Fast and Efficient Reconstruction of Digitized Frescoes

Nicolas Lermé[a,**], Sylvie Le Hégarat-Mascle[a], Boyang Zhang[a], Emanuel Aldea[a]

[a] *Université Paris-Saclay, SATIE Laboratory UMR 8029, 91190 Gif-sur-Yvette, France*

ABSTRACT

Virtually recomposing destroyed frescoes is of great importance for heritage conservation. Given a digitized fresco image and a digitized set of fragments, such a problem is challenging due to the potentially large number of fragments, their irregular shape, uniqueness and non-overlapping constraints, the possible absence of fragments and the possible presence of small, homogeneous, eroded and/or spurious fragments. To cope with these specific features, we propose in this paper a fast and efficient non-dense approach benefiting from previous developments in pattern matching. Preliminary experiments led on simulations exhibit a mean accuracy above 90% with a mean translation error of less than 4 pixels and a mean orientation error of about 1 degree. An analysis of fresco and fragment features impacting the algorithm is also provided. Compared to a dense approach and the recent DeepMatch approach, the proposed one remains competitive both in running time and accuracy.

1. Introduction

Earthquakes can cause damage to a nation not only with respect to the properties of individuals and companies, but also with respect to its the artistic heritage. For instance, when old churches or monuments are destroyed, the frescoes painted on their walls or ceilings suffer the same fate.

Fortunately, nowadays, images of almost all artistic works have been acquired, so that the reconstruction of these frescoes from the set of their pieces is a task not only possible for humans but even automated using computer vision. Virtual reconstruction also maximizes the chances of succeeded restorations by minimizing the handling of fragments.

Let I denote the image of the fresco before the destruction and let $\mathcal{F}$ be the set of the fragments of the fresco. The considered problem can be formulated as follows: Estimate for each fragment in $\mathcal{F}$, the parameters of the geometrical transformation that, when applied to the fragment, generates an image $\widetilde{I}$ as close as possible to the model image I. Such a formulation allows for the presence of spurious fragments for which the derived set of geometrical parameters should be the empty set.

Most of the proposed works for solving jigsaw puzzles exploit fragment border consistency [1] and/or fragment shapes [2]. However, our problem includes missing fragments representing up to 50% of the fresco area and fragment erosion has degraded their shape. On the other hand, the fresco model is available and fragment radiometry is preserved so that fresco reconstruction can be handled as a pattern matching problem with specific constraints. Pattern matching applications range from estimating geometrical relationships between images or frames (e.g. for image co-registration, e.g. [3]) to object recognition (e.g., [4]) within an image. We distinguish three main kinds of approaches.

The first one is based on the evaluation of the relevance of a geometric transformation between two images, via a criterion computed on the whole image fields, such as normalized cross-correlation [5]. While these approaches are robust against corrupted data, a core issue is the derivation of the subset of transformations to evaluate.

The second kind of approaches is based on hand-crafted features both for the detection of keypoints and for their description

The third kind of approaches is inherited from the success of deep networks for the past decade. For pattern matching, DeepMatch [6, 7] has been proposed and is widely used for instance for object tracking. It applies a multi-layer deep convolutional architecture to yield possibly non-rigid matches between a pair of images. Thus, despite its high performance, the fact that matching constraints (object rigidity, non-overlapping) are different may be a challenge for its use in our fresco reconstruction

**Corresponding author: Tel.: +33(0)1.69.15.78.04
e-mail: nicolas.lerme@universite-paris-saclay.fr (Nicolas Lermé)

problem.

Let us recall the specific features of fresco reconstruction: (i) the number of fragments ranges from several hundreds to a few thousands, definitively excluding exhaustive exploration of the solution space; (ii) a fragment does not correspond to a rectangular tile, i.e. in the bounding box containing the fragment, only a subset of the pixels have to be considered; (iii) in the solution, the fragments shall not overlap. Even more, there may be a gap between adjacent fragments depending on their level of erosion; (iv) some parts of the fresco may be lost (no fragment covers them); (v) any fragment must appear at most one time in the solution: 0 times for the spurious fragments and 1 time for the others; (vi) some fragments may be very small (less than a few tens of pixels) and/or homogeneous. The algorithm proposed in this study takes into account all these features while benefiting from previous developments in pattern matching.

