pipeline is amendable to a fully 3D setting, where projecting into 2D raster images is unnecessary, by using 3D surface feature descriptors (Bogacz and Mara, 2018; Fey et al., 2018).

ACKNOWLEDGMENTS

We genuinely thank and greatly appreciate the efforts of Maria Anastasiadou supervising the “Corpus der minoischen und mykenischen Siegel” (CMS). We sincerely thank Markus Kühn for his contributions to our tooling, Katharina Anders for her feedback on related work, and ZuK 5.4 and BMBF eHeritage II for partially funding this work.

REFERENCES

- Bogacz, B. and Mara, H. (2018). Feature Descriptors for Spotting 3D Characters on Triangular Meshes. *International Conference on Frontiers in Handwriting Recognition*.
- Chen, C. and Li, Y. (2012). A robust method of thin plate spline and its application to DEM construction. *Computers & Geosciences*.
- Chui, H. and Rangarajan, A. (2003). A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical

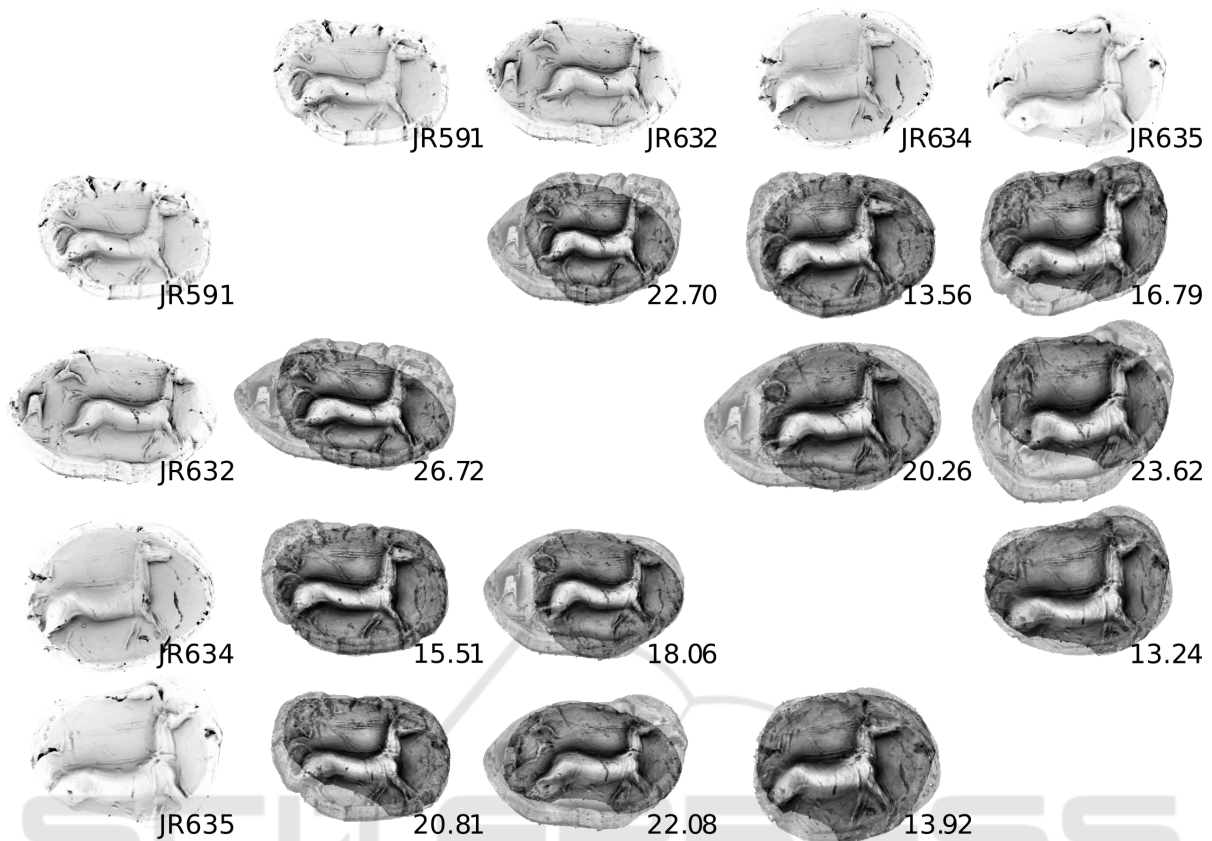


Figure 9: Pairwise comparison of four supposedly identical sealings in our dataset. Comparison is not symmetric as in one case the left sealing is being deformed and in the other case the top sealing is being deformed. Shown are additive sealing overlays where the left sealing is deformed to match the top sealing. Subject pose is therefore constant in columns and varying in rows. The feature descriptor used is DAISY. Mean of pixel-wise differences, with each pixel in $[0, 255]$, is given on the bottom left of the respective overlaid sealing pair.

- Image Database. *Computer Vision and Pattern Recognition*.
- Fei-Fei, L. and Perona, P. (2005). A Bayesian Hierarchical Model for Learning Natural Scene Categories. *Computer Vision and Pattern Recognition*.
- Fey, M., Lenssen, J. E., Weichert, F., and Müller, H. (2018). SplineCNN: Fast Geometric Deep Learning with Continuous B-Spline Kernels. *Computer Vision and Pattern Recognition*.
- Fischler, M. A. and Bolles, R. C. (1980). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *SRI International*.
- Ham, B., Cho, M., Schmid, C., and Ponce, J. (2016). Proposal Flow. *Computer Vision and Pattern Recognition*.
