MACHINE LEARNING FOR FRESCO RECONSTRUCTION

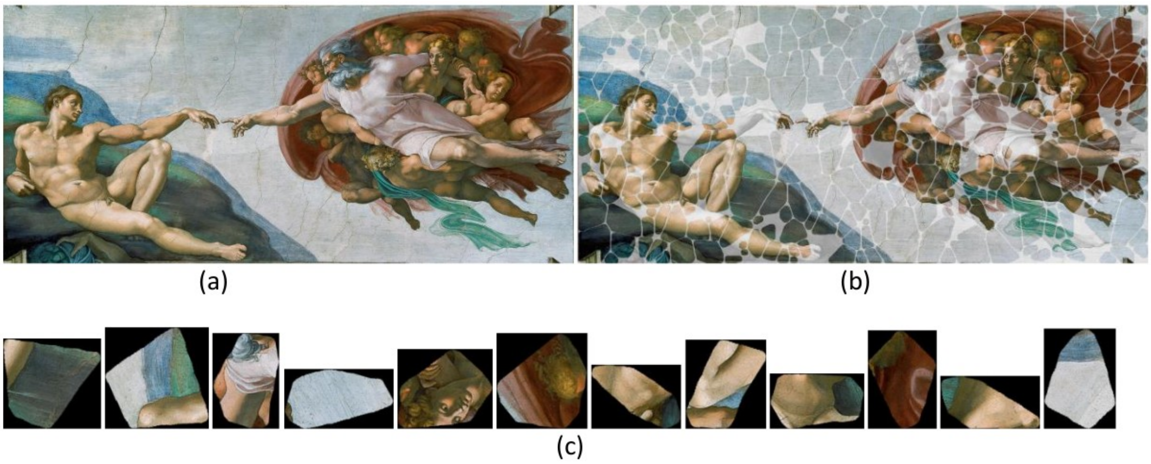

Fig. 1: (a) [input] fresco model, (b) reconstructed fresco and (c) collection of input fragments.

Image-based reconstruction consists in determining the optimal organization between the parts of an object of interest characterized by structured visual information. Such a general problem includes reconstruction of frescoes / paintings or mosaics from their fragments. It has thus straightforward applications in heritage restoration and archeology. In the case of heritage restoration, it mainly deals with the reconstruction of frescoes destroyed during earthquakes or wars, and for which having a model (picture before destruction according to documentary source) appears likely. In this context, the DAFNE challenge [Dondi et al., 2020] proposed simulated data sets for the evaluation of algorithms knowing the fresco model. In the case of applications in archeology, the problem is much more complex since there is no model and the support can be 3D (case of anastylosis, e.g. objects found during excavations or monuments). Compared to a classic "puzzle" or jigsaw problem, the targeted reconstruction presents several additional difficulties: (i) the gigantic number of pieces (fragments), (ii) the variable characteristics of the latter, whether in size or image content, (iii) the irregular shape of the pieces and the deterioration of their edges (erosion of the latter inducing discontinuities between the pieces), (iv) the presence of pieces not belonging to the puzzle, (v) the loss of some pieces preventing the complete reconstruction.

Having successfully participated to the DAFNE challenge, we have already developed some approaches for the case where the fresco model is available. Specifically, we develop a simple but efficient approach relying on key point matching [Lermé et al., 2020], that we compare with an alternative based on maximization of the normalized cross-correlation [Padfield, 2011]. In both cases, a prior estimate of location likelihood maps (derived by backprojection or based on histogram distance) is necessary to control the computation time and raise several ambiguities. These approaches require the adjustment of many parameters. As an alternative, we also develop an approach based on the DeepMatch network [Revaud et al., 2016]. It makes it possible to match images even in the case of non-rigid deformations.

However, in the previous approaches only unary terms, i.e. considering only one fragment at a time, are considered. They allow us to assess (according to the criterion retained by the approach considered) the relevance of a transformation (rotation + translation) applied to the fragment. Now, firstly, these terms cannot be calculated in the absence of the fresco model, and secondly, in homogeneous regions, they do not allow for the reliable estimation of the fragment transformation.

Thus, the objective of this thesis is to develop approaches taking into account binary terms (or even more) to model the interactions between the fragments and to assess the relevance of their placement side by side. We will compare so-called classical approaches and learning approaches. In the first case, the problem is reformulated through a functional to be optimized, functional which potentially includes some data attachment terms (corresponding to unary terms), but also some interaction terms between the fragments (corresponding to binary terms), and possibly some terms corresponding to geometric priors such as the non-overlapping of fragments, the minimization of the area occupied by a given set of fragments. The algorithm for optimizing this functional will be defined jointly with the latter to ensure the possibility of obtaining solutions in a reasonable time. Specifically, for an optimization using graph cuts, the properties of the sub-modularity of the functional must be verified. We also may take inspiration from [Cho et al., 2010; Pomeranz et al., 2011; Gallagher, 2012]: [Gallagher, 2012] proposes a Mahalanobis-like distance for color images while [Cho et al., 2010] compares five consistency measures between adjacent fragments and [Pomeranz et al., 2011] describes how to optimize the “best” one chosen.

