3D surface Approximation of the Entire Bayeux Tapestry for Improved Pedagogical Access

Marjorie Redon Matthieu Pizenberg Yvain Quéau Abderrahim Elmoataz
Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France
{marjorie.redon,matthieu.pizenberg,yvain.queau,abderrahim.elmoataz-billah}@unicaen.fr

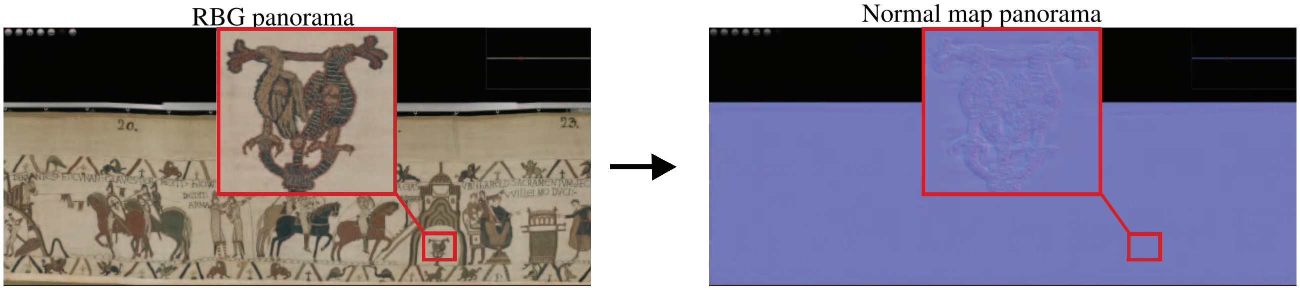


Figure 1: We present a pipeline for enabling the fine-scale 3D-reconstruction of a cultural heritage masterpiece - the Bayeux Tapestry. The proposed system transforms an existing RGB panorama (left) of this embroidery into a normal map panorama (right), which reveals the thin geometric details of the artwork.

Abstract

The Bayeux Tapestry is an exceptional cultural heritage masterpiece by its size and the finesse of its details. Digitizing it raises a challenge, knowing that it is extremely fragile and thus lasers or invasive techniques are out of scope. In this work, we address this 3D-reconstruction challenge by introducing a pipeline to generate a high-resolution panorama of the Tapestry's geometry. It is based on a deep learning architecture that converts the RGB images of a pre-existing 2D panorama into a 2.5D normal map panorama. With a view to facilitating the Tapestry inclusive accessibility, we further show that coupling our 3D-reconstruction pipeline with a segmentation method allows the affordable and rapid creation of 3D-printed bas-reliefs, which can be explored tactilely by visually impaired people.

1. Introduction

Art conservation has become of paramount importance since art itself exists [13], and technology has a crucial role to play in this domain. The emergence of cameras first enabled the creation of back-ups of pictured artworks [41]. Then, with the rising of computers and the internet, online museums such as the "Virtual Museum of Computing" started to develop [6]. The technologies' improvement then progressively allowed more complex digitization systems which go beyond 2D, e.g. laser scanners [30], photogrammetry [3], structured light [44] or time-of-flight [33]. Nowadays, 3D digitization is considered as a common practice in the cultural heritage domain [28], and high-quality results have been demonstrated [8]. Thus back-ups are essential to keep a trace of time deterioration and more important to the reconstruction of cultural heritage art or monuments after accidents such as fire destruction [32].

In the present work, we tackle the 3D digitization of a particular large-scale and fragile artwork: the Bayeux Tapestry. This masterpiece from the eleventh century tells the epic of William, Duke of Normandy, through a $70m$ long and $50cm$ high *embroidery*. Being made of wool strings on a linen canvas, it exhibits fine-scale surface details whose perception remains out of scope for the public - the Tapestry is protected by glass. The exceptional size of this masterpiece, along with the finesse of its details, makes its 3D digitization particularly challenging.

Contribution – The deep learning-based solution we introduce turns a pre-existing RGB panorama [1] into a 2.5D normal map panorama which reveals the thin geometric variations of the artwork (Fig. 1). This allows not only the virtual inspection of the embroidery's details by curators or the general public but also the automated creation of 3D-printed objects enabling the actual perception of the micro-relief. As an application, we show how coupling our 3D-reconstruction pipeline with a segmentation framework allows the creation of 3D objects which can be explored tactilely by Visually Impaired People (VIPs).

2. Related work

Digitization of real-world assets in three dimensions is an open problem where choosing a reconstruction approach requires balancing the advantages and drawbacks of each method. The main criteria are the precision of the reconstruction, the subject scale, and fragility, the setup cost and difficulty, as well as specific accessibility constraints. We are especially interested in digitizing the Bayeux Tapestry, which, as previously mentioned, is a thousand-year-old wool embroidery on a linen canvas. Preservation of this piece of European history is the main constraint, so coordinate measuring machines, with their probe-mounted moving arms, are obviously out of the equation.

Digitizing large-scale cultural artifacts very often involves lasers [15, 20, 30], however, this is also out of question for fragile artworks such as the Bayeux Tapestry. Alternative non-destructive methods based on multi-image 3D-reconstruction have been proposed [29]. Yet, this method has been thought of for building reconstruction, not for fine details of structures such as embroidery. One could imagine resorting to structured light sensing as in [44] However, the presence of the protective glass prevents such a technique from working correctly. A viable solution could be photogrammetric techniques [3]. Those kinds of techniques allow representing the shape of large-scale artworks, e.g. the "Meissen Fountain Table" presented at the Victoria and Albert Museum in London [36]. Still, the 3D-reconstruction of high-frequency geometric variations remains limited.

