







Towards the Tactile Discovery of Cultural Heritage with Multi-approach Segmentation

Ali Souradi¹, Christele Lecomte¹, Katerine Romeo¹ ,
Simon Gay¹ , Marc-Aurele Riviere¹ , Abderrahim El Moataz²,
and Edwige Pissaloux¹ 

¹ Rouen Normandy University, Saint Etienne du Rouvray, France
katerine.romeo@univ-rouen.fr

² Caen Normandy University, Caen, France

Abstract. This paper presents a new way to access visual information in museums through tactile exploration, and related techniques to efficiently transform visual data into tactile objects. Accessibility to cultural heritage and artworks for people with visual impairments requires the segmentation of images and paintings to extract and classify their contents into meaningful elements which can then be presented through a tactile medium. In this paper, we investigate the feasibility and how to optimize the tactile discovery of an image. First, we study the emergence of image comprehension through tactile discovery, using 3D-printed objects extracted from paintings. Later, we present a dynamic Force Feedback Tablet (F2T) used to convey the 2D shape and texture information of objects through haptic feedback. We then explore several image segmentation methods to automate the extraction of meaningful objects from selected artworks, to be presented to visually impaired people through the F2T. Finally, we evaluate how to best combine the F2T's haptic effects in order to convey the extracted objects and features to the users, with the aim of facilitating the comprehension of the represented objects and their affordances.

Keywords: Accessibility to artworks · Cultural heritage · Visual impairment · Image segmentation · Haptic interface

1 Introduction

Access to art and culture for visually impaired people (VIP) is often complex, with most of the artworks exhibited in museums relying on visual (2D) content. Currently, the most common solutions to overcome this problem are audio descriptions and 3D models for tactile exploration. However, these solutions convey limited information and have several drawbacks: audio descriptions are sequential, passive, and they monopolize the attention and the listening of the user. An active exploration (gaze or finger-guided) is paramount to form a holistic and coherent mental picture of the explored content (and thus appreciate its beauty), which is often not compatible with the linearity and passivity of audio descriptions. On the other hand, 3D-printed or thermoformed objects are usually expensive to manufacture (due to the transposition of

the artwork to its tactile representation often done manually by artists), and usually provides too much detail for an efficient tactile exploration.

Allowing VIP to explore tactile representations of artworks autonomously is a challenge that requires both the automated extraction of the meaningful content from an image and its adaptation to the specificities of haptic exploration. Each artwork being specific, classical automatic methods fail to provide a universal solution.

In order to improve accessibility to art and culture for VIP, we want to design and prototype a solution allowing to display transposed artworks, combining tactile and kinesthetic perceptions, which can be actively explored by the user. This solution will cover the two main problematics linked with artwork accessibility: (1) the reliable automatic extraction of meaningful objects from artworks and (2) their efficient transposition into haptic representations allowing VIP to understand their meaning and appreciate their beauty.

This paper first presents how a tactile representation of artworks is perceived on a printed object and the difficulties of understanding the shapes (Sect. 2). In Sect. 3, we introduce the semantic segmentation methods we explored in order to extract meaningful content from an image, and their results on the Bayeux Tapestry. The 4th section presents the F2T, a multimodal (audio-tactile) interface for the active exploration of graphical content, while the 5th section explores the possible applications of our system to cultural heritage of Bayeux Tapestry. Finally, we discuss the results and suggest possible future improvements.

2 Tactile Representation of Artwork

Tactile representations of artworks must be tailored to the specificities of haptic perception in order to be easily accessible to VIP [1, 2]. It must have, above all, outlines highlighted to detect the represented objects, and some indication should be given on their number and their location relative to the neighboring objects [3, 4]. This will allow users to better understand the explored object and its meaning in relation to the rest of the scene [5]. Furthermore, transposing information from one modality to another (i.e. visual to tactile) is an extremely complex process [6], comprised of two main problematics: selecting the most useful characteristics to convey (here, the spatial and visual features of an object), and finding the optimal way to encode those features in the space of the output modality (i.e. the tactile stimulation parameters of the haptic device used as an interface). Although it is an inclusive approach it involves a creation process towards tactile representation in order to communicate messages to users.