In the second case, the problem is reformulated as a learning one. For learning the term unary, we can rely on the architectures proposed for pattern matching, e.g. [Luo et al., 2016; Subramaniam et al., 2016], for applications of stereovision (calculation of the disparity map) and re-identification of people, which will then be compared to DeepMatching [Revaud et al., 2016]. For non-unary terms, we will draw inspiration from the work of [Doersch et al., 2015; Noroozi & Favaro, 2016; Paumard et al., 2018, Paumard et al., 2020]. In [Doersch et al., 2015], the authors train a network to predict the relative position of patches extracted in an image, this in order to get rid of the labeling of objects while learning the more relevant features for object detection / recognition. However, the patches considered are non-contiguous so that high level spatial organization criteria prevail (e.g., the eyes are "above" the mouth, the wheels are "below" the windows etc.). In this sense, it can be linked to the work carried out previously in spatial reasoning [Bloch, 2005; Yang et al., 2017]. [Noroozi & Favaro, 2016] is an extension of [Doersch et al., 2015] allowing one to simultaneously consider the set of patches, while [Gur & Ben-Shahar, 2017] presents an application close to our puzzle problem but the fragment shapes that are rectangular tiles without erosion at the edges. [Paumard et al., 2018 & 2020] apply similar idea to painting fragments relative positioning.

Note that the generalization of the previous approaches taking into account the specificities of our problem and its greater degree of complexity may require modifications to the proposed architectures. During the thesis, the two approaches will first be studied independently (first two years of the thesis). Then (last year of the thesis), they can be combined so as to exploit the strengths of each and / or validate or not their individual decisions.

Location of the Ph.D.: The student will work in the SATIE lab of the Ecole Normale Supérieure at the University Paris-Saclay, first top-ranked french university and in ranked 13th position according to the Shangai academic ranking. Postal address: DIGITEO building (660), avenue des sciences, University Paris-Saclay, 91190 Gif-sur-yvette.

Contact:

- Supervisors: Nicolas Lermé, Sylvie Le Hégarat-Mascle
- E-mails: {nicolas.lerme,sylvie.le-hegarat}@universite-paris-saclay.fr

References

[Bloch, 2005] Bloch, I. (2005). Fuzzy spatial relationships for image processing and interpretation: A review. Image and Vision Computing, 23(2), 89-110.

[Cho et al., 2010] Cho, T. S., Avidan, S., & Freeman, W. T. (2010). A probabilistic image jigsaw puzzle solver. In IEEE Conference on Computer Vision and Pattern Recognition, 183-190.

[Doersch et al., 2015] Doersch, C., Gupta, A., & Efros, A. A. (2015). Unsupervised visual representation learning by context prediction. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1422-1430).

[Dondi et al., 2020] Dondi, P., Lombardi, L., & Setti, A. (2020). DAFNE: A dataset of fresco fragments for digital anastylosis. Pattern Recognition Letters, 138, 631-637.

[Gallagher, 2012] Gallagher, A. C. (2012). Jigsaw puzzles with pieces of unknown orientation. In IEEE Conference on Computer Vision and Pattern Recognition, 382-389.

[Gur & Ben-Shahar, 2017] Gur, S., & Ben-Shahar, O. (2017). From square pieces to brick walls: The next challenge in solving jigsaw puzzles. In Proceedings of the IEEE International Conference on Computer Vision, 4029-4037.

[Lermé et al., 2020] Lermé, N., Le Hégarat-Mascle, S., Boyang, Z. & Aldea, E. (2020). Fast and

Efficient Reconstruction of Digitized Frescoes. Pattern Recognition Letters, 138, 417-423.

[Luo et al., 2016] Luo, W., Schwing, A. G., & Urtasun, R. (2016). Efficient deep learning for stereo matching. In IEEE Conference on Computer Vision and Pattern Recognition, 5695-5703.

[Noroozi & Favaro, 2016] Noroozi, M., & Favaro, P. (2016). Unsupervised learning of visual representations by solving jigsaw puzzles. In European Conference on Computer Vision, 69-84.

[Padfield, 2011] Padfield, D. (2011). Masked object registration in the Fourier domain. IEEE Transactions on Image Processing, 21(5), 2706-2718.

[Paumard et al., 2018] Paumard, M. M., Picard, D., & Tabia, H. (2018). Jigsaw puzzle solving using local feature co-occurrences in deep neural networks. In IEEE International Conference on Image Processing, 1018-1022.

[Paumard et al., 2020] Paumard, M. M., Picard, D., & Tabia, H. (2020). Deepzzle: Solving Visual Jigsaw Puzzles with Deep Learning and shortest path optimization. IEEE Transactions on Image Processing, 29, 3569-3581.

[Pomeranz et al., 2011] Pomeranz, D., Shemesh, M., & Ben-Shahar, O. (2011). A fully automated greedy square jigsaw puzzle solver. In IEEE Conference on Computer Vision and Pattern Recognition, 9-16.

[Revaud et al., 2016] Revaud, J., Weinzaepfel, P., Harchaoui, Z., & Schmid, C. (2016). Deepmatching: Hierarchical deformable dense matching. International Journal of Computer Vision, 120(3), 300-323.

[Subramaniam et al., 2016] Subramaniam, A., Chatterjee, M., & Mittal, A. (2016). Deep neural networks with inexact matching for person re-identification. In Advances in Neural Information Processing Systems, 2667-2675.

[Yang et al., 2017] Yang, Y., Atif, J., & Bloch, I. (2017). Raisonnement abductif en logique de description exploitant les domaines concrets spatiaux pour l'interprétation d'images. Revue d'Intelligence Artificielle, 31(1-2), 11-39.