Since we are interested in the representation of such fine-scale embroidery variations, photometric techniques seem preferable over photogrammetry. Shape-from-shading [22] has for instance been considered in [19, 21]. Yet, these works are interested in giving a 3D interpretation of paintings, which are subject to the "trompe l'oeil" ambiguity, while we are interested in an object presenting real micro-variations on its surface. The photometric stereo technique (PS) [42] provides a solution to this high-frequency 3D-reconstruction problem. Furthermore, it has successfully been applied in various cultural heritage applications [17, 31]. Therein, a per-pixel surface normal map is estimated from a series of photos taken under the same viewing angle, while varying the lighting directions to create shading effects in the images. The precision of the 3D scan is thus mostly dependent on the sensor resolution. With the emergence of new technologies, the resolution of recent cameras tends to dramatically increase: we passed from a resolution of 4 megapixels at the beginning of the century to a resolution of 45 megapixels in 2023. Hence, one may have good hopes to recover structures as thin as the wool strings of the Tapestry. Besides, the setup for photometric stereo is quite straightforward and relatively cheap. It only requires a remotely triggered camera, a tripod, and spherical objects used for calibrating the light directions.

However, applying this solution to the entire Tapestry would be extremely time-consuming, as this would require successively setting a PS setup for each scene of the $70m$ artwork. On the other hand, we already have at our disposal an RGB panorama of the artifact [1]. Therefore, we propose to use PS on a small subset of the Tapestry to construct a ground truth database of (image, geometry) pairs. Then, from this ground truth database, a neural network will be trained to turn a single RGB image into a normal map. Consequently, the entire RGB panorama can be converted into a normal map panorama, yielding the fine-scale 3D-reconstruction of the large-scale artwork. This is detailed in the next section.

3. 3D-reconstruction of the Tapestry surface

Our aim in this section is to turn each RGB image of the panorama [1] into a representation of the surface geometry. Obviously, this is an ill-posed inverse problem since it is impossible to separate color from shape without a priori information [2]. Yet, insofar as all considered images represent a unique embroidery (which is a quasi-planar surface with thin surface variations), we hope that a trained neural network will manage to generalize correctly i.e., find a shape that is qualitatively adequate (our work has a pedagogical, rather than a metrological, aim). To encode the thin geometric variations, we represent the 3D shape by a normal map. The 3D-reconstruction problem thus becomes an image translation one, where the input image must be converted into a normal map. We propose to solve this problem using a Generative Adversarial Network (GAN) inspired by Pix2Pix [24]. As illustrated in Fig. 2, such a network consists of a generator that turns the RGB image into a normal map and a discriminator that compares the generated map with the ground truth one.

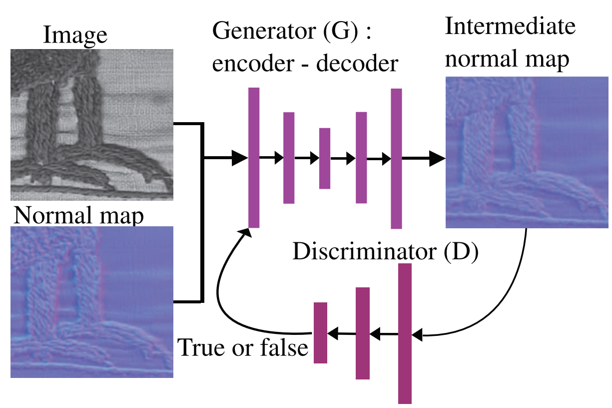


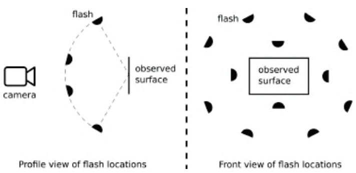
Figure 2: GAN architecture: pairs (image, normal map) are sent to an encoder-decoder generator (G). It creates a new normal map which is compared with the real one by the discriminator (D) to minimize the discrepancy between them.

3.1. Training set creation using photometric stereo

The proposed network must be trained on a dataset of ground truth (image, geometry) pairs. In order to create this training set, we achieved the 3D-reconstruction of a subset of the Tapestry through photometric stereo, as illustrated in Fig. 3.

(a) Digitization of the scene 55. The camera is fixed on a tripod, and the flash is held on the side, directed toward the captured scene.

(b) Calibration sphere stick on the glass panel.



(c) Distribution of flash locations on a hemisphere around the subject.

(d) 3D-scan equipment.

Figure 3: Equipment and protocol for the 3D-digitization of the Bayeux Tapestry by photometric stereo.

Acquisition protocol The PS technique infers geometry from the brightness variations arising in the images as the light direction changes. Getting sharp images without noise is key to the reconstruction quality. We thus decided to use a DSLR camera with high light sensibility, which can be remotely triggered and connected to a flashlight (cf Fig. 3d). An important point to mention is that a pre-study of the digitization protocol was performed and submitted to the "Direction régionale des affaires culturelles" (DRAC Normandie) to ensure there was no deterioration risk. Therein, we established that the energy brought by our flashlight amounts to 6 seconds of the Bayeux Tapestry's normal lighting conditions in its exhibition room.

PS assumes that all images are captured from the exact same point of view. However, at the scale of a pixel, any sensor vibration due to a manual trigger or floor vibrations due to walking around can create misalignments in the sequence of images. A strong and stable tripod, as well as remote triggers, are thus highly recommended. In order to calibrate the location of the flashlight for each image of a sequence, we placed small reflective spheres at the images corners. Since the artifact was kept behind its protective glass, we designed cheap, hand-made calibration spheres, by joining plastic door knobs and bathroom wall suction cups (Fig. 3b), enabling easy displacements of the spheres.