To experience how variations of the chosen tactile parameters might influence the recognition of the selected scene elements, we segmented (using inkscape) salient objects from a painting (Fig. 1a), such as a spinning wheel (Fig. 1b), and then 3D-printed it as a slightly elevated relief on a 2D plate, as shown in Fig. 1c. Visually, it is easily recognizable, but blindfolded participants were not able to locate or identify any part of the object without external help. We then printed a 3D model of this object (as shown in Fig. 1d) and presented both prints (2.5D and 3D) to blindfolded participants, explaining they were two representations of the same object. Again, the correspondence between the two models was not perceived through touch alone. As mentioned

by Hatwell in [3], tactile recognition can be acquired with training, but is usually an arduous process due to the interference of vision-specific elements, such as perspective cues, which will hinder recognition [7]. A drawn object with added relief information (2.5D, see Fig. 1c) has very little actual correspondences with the tactile experience of the actual object (in 3D, see Fig. 1d). Indeed, the projection of an object on a surface relies on perspective cues which are specific to vision. VIP thus have to learn to interpret those perspective cues in order to make sense of this type of images. A tactile representation must be simplified compared to the visual object it is representing, keeping only the essential information to allow its recognition. It must “preserve the overall meaning” (or gist) of the represented object [8, 9].

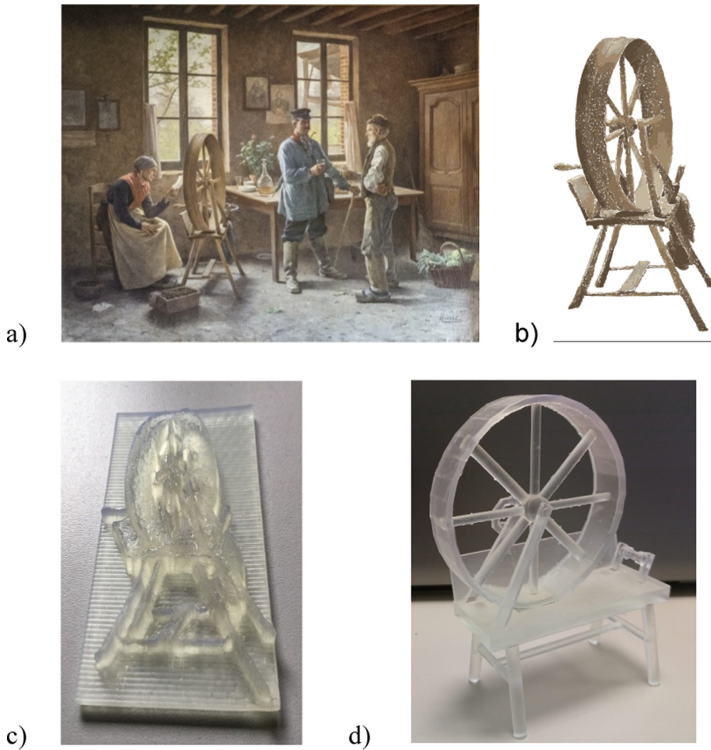


Fig. 1. a) Painting of L. Minet, Chateau de Martainville; b) The spinning wheel in the painting, c) printed with a small relief (2.5D); d) printed as a 3D model.

3 Semantic Segmentation of an Artwork

In this section, we present several segmentation methods we selected in order to extract meaningful elements from the cultural heritage of Bayeux Tapestry. The particularity of this piece of art is its lack of perspective or shadows. It is comprised of a multitude of

characters with their accessories, but also animals and some simplified buildings, depicting important scenes of the joint history of England and Normandy.

Semantic segmentation allows extracting and grouping elements of an image into meaningful categories. Several approaches exist, which could be categorized as: (1) segmentation based on edge detection; (2) segmentation based on the perceived similarities of spatially closed pixels (i.e. clustering), and (3) the cooperative segmentation of contours and regions.