The rest of this paper is organized as follows. The proposed approach is detailed in Section 2. Section 3 presents both a comparison with alternative approaches (even if much more various alternatives have been tested during the aforementioned competition) and an analysis of the fresco and fragment features impacting the algorithm performance. Section 4 concludes by raising the main perspectives of this work.

2. Proposed approach

Due to the important number of fragments to process, we oriented our solution as the second kind of approaches mentioned in Section 1, namely based on hand-crafted features. Indeed, it allows us to estimate the geometrical transformation directly from matched keypoint pairs saving numerous calculations. However, such an approach is sensitive to outliers. Besides, we have to cope with border issues due to the fact that some fragments are very small (less than a few tens of pixels) and finally to carry out the uniqueness and non-overlapping constraints. Before detailing how these points have been addressed, let us introduce the notations for the rest of the paper.

For a positive integer $c > 0$, let us denote by $D = [0, 1]^c$ the normalized image intensities (in our experiments, only color images are handled, so $c = 3$). We define the fresco image as $I \in D^Y$, where the domain Y denotes a connected subset of $\mathbb{R}^2$. Let $\mathcal{F}$ be the finite set of fragment images whose cardinality is denoted $\sharp\mathcal{F}$. For any $j \in \{1, \ldots, \sharp\mathcal{F}\}$, a fragment image is defined as $f_j \in D^{X_j}$, where the domain X_j denotes a connected subset of $\mathbb{R}^2$. Then, our problem boils down to finding, for each fragment image f_j, (i) the translation vector $\mathbf{t}_j$ and the rotation angle θ_j allowing for the fragment placement in the fresco domain Y which is the most consistent with the whole fresco image I and (ii) its label $l_j \in \{0, 1\}$ representing its belonging or not in the fresco (fragment f_j belongs to the fresco if and only if $l_j = 1$) and therefore the relevance of $\left(\mathbf{t}_j, \theta_j\right)$.

2.1. Using color information

The goal of this step is to improve the matching between keypoints. The point for such purpose is to restrict the considered area in the fresco with respect to the considered fragment. Indeed, by doing so, we will drastically reduce the amount of putative keypoints in the fresco and thus reduce both matching ambiguities and matching errors. This can be simply achieved if we are able to derive, for each fragment, a rough estimation of its location in the fresco. In addition to computational efficiency, we set two constraints for this rough estimation. It shall be (i) based on information that is complementary to the one used in the keypoint detector/descriptor; (ii) rotation invariant since, the fragment orientation is still unknown.

Then, regarding the first constraint, let us remark that the considered data images (fresco and fragments) are in color, whereas most keypoint descriptors are defined for grayscale images. The color extension of these latter ones is generally a simple concatenation of the descriptors computed for each of the color channels, respectively. However, actual 3D color histograms are much more discriminating than concatenated ones.

Dealing with the second constraint, histograms provide a global description of an area without regards to the spatial organization of its elements. However, such a description remains sensitive to rotation depending on the shape of the area considered for histogram estimation. Specifically, the only shape that is rotation invariant is a disk. In this study, we have approximated the disk shape by a number of rectangles (denoted by $N \in \mathbb{N}_{>0}$) inscribed in it, whose sizes are determined by uniformly sampling angles along the arc of one of the quadrants of the disk. This strategy allows us to benefit from the computational efficiency of integral images or integral histograms [8], while drastically reducing sensitivity to rotation. Then, from the histogram representation, most popular approaches for deriving confidence maps of pattern location in an image are based on histogram similarity or distance computation [9]. On the one hand, these techniques are often designed for robustness to occlusion phenomena, which are completely absent in our case. One the other hand, they do not handle actual 3D histograms. Therefore, we prefer an older approach, namely histogram backprojection [10], that meets our application conditions while being simple and efficient.