Once the equipment is gathered, the digitization steps are the following: 1) place the camera on its tripod and adjust the frame; 2) place the reflective spheres on the image corners; 3) focus on the subject; 4) remotely launch a sequence of shots while changing the flash location between each shot. A minimum of 3 shots with non-co-planar light sources is required for photometric stereo, but we went with 12 shots to enable robustness to outliers such as cast shadows and caustics induced by the spheres and glass panel. As illustrated in Fig. 3c, eight of the sources were uniformly placed at an elevation angle of roughly $50°$ (the optimal value for the digitization of quasi-planar artefacts [7]), while four additional shots were taken with a smaller elevation angle to increase the number of outlier-free images.

Pre-processing Before proceeding to the 3D-reconstruction, it is necessary to calibrate the incident lighting using the reflective spheres. It is straightforward to infer the illumination direction from the location of the saturated pixels on the sphere. However, the flash source being non-directional, each sphere provides a different estimate: we simply average them to have a reliable estimate at the center of the scene.

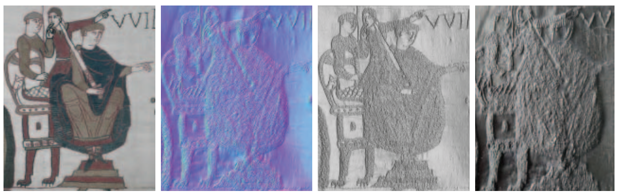
Despite relying on a stable tripod, there might still be slight sensor displacements between the shots. To cope with this issue, we automatically register the image sequence using a low-rank approximation procedure [34]. The Tapestry has a diffuse reflectance, hence the light-geometry interaction is fully governed by Lambert's law. Undesirable caustics and shadows cast by the calibration spheres, visible in Fig. 4a, can thus be considered as outliers to the linear Lambertian model. In addition to compensating for misalignments, the low-rank approximation procedure allows us to automatically remove such artifacts from the input images.

3D-reconstruction algorithm Given the pre-processed images and calibrated illumination, we iteratively estimate the reflectance, normals, and source intensities by semi-calibrated PS [12]. This yields a normal field which is integrated into a depth map using discrete cosine transform [35]. Since we wrongly assumed directional lighting for simplicity, the depth map exhibits a well-known low-frequency bias ("potato chip" effect [23]). We remove this bias by fitting and subtracting a low-dimension polynomial from the depth map, which is simpler to implement than, e.g., resorting to photogrammetry [26]. Next, we use finite difference to obtain the normal map which will finally serve as the ground truth in our deep learning-based algorithm. The quality of this normal map can be qualitatively verified by integrating it again into a depth map, as in Figs. 4b and 5c. Besides, such depth maps can be meshed and turned into a printable 3D volume, as illustrated in Fig. 5d.

(a) Two of the input images. Note the shadows and caustics induced by the spheres and the glass.

(b) Highly detailed 3D-reconstruction, overlaid on one pre-processed input image.

Figure 4: Photometric stereo-based generation of a ground truth (image, geometry) pair, illustrated on the death of Harold sequence (scene 57). The normal map corresponding to the PS reconstruction, coupled with any of the pre-processed images, will serve as ground truth in our training set.



(a) Image (b) Normals (c) Depth (d) 3D-print

Figure 5: Photometric stereo-based generation of a ground truth (image, geometry) pair, illustrated on the Duke William sequence (scene 27). Besides constituting our training set, the estimated geometries can be 3D-printed in view of tactile experiments (see Sect. 4).

Complete database The previous process resulted in the high-resolution 3D-reconstruction of 16 scenes among the 58 scenes of the Bayeux Tapestry. Therein, each normal map is of size 5568×3712 px, and is pixel-accurate registered with 12 gray-level images. On those images shadows or caustics are not present thanks to the pre-processing procedure. To constitute our ground truth database, we empirically picked one gray-level-converted pre-processed image. Remark that since we have the estimated reflectance, we could have rendered new synthetic images under novel illumination. We left such an approach as perspective. Then, we split these 16 (image, normal map) pairs into 13 for training and 3 for testing. Lastly, each pair was randomly cropped into 500 thumbnails of size 256×256 px, providing a total of 6500 data for training and 1500 for testing.

3.2. GAN-based normal estimation framework

Base architecture Given the previously described ground truth database, we first trained a standard GAN architecture (Fig. 2) to minimize the 1-norm difference between real and simulated normal maps, using the Adam algorithm, which we let iterate during 50 epochs.

Concave-convex ambiguity This basic GAN architecture provided somehow reasonable results, with an average angular deviation of $13.30 \pm 1.07°$ on the test set. However, by integrating the resulting normal maps, we noticed that some structures were incorrectly reconstructed. Indeed, as illustrated in Fig. 6, several yarns appear concave, although by nature the embroidery exhibits only convex structures. This is after all not that surprising: inferring geometry from a single image comes down to solving the shape-from-shading problem, which is fundamentally prone to the concave-convex ambiguity [18].

Here, we can interpret the concave aspect of the generated yarn as an over-interpretation of the high frequencies by the GAN. In fact, frequency artifacts are well known in deep fake detection [25]. They appear during both the generation and discrimination steps of the learning process. During the image generation step, a U-Net architecture is used [24, 39]. This neural network architecture is based on several mathematical operations, including convolutions and max pooling. During the contraction process, those increase the pixel classification and decrease its localization. In other words, we know if the pixels are concave or convex thanks to their closest neighbors, so the fine structures of the embroidery are well generated but not the general structures like yarns form.