These classes of methods all aim to segment the image using low-level information such as pixel color and its gradients along the image's dimensions. Selecting the appropriate methods will depend on the target images characteristics and on the goal of the segmentation. Our final objective is to segment the image elements in way that is relevant to the tactile transposition (and later, comprehension) of its content.

3.1 Contour Detection

Edge detection consists in the identification of delimitations between meaningful elements in an image, and often relies on the identification of strong oriented gradients of chrominance or luminance. Conventional methods such as high-pass or Canny filters often produce discontinuous contours, which does not allow users to follow contours in order to identify an object. They also tend to over-segment the image by producing contours and noise where no meaningful borders are present due to their sensitivity to lighting and exposure variations. We applied the following methods to the Bayeux tapestry (scene N°16 in Fig. 2a):

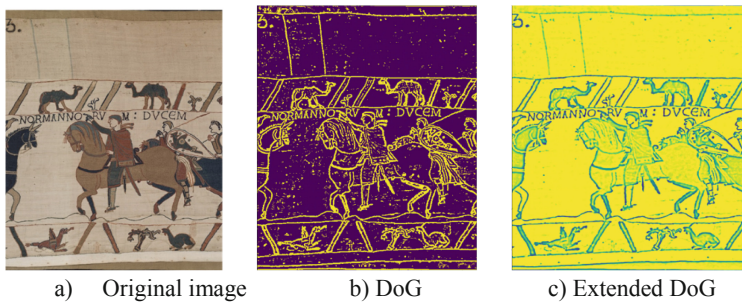


Fig. 2. Bayeux Tapestry contour segmentation with Difference of Gaussians method.

- 1) **The Difference of Gaussians (DoG)** is an edge extraction algorithm that relies on computing the difference between two images blurred with Gaussian kernels of different spread, producing an image where the areas most affected by the Gaussian blurring (i.e. high-frequency spatial information) are over-represented. The DoG algorithm produced continuous edges suitable for tactile transposition (Fig. 2b). Its main inconvenient is the exacerbation of background noise which will result in a grainy tactile transposition if kept. The Extended Difference of Gaussian (XDoG) [10] gives cleaner results (Fig. 2c).

- 2) **The HED (Holistically-Nested Edge Detection)** [11] is an end-to-end deep neural network architecture inspired from Fully Convolutional Networks. Based on human perception in the search for the contours of objects, it uses different levels of perception, structural information and context (Fig. 3b). Nested multi-scale feature learning is taken from deeply supervised nets. The output is combined from multiple scales. The details of the figures on the tapestry are correctly segmented by this method.

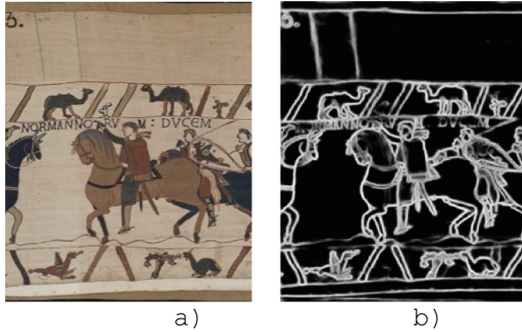


Fig. 3. **a:** A detail of the Bayeux Tapestry (scene n°16), **b:** Contour image obtained with HED, Holistically Nested Edge Detection method applied to the image (right).

- 3) **DeepLab V3** (Fig. 4b) [12]: deep convolutional neural network architecture used for semantic segmentation of natural images. It is characterized by Encoder-Decoder architecture. Despite being pre-trained to detect human in real-world images, it transfers quite well to painting datasets (such as IconArtV2), but less so to the extremely stylized figures of the Bayeux Tapestry. This method would require a complete retraining or a fine-tuning on an image dataset closer to the domain content of the Tapestry, which could be the object of future endeavors.