The chosen strategy is as follows. Firstly, the fresco 3D color histogram is computed in a chosen color space [1] using $B \in \mathbb{N}_{>0}$ bins per channel. Then, for any fragment f_j, the optimal inscribed (denoted by $(c_j, r_j) \in (\mathbb{R}^2 \times \mathbb{R}_{>0})$) and circumscribed (denoted by $(C_j, R_j) \in (\mathbb{R}^2 \times \mathbb{R}_{>0})$) circles are computed.

Secondly, the estimated inscribed circle is approximated by N tiles and the 3D color histogram (B^3 bins) of the fragment (restricted to N tiles) is derived and computed in the same color space as I using again B^3 bins (B bins per channel).

Thirdly, the consistency between the fragment histogram and the fresco local ones is mapped. Using for this the backprojection technique of [10], the 'consistency' measure relies on an histogram ratio which is spatially low-pass filtered. Note that using an averaging filter allows us to benefit from integral images [11] and to make the complexity for such step independent of the fragment size.

Fourthly, the resulting confidence map is thresholded to derive a binary map. The threshold, denoted by $\tau_{cm} \in [0, 1]$, is

[1] This choice is not discussed here. Even if the results will correspond to RGB color space, the proposed approach applies to any 3D color space.

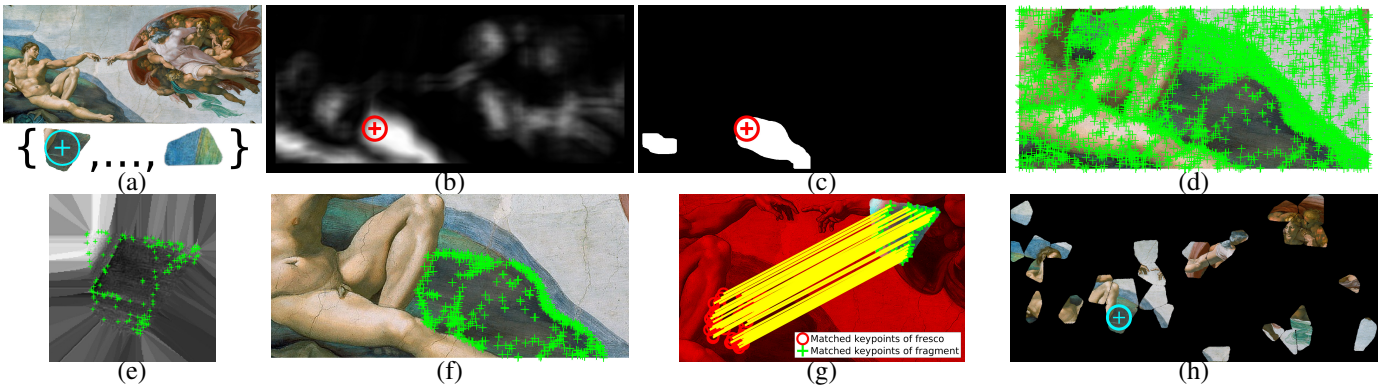

Fig. 1. Steps of the proposed approach for reconstructing digitized frescoes from a given fresco image and a set of fragments (a). For any fragment (the considered one appears in cyan here), a confidence map is computed indicating for any pixel the amount of similarity in terms of image intensities between the fresco and the fragment from the area covered by the inner disk of the latter centered at that pixel (b). This map is thresholded and a morphological dilation is applied to take into account the spatial extent of the fragment (c). In (b) and (c), the red cross denotes the true location of the inscribed circle of the fragment. Borders of the fresco image and borders of any fragment are extended and keypoints (green crosses) are extracted from both of them (d),(e). Using (c), keypoints from (d) are discarded (f). Resulting keypoints are matched (g) and fragments satisfying specific constraints are greedily placed (h).

set to ensure that the actual match of the fragment is included in the binary map. The benefit of using binary maps is that they can be efficiently stored to minimize the memory footprint.

Lastly, we propose a final step that carries out a refinement of the previous binary map to be able to use it as a research area when matching fresco and fragment keypoints (as presented in the next section). Indeed the computed confidence values correspond to a fragment centered on its