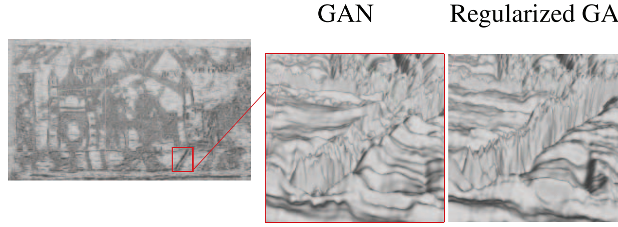


Figure 6: GAN-based 3D-reconstruction of the King Edward sequence (scene 1). Without regularization, some of the embroidered yarns are wrongly inferred as concave.

Low-frequency regularization During the discrimination step, the generated normal is compared to the input one by combining two losses: L_1 and L_{GAN}. Here, L_{GAN} is a Markovian discriminator (PatchGAN) which captures the high-frequency structures, while the L_1 term is expected to enforce low-frequency correctness. However, the previously observed concave-convex ambiguities show that the low frequencies are not sufficiently well generated in our particular context. This drove us to add a third term to further encourage low-frequency correctness. To this end, we added an L_1 regularization on the low-pass filtered normals:

$$L_{\text{reg}}(N, \tilde{N}) = \|F(N) - F(\tilde{N})\|_1 \quad (1)$$

where F is a low-pass filter:

$$F(N) = \mathcal{F}^{-1}(\mathcal{F}(N) \cdot H) \quad (2)$$

$$H(u,v) = \begin{cases} 1 & \text{if} \quad D(u,v) \leq D_0 \\ 0 & \text{if} \quad D(u,v) > D_0 \end{cases} \quad (3)$$

$$D(u,v) = \sqrt{\left(u - \frac{M}{2}\right)^2 + \left(v - \frac{N}{2}\right)^2} \quad (4)$$

with $\mathcal{F}$ the Fourier transform, (M, N) the image size and D_0 a parameter empirically set to 50. Fig. 7 summarizes the architecture of the proposed regularized GAN.

Evaluation The regularized GAN was trained in the same conditions as above, on the same dataset. This required about 12 computation hours on an Intel Xeon E5-26-40 v4 processor at 2.4 GHz equipped with an MSI Ge Force GTX 1080 Ti 3584 Cores GPU with 11GB of RAM. Quantitatively, we obtain an average angular deviation of $13.26 \pm 1.11°$ on the test set, which is slightly better than the unregularized version. As expected, the differences between the unregularized and the regularized are sparse, and located on the few yarns which were incorrectly inferred as convex. A concavity correction example is provided in Fig. 6.

In Fig. 8, we provide a few qualitative normal reconstruction results. The results on the test set are very satisfactory. This is not very surprising since the illumination resembles that in the training set. Still, it is worth noticing some imperfections: in the first reconstruction, a spot is wrongly interpreted in terms of shape rather than color variations. Indeed, no such spot is present in the training set, which should be enlarged if we want to improve robustness to such effects. Also, the second reconstruction is slightly blurred: this is likely due to the encoder-decoder architecture, which builds upon convolution layers. Interestingly, the proposed approach seems robust to illumination variations: the third image was directly taken from the online panorama of the Tapestry [1]. In this case, the illumination is uncontrolled and we have no ground truth, yet the results are qualitatively satisfactory.

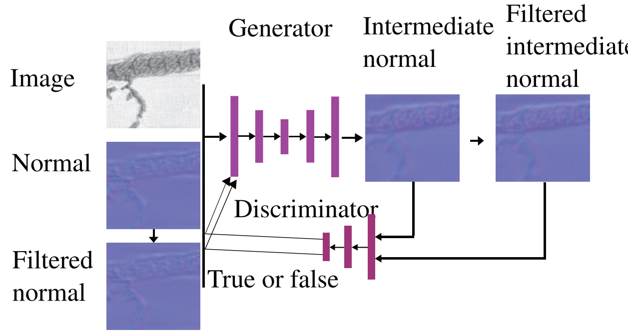


Figure 7: Regularized GAN architecture. In comparison with the standard GAN architecture (Fig. 2), a low-pass filter is applied to both the ground truth and the generated normals, and the loss in the discriminator combines distance between both the normals and the filtered normals, to further encourage low-frequency correctness.

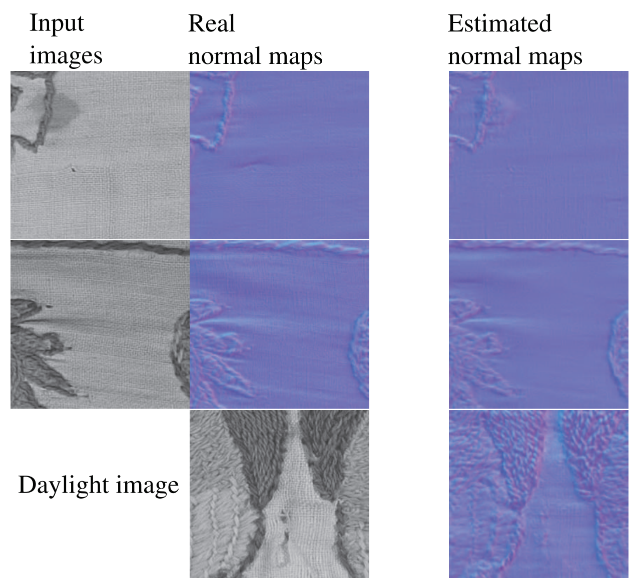


Figure 8: Normal map reconstruction by the regularized GAN architecture on images from the test database (top rows) and on a photograph acquired under uncontrolled lighting (bottom row).