- Holden, M. (2008). A Review of Geometric Transformations for Nonrigid Body Registration. *Transactions on Medical Imaging*.
- Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *Computer Vision and Pattern Recognition*.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*.
- Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. *Computer Society Conference on Computer Vision and Pattern Recognition*.
- Lowe, D. (2004). Distinctive Image Features from Scale Invariant Keypoints. *Computer Vision*.
- MacQueen, J. B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Berkeley Symposium on Mathematical Statistics and Probability*.
- Mara, H. (2016). Made in humanities: Dual integral invariants for efficient edge detection. *Journal on Information Technology*.
- Mara, H. and Krömker, S. (2017). Visual Computing for Archaeological Artifacts with Integral Invariant Filters in 3D. *Eurographics Workshop on Graphics and Cultural Heritage*.
- Mikolajczyk, K. and Schmid, C. (2005). A Performance

- Evaluation of Local Descriptors. *Pattern Analysis and Machine Intelligence*.
- Miller, G. (1994). Efficient Algorithms for Local and Global Accessibility Shading. *Computer Graphics and Interactive Techniques*.
- Mongus, D. and Žalik, B. (2012). Parameter-free ground filtering of LiDAR data for automatic DTM generation. *Journal of Photogrammetry and Remote Sensing*.
- Razavian, A. S., Azizpour, H., Sullivan, J., and Carlsson, S. (2014). CNN Features off-the-shelf: an Astounding Baseline for Recognition. *Computer Vision and Pattern Recognition*.
- Rocco, I., Arandjelović, R., and Sivic, J. (2017). Convolutional Neural network architecture for geometric matching. *Computer Vision and Pattern Recognition*.
- Tennakoon, R. B., Bab-Hadiashar, A., Suter, D., and Cao, Z. (2013). Robust Data Modelling Using Thin Plate Splines. *Digital Image Computing: Techniques and Applications*.
- Tola, E., Lepetit, V., and Fua, P. (2010). DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo. *Pattern Analysis and Machine Intelligence*.
- Tran, Q.-H., Chin, T.-J., and M. S. Brown, G. C., and Suter, D. (2012). In Defence of RANSAC for Outlier Rejection in Deformable Registration. *European Conference on Computer Vision*.