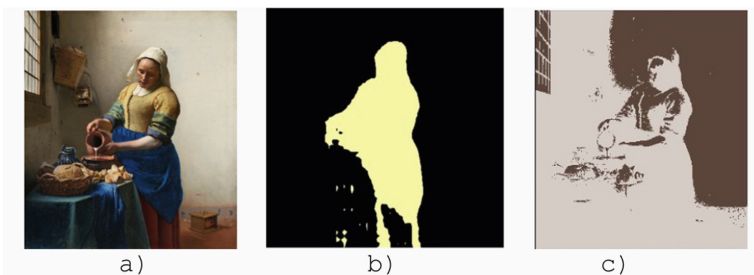


Fig. 4. **a** The painting milkwoman by Jan Vermeer, **b** DeepLab gives the region where the woman is detected in the image. **c** GMM with 2 regions (dark/light)

For the extraction of contours, HED seems to provide the most relevant contours, with the least discontinuities and thus the easiest to follow.

3.2 Clustering

We also considered several clustering-based approaches to segment the Tapestry's content in an unsupervised manner:

- 1) ***K-means clustering algorithm*** aims to partition pixels into a preset number of regions based on their perceived similarity. For the Tapestry, the number of clusters giving the best segmentation results is between 4 to 6 clusters. But some segmentation errors occur where adjacent pixels of different objects have low contrast.
- 2) ***Gaussian mixture model*** (GMM) [13] is a latent probabilistic clustering method which is used to estimate the parameters of distributions representing features potentially involved in the generation of the observed data (i.e. the content of the image), by modeling them as a finite mixture of Gaussians. The results obtained with this method are presented in Fig. 4c.
- 3) ***Slic Superpixels*** [14]: is a clustering method that divides the image into groups of connected pixels based on a weighted combination of their chrominance similarity and their spatial proximity. These superpixels can later be merged in bigger regions based on their semantic meaning. This solution could make a useful interactive tool to help exhibition curators when manually segmenting paintings.

Between the aforementioned methods, K-means and GMM are better suited for the segmentation of the Bayeux Tapestry than DeepLab and Superpixels. Indeed, DeepLab requires further domain specialization for the specific painting style of the tapestry.

K-Means and GMM will therefore be used in our approach for preliminary segmentations. The proposed image processing pipeline is based on several of the methods previously evaluated:

1. Preprocessing is done to choose the object semi-automatically:
 - 1.1. Mean-Shift method for minimizing the number of color clusters and eliminating the noisy background (fabric),
 - 1.2. Grabcut to extract relevant objects and reduce errors due to the influence of spatial proximity during clustering.
2. Integrating extracted contours and clusters:
 - 2.1. We apply a GMM on the extracted image (since K-means resulted in more noisy regions) to obtain better clusters. In Fig. 5b, the methods of K-Means with the cluster number of 6, is applied to the extracted Horseman of Fig. 5a. GMM gives a better result with 4 clusters; the horse and the horseman are better separated (Fig. 5c).
 - 2.2. To produce a more intuitive tactile image with relevant details and relief, we then apply HED to the GMM output. Then contours obtained that way are cleaner and the overall output less noisy than after applying HED to the original image, (Fig. 6).

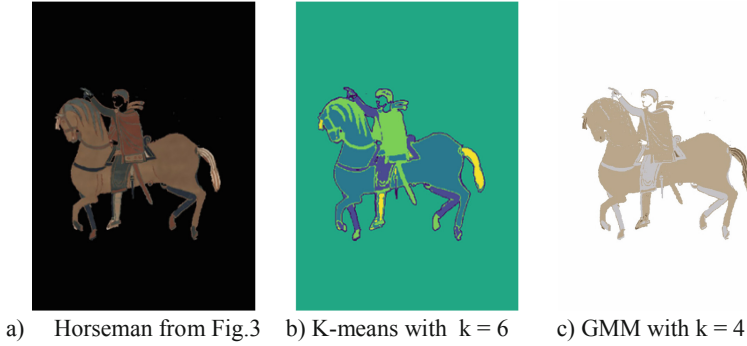


Fig. 5. K-Means and GMM are applied on a Horseman in Bayeux Tapestry.