3.3. Normal map panorama

An image digitization campaign of the Tapestry was carried out by “La Fabrique de Patrimoines en Normandie” in January 2017. This digitization had several goals: serve as a reference for studying the future evolution of the artifact; serve as the basis for different research programs; feed a geo-referenced image database; and provide high-resolution images of the artifact for research, cultural, education, and communication purposes. The campaign resulted in a series of 86 high-resolution images (roughly $8000 \times 5500\ px$).

Figure 9: A view of the web interface which allows the exploration of the geometric panorama of the Tapestry.

Those daylight images enabled the constitution of a high-definition digital panorama of the Tapestry, of size 680.000×5500 pixels. This RGB panorama is readily available online, allowing one to remotely explore the artwork [1]. Using our regularized GAN architecture, we converted each of the input RGB images into a normal map and re-assembled all the obtained normals into a new panorama. This new geometric panorama of the Tapestry can be explored in a web interface (Fig. 9) which will be made publicly available, allowing anyone to explore the thinnest geometric details of this cultural heritage masterpiece. Besides online exploration, reconstructed geometry can also serve as a basis for creating 3D-printed objects in view of tactile experiments, in particular for visually-impaired people. Such an application is explored in the next section.

4. Application: 3D-prints for blind people

In 2020, M. J. Burton et al. [9] estimated that 1.1 billion people have a visual disability. For them, the visual artwork's perception remains difficult. In order to be more inclusive, some museums make visual artworks accessible through another sense – hearing or touch. For example, the Petite Galerie du Louvre presents artworks with audio descriptions, while the exhibition "Touching the Prado" combines paintings with Braille descriptions and three-dimensional plates. The 3D-reconstruction of the Tapestry which was described previously allows us to follow a similar track, by enabling tactile exploration of the artwork through 3D printing. To this end, we embed the previous 3D-reconstruction framework into a larger pipeline aiming at the automated generation of 3D objects from a single image of the Tapestry, as illustrated in Fig. 11. This pipeline combines the previous 3D-reconstruction module with a segmentation tool so that the semantically important structures can be emphasized in the 3D-printed bas-relief.

4.1. Detection and segmentation

We developed a module for detecting, segmenting, and classifying the elements present in the input image. We know from previous studies of the Tapestry [5] that there is a total of 1515 elements, which can be grouped into 10 different categories. We decided to focus on four: letters, animals, people, and boats. The module aims to locate, precisely segment, and classify each of these elements. For this, we use the Mask-RCNN [37] algorithm implementation provided in Detectron2 [43], cf. Fig. 10.

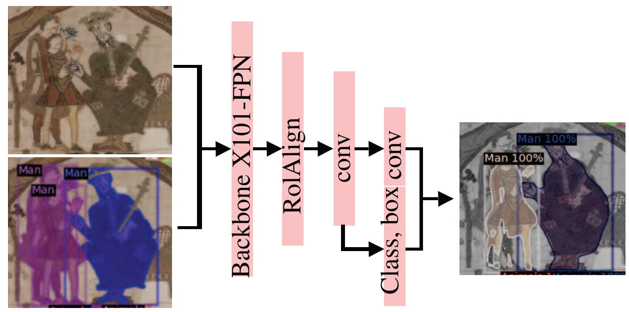


Figure 10: Mask-R-CNN neural network architecture. Pairs (images, annotations) are fed to the neural network, which is trained to locate, segment, and classify the elements present in the image. Here, segmenting the three characters is difficult, as one is partially occluded by another.

To constitute the database, we chose 12 representative images of size 8100×5600 px. Using the CVAT [14] tool, we annotated the four class elements. The database is augmented by simulating camera translations through random crops of size 5500×5500 px. We then rescale them to 1000×1000 px by linear interpolation and apply random modifications of brightness, saturation, luminosity, and contrast. This yields a total of 56000 pairs (image, annotations) for training, and 50 for testing. Out of the 8 models provided in Detectron2 [43], we chose the X101-FPN model, whose weights pre-trained on COCO were refined using our learning database. For the optimization, we used an inertial stochastic gradient descent algorithm, which is stopped after 50000 epochs. The batch size is 5 images per epoch. Refining the weights requires about 30 computation hours on the same processor as in Sect. 3.

The results, illustrated in Fig. 12, are in line with what is expected on simple and non-overlapping elements. Yet, as soon as the elements are superimposed, the approach limits become visible. Under-segmentation (right character head in the second image), misclassification (left horse classified as a man in the last image) and over-segmentation (second horse detected in two parts in the bottom-left image) indeed appear. A possible workaround would be to refine the automatic segmentation by a semi-supervised technique such as GrabCut [40] or SAM [27].

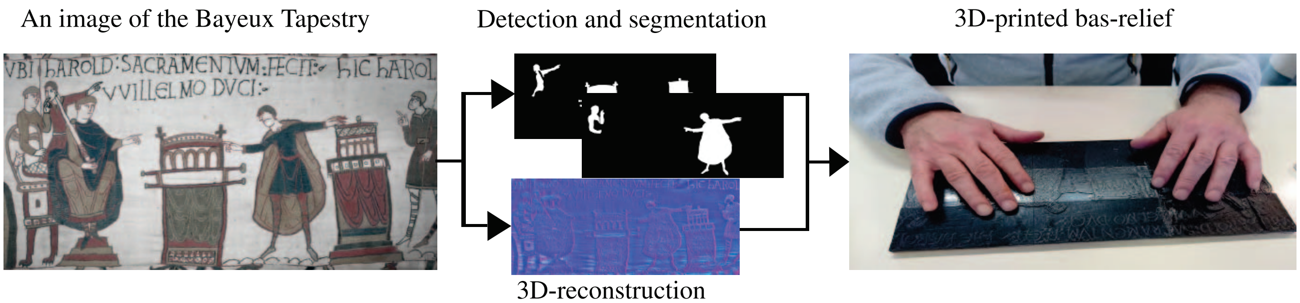


Figure 11: Complete pipeline for enabling the tactile exploration of the Bayeux Tapestry. The proposed system transforms an image (left) of this embroidery into a 3D-printed bas-relief that can be explored tactilely (right). A detection and segmentation module isolates each element of the scene, while the 3D-reconstruction module transforms the image into a normal map.

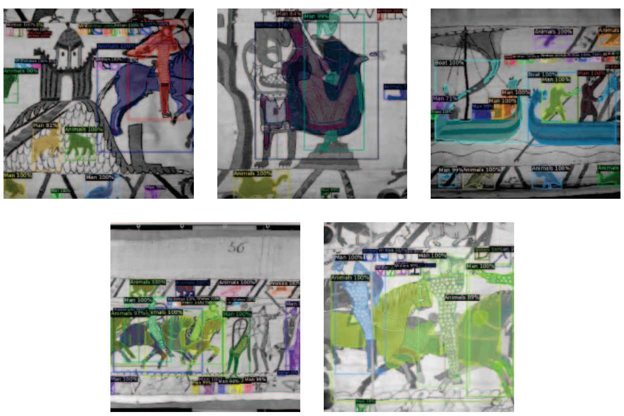


Figure 12: Detection and segmentation of the four chosen categories. The results are globally satisfactory, except on overlapping elements.

Figure 13: Setting each element's average height, from the detected and segmented masks. Here the background is at level zero, the letters at 2 mm, and the comet at 3 mm.

4.2. Creation of the shapes

Using the masks found during the first stage, the average height of each element is fixed manually (Fig. 13), so that the different semantic structures can easily be identified in tactile experiments. Then, the surface thin variations are transferred into these elements by integrating the normals obtained from the previously described 3D-reconstruction module. The result is a 2.5D representation where each element can be identified by its height, and the inside of each element is "textured" according to the actual geometry of the embroidery in order to give a feeling of the thinness of the artwork. Fig. 15 shows some examples of 3D models that have been created in this way.

4.3. Experience feedback

To empirically validate the interest of our pipeline, we printed a series of bas-reliefs and, as shown in Fig. 11, we provided them to volunteers during a tactile experience. This experiment was carried out with the support of one "sighted" person, three partially blind and three blind ones, to receive feedback from people with different perceptions.

General impressions We have received very positive initial feedback. In fact, these supports have been perceived as very complementary to the tactile plates handmade by Rémy Closset (Fig. 14). Whereas those are sculpture-like smooth, ours transcribe the artwork's thin geometry. However, for both of them, the addition of an audio description is essential to accompany the touched discovery and to explain its semantic content, as is also analyzed in [38]. In our future prints, we will therefore amplify it before the integration. A major difficulty remains in the perception of partially occluded objects. In this case, we can either respect the original work or transform it, for example by taking the liberty of not representing the occluding element. The latter solution eases the elements' understanding and was more appreciated.

Figure 14: Bas-relief handmade by M. Closset [16], representing Harold's oath sequence.