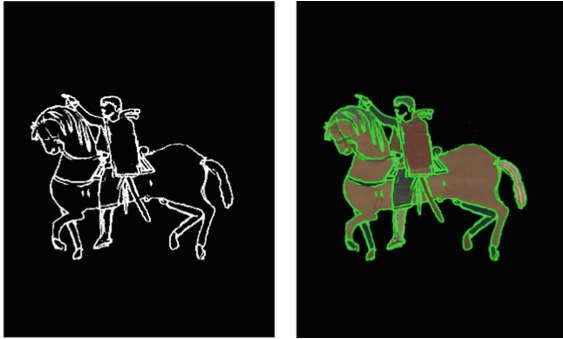


Fig. 6. A horseman with HED applied to the GMM image from Fig. 5c (left), HED contours are highlighted with the integration of region information from GMM image.

4 Audio and Tactile Interfaces

Various interfaces have been designed to allow haptic-based communication, such as taxel matrices and vibrating surfaces. However, most of those devices do not allow the simultaneous display of contours and textures, and are often expensive. After reviewing and evaluating different technologies, we developed our own device: a Force Feedback Tablet (F2T) shown in Fig. 7.

When using the F2T, the exploration is controlled by a micro joystick on a motorized mobile support, which provides information on the underlying image through force-feedback variations, making it possible to create different passive (in response to user movements) and active (guidance) effects. Furthermore, audio information (such as audio-descriptions or ambient sounds) can be linked to the movements of the user to provide additional semantic information. Preliminary evaluations of the F2T were carried out on the recognition of simple geometric shapes, directions, perceived angles and spatial structures (arrangement of rooms) [15].

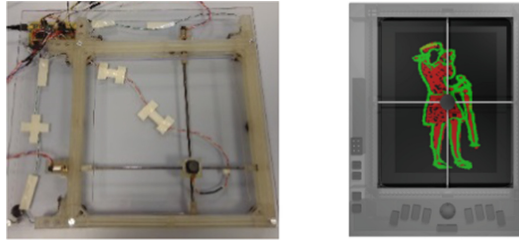


Fig. 7. F2T Prototype (left), concept of a tactile tablet (right).

5 Application to Cultural Heritage

After obtaining the final segmented image (See Fig. 6), we attribute different tactile effects to each cluster based on its meaning (human, horse, etc.). Figure 8 shows the regions and contours of the horseman linked to different tactile textures (i.e. different friction effects of the F2T), making it possible to feel the difference between the two regions. The contours in the image (green overlay) are simulated as domed surface with a small slope which is perceived as an edge during exploration with the F2T.

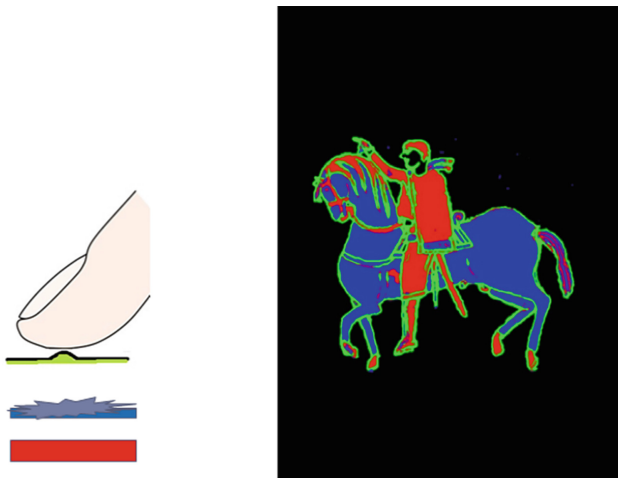


Fig. 8. Application on the Force Feedback Tablet with the contours in green producing a slope effect under the moving touch of a finger, the texture shown in red is simulated with a liquid friction effect and the texture shown with blue with a solid friction effect. (Color figure online)

Future evaluations will be conducted to assess the efficiency of each tactile effect (and its intensity) on the overall understanding of the displayed image.