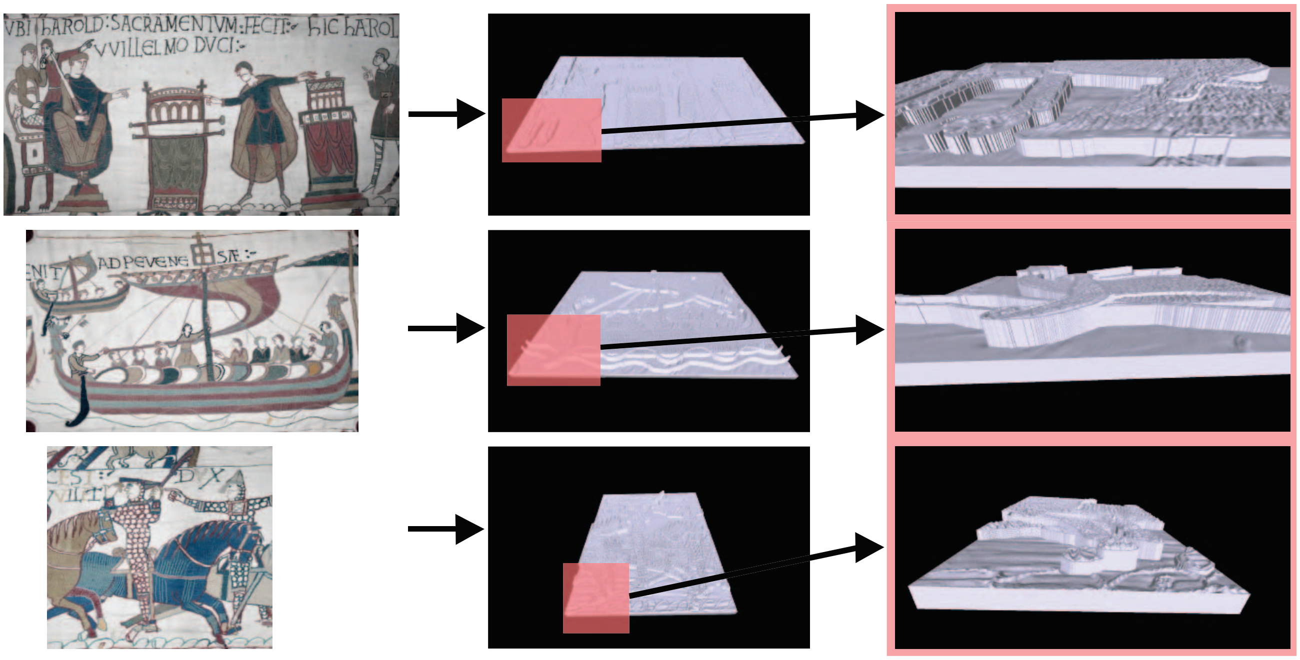


Figure 15: Three scenes from the Bayeux Tapestry (left), and their 3D representation obtained following the proposed approach. Attributing a different height to the segmented elements allows highlighting them while transferring the embroidery geometry enabling the perception of the artwork micro-relief.

Possible extensions of the device It seems very important to pay attention to the represented elements' spacing. So far, to not betray the original artwork, we kept its layout. This may not be relevant when the spacing is insufficient to feel the separation between two elements. In future printings, we will make sure that all elements are separated well. Another possibility is to allow elements to be detached from their support, like a puzzle. This idea would ease the elements outline's discernment, and make at the same time the experience more interactive.

We also proposed a color addition to the prints. Although using colors for VIPs may not seem relevant, this is not true for partially blind people. High-contrast colors would help them to distinguish elements. Furthermore, during the Bayeux Tapestry creation, only 10 natural dyes have been used, which makes the print's colorization doable. Work in progress [11] is aimed at identifying the wool composition used during the Bayeux Tapestry creation, and so its original colors. It would be interesting to add them to our prints, in order to show what the work might have looked like at its creation end.

A last proposal is to enhance tactile exploration with haptic feedback and sound. For instance, an audio description could be launched when an object is touched twice, in a similar way hand gestures are recognized in deaf-blind communication systems [4].

5. Conclusion and perspectives

We have proposed a framework for the fine-scale 3D approximation of the entire Bayeux Tapestry's surface, with a particular view at improving its pedagogical access. A regularized GAN architecture was proposed, which infers the geometry of the embroidery's surface from a single image. This allowed us to transform an existing RGB panorama into a normal map panorama which can be used for remote inspection of the embroidery's geometric details. As an application, we also showed how such a 3D-reconstruction pipeline can be coupled with automatic segmentation tools for generating 3D-printed objects which enable inclusive tactile explorations of the artwork. In addition to the avenues of improvement mentioned in the previous section, we plan to improve the device and integrate it into a multimodal inclusive system, in the manner of Q. Cavazos et al. [10].

Acknowledgment

This work was supported by the ANR grant "Guide Muséal Inclusif" (ANR-20-CE38-0007). The authors thank C. Berthelot of the Musée de la Tapisserie de Bayeux for supervising the image acquisition campaign, and the Valentin Haüy Association for experiments with VIPs.

References

[1] Official digital representation of the Bayeux Tapestry - XIth century. Credits: City of Bayeux, DRAC Normandie, University of Caen Normandie, CNRS, ENSICAEN. Photos: 2017 – La Fabrique de patrimoines en Normandie. . https://www.bayeuxmuseum.com/en/the-bayeux-tapestry/discover-the-bayeux-tapestry/explore-online/, 2021. [Online; accessed July 18th, 2023].

[2] E. H. Adelson and A. P. Pentland. The perception of shading and reflectance. *Perception as Bayesian inference*, 409:423, 1996.

[3] I. Aicardi, F. Chiabrando, A. M. Lingua, and F. Noardo. Recent trends in cultural heritage 3d survey: The photogrammetric computer vision approach. *Journal of Cultural Heritage*, 32:257–266, 2018.

[4] G. Airò Farulla, L. Russo, C. Pintor, D. Pianu, G. Micotti, A. Salgarella, D. Camboni, M. Controzzi, Ch. Cipriani, C. Oddo, S. Rosa, and M. Indaco. Real-time single camera hand gesture recognition system for remote deaf-blind communication. In *Augmented and Virtual Reality*, pages 35–52, 2014.

[5] S. Bertrand. *La Tapisserie de Bayeux et la manière de vivre au onzième siècle*. Zodiaque, 1966.

[6] J. P. Bowen. A brief history of early museums online. *The Rutherford Journal*, 3, 2010.

[7] S. Brenner, S. Zambanini, and R. Sablatnig. An investigation of optimal light source setups for photometric stereo reconstruction of historical coins. In *Eurographics Workshop on Graphics and Cultural Heritage*, pages 203–206, 2018.

[8] D. L. Buglio and L. De Luca. Critical Review of 3d Digitization Methods and Techniques Applied to the Field of Architectural Heritage: Methodological and Cognitive Issues. In *VAST: International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage - Short and Project Papers*, 2011.

[9] M. J. et al. Burton. The lancet global health commission on global eye health: vision beyond 2020. *The Lancet Global Health*, 9(4):489–551, 2021.

[10] L. Cavazos Quero, J. Iranzo Bartolom'/e, and J. Cho. Accessible visual artworks for blind and visually impaired people: comparing a multimodal approach with tactile graphics. *Electronics*, 10(3):297, 2021.

[11] C. Chavanne, A. Verney, C. Paquier-Berthelot, M. Bostal, P. Buléon, and P. Walter. Bayeux tapestry: First use of early synthetic dyes for the restoration of a masterpiece. *Dyes and Pigments*, 208:110798, 2022.

[12] D. Cho, Y. Matsushita, Y. Tai, and I. Kweon. Semi-calibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1):232–245, 2018.

[13] A. Conti and H. Glanville. *History of the Restoration and Conservation of Works of Art*. Routledge, 2007.

[14] CVAT.ai Corporation. Computer Vision Annotation Tool (CVAT), 2022.

[15] N. Crombez, G. Caron, and E. Mouaddib. Using dense point clouds as environment model for visual localization of mobile robot. In *2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pages 40–45, 2015.

[16] Association Valentin Haüy / Musée de Bayeux. Bas-relief by M.Closset: Harold swears on the relics, 2019.

[17] T. G. Dulecha, R. Pintus, E. Gobbetti, and A. Giachetti. SynthPS: a Benchmark for Evaluation of Photometric Stereo Algorithms for Cultural Heritage Application. In *Eurographics Workshop on Graphics and Cultural Heritage*, pages 13–22, 2020.

[18] J.-D. Durou, M. Falcone, Y. Quéau, and S. Tozza. A comprehensive introduction to photometric 3d-reconstruction. *Advances in Photometric 3D-Reconstruction*, pages 1–29, 2020.

[19] R. Furferi, L. Governi, Y. Volpe, L. Puggelli, N. Vanni, and M. Carfagni. From 2D to 2.5D i.e. from painting to tactile model. *Graphical Models*, 76(6):706–723, 2014.

[20] L. Gomes, O. Regina Pereira Bellon, and L. Silva. 3d reconstruction methods for digital preservation of cultural heritage: A survey. *Pattern Recognition Letters*, 50:3–14, 2014.

[21] L. Governi, R. Furferi, Y. Volpe, L. Puggelli, and N. Vanni. Tactile exploration of paintings: an interactive procedure for the reconstruction of 2.5 d models. In *Mediterranean Conference on Control and Automation*, pages 14–19, 2014.

[22] B. K. P. Horn. *Shape from Shading: a Method for Obtaining the Shape of a Smooth Opaque Object from One View*. PhD Thesis, MIT, 1970.

[23] X. Huang, M. Walton, G. Bearman, and O. Cossairt. Near light correction for image relighting and 3d shape recovery. In *Digital Heritage*, volume 1, pages 215–222, 2015.

[24] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial nets. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5967–5976, 2017.

[25] Y. Jeong, D. Kim, Y. Ro, and J. Choi. Frepgan: Robust deepfake detection using frequency-level perturbations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(1):1060–1068, 2022.

[26] A. Karami, F. Menna, and F. Remondino. Investigating 3d Reconstruction Of Non-Collaborative Surfaces Through Photogrammetry And Photometric Stereo. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:519–526, 2021.

[27] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023.

[28] A. Koutsoudis, K. Stavroglou, G. Pavlidis, and C. Chamzas. 3dsse – a 3d scene search engine: Exploring 3d scenes using keywords. *Journal of Cultural Heritage*, 13:187–194, 2012.

[29] A. Koutsoudis, B Vidmar, G. Ioannakis, Arnaoutoglou F., G. Pavlidis, and C. Chamzas. Multi-image 3d reconstruction data evaluation. *Journal of Cultural Heritage*, 15(1):73–79, 2014.

[30] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project:

3d scanning of large statues. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, page 131–144, 2000.

[31] J. Mélou, A. Laurent, C. Fritz, and J.-D. Durou. 3D Digitization of Heritage: Photometric Stereo can Help. In *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 145–152, 2022.

[32] E. Mouaddib, A. Pamart, M. Pierrot-Deseilligny, and D. Girardeau-Montaut. 2D/3D data fusion for the comparative analysis of the vaults of Notre-Dame de Paris before and after the fire. *Journal of Cultural Heritage*, 2023.

[33] M. Pieraccini, G. Guidi, and C. Atzeni. 3d digitizing of cultural heritage. *Journal of Cultural Heritage*, 2:63–70, 2001.

[34] M. Pizenberg, Y. Quéau, and A. Elmoataz. Low-rank registration of images captured under unknown, varying lighting. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 153–164, 2021.

[35] Y. Quéau, J.-D. Durou, and J.-F. Aujol. Normal integration: a survey. *Journal of Mathematical Imaging and Vision*, 60:576–593, 2018.

[36] A. Reichinger, H. G. Carrizosa, and C. Travnicek. Designing an interactive tactile relief of the meissen table fountain. In *International Conference on Computers Helping People with Special Needs*, pages 209–216, 2018.

[37] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2015.

[38] K. Romeo, H. Thompson, and M. Chottin. Inclusive multimodal discovery of cultural heritage: Listen and touch. In *Computers Helping People with Special Needs*, pages 278–285, 2022.

[39] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. Wells, and A. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, pages 234–241, 2015.

[40] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23:309—314, 2004.

[41] F. Tietjen. Photography in art conservation. pages 1102–1104. Routledge, 2008.

[42] R. J. Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Image understanding systems and industrial applications I*, volume 155, pages 136–143, 1979.

[43] Y. Wu, A. Kirillov, F. Massa, W. Lo, and R. Girshick. Detectron2. https://github.com/facebookresearch/detectron2.

[44] T. Zaman, P. Jonker, B. Lenseigne, and J. Dik. Simultaneous capture of the color and topography of paintings using fringe encoded stereo vision. *Heritage Science*, 2(1):1–10, 2014.