6 Conclusion and Perspectives

This paper introduced preliminary results on a system facilitating the transposition of graphical artworks into haptic representations, using a specific interface, the F2T, which could improve the accessibility of Museums to VIP and facilitate the discovery of cultural heritage. This innovative audio-haptic interface allows VIP to actively and autonomously explore simplified version of artworks, where meaningful objects are extracted from the overall scene by a combination of edge detection and semantic segmentation algorithms.

Further research and development will be conducted to improve the segmentation pipeline, and to further optimize the conversion of segmented objects into haptic representations. We plan to develop solutions to facilitate a collaborative and iterative segmentation between the selected algorithms and domain specialists such as museum curators, in order to better capture the intentions of artworks' authors.

Acknowledgment. This research work is the result of an ongoing collaboration between the University of Rouen-Normandy, the University of Caen-Normandy, the Region of Normandy, the Department of Seine Maritime, the city of Bayeux, the Museum of Bayeux Tapestry, the Museum "Château de Martainville", the Museum of Quai Branly and several Associations of Visually Impaired People (FAF, ANL, AVH).

References

1. Chen, Y.: Analyse et interpretation d'images à l'usage des personnes non-voyantes, Application à la generation automatique d'images en relief à partir d'équipements banalisés. Thesis, University of Paris 8 (2015)
2. Chen, Y., Haddad, Z., Lopez Krahe, J.: Contribution to the automation of the tactile images transcription process. In: Miesenberger, K., Fels, D., Archambault, D., Peñáz, P., Zagler, W. (eds.) ICCHP 2014. LNCS, vol. 8547, pp. 642–649. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08596-8_99
3. Hatwell, Y.: Le développement perceptivo-moteur de l'enfant aveugle. *Enfance* 1 **55**, 88–94 (2003)
4. Hatwell, Y., Streri, A., Gentaz, E.: Toucher pour Connaître-Psychologie cognitive de la perception tactile manuelle. *Psychologie et science de la pensée*, Presse Universitaire de France, p 193 (2000)
5. Eriksson, Y.: The Swedish Library of Talking Books and Braille. How to make tactile pictures understandable to the blind reader, Disability Information Resources, JSRPD. https://www.dinf.ne.jp/doc/english/Us_Eu/conf/z19/z19001/z1900116.html. Accessed 29 Mar 2020
6. Strickfaden, M., Vildieu, A.: Questioning communication in tactile representation. In: International Conference on Inclusive Design, Include 2011, London (2011)
7. Ansaldi, B.: Perspective and the blind. In: Cocchiarella, L. (ed.) ICGG 2018. AISC, vol. 809, pp. 541–552. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-95588-9_44
8. Romeo, K., Chottin, M., Ancet, P., Pissaloux, E.: Access to artworks and its mediation by and for visually impaired persons. In: Miesenberger, K., Kouroupetroglou, G. (eds.) ICCHP 2018. LNCS, vol. 10897, pp. 233–236. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94274-2_32

9. Romeo, K., Chottin, M., Ancet, P., Lecomte, C., Pissaloux, E.: Simplification of painting images for tactile perception by visually impaired persons. In: Miesenberger, K., Kouroupetroglou, G. (eds.) ICCHP 2018. LNCS, vol. 10897, pp. 250–257. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94274-2_35
10. Winnemöller, H., Kyprianidis, J.E., Olsen, S.C.: XFoG: an eXtended difference-of-Gaussians compendium including advanced image stylization. *Comput. Graph.* **36**(6), 720–753 (2012)
11. Xie, S., Zhuowen, T.: Holistically-nested edge detection. *Int. J. Comput. Vis.* **125**(1–3), 3–18 (2017). <https://doi.org/10.1007/s11263-017-1004-z>
12. Chen, L.-C., et al.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2017)
13. Reynolds, D.A.: Gaussian mixture models. In: *Encyclopedia of Biometrics* (2009)
14. Achanta, R., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC superpixels, Technical report, EPFL, 149300 (2010)
15. Pissaloux, E., Romeo, K., Chottin, M., Ancet, P.: TETMOST-Toucher et Etre Touché, Les vertus inclusives du Mouvement et de la Sensibilité Tactiles. *HANDICAP 2018*, Paris (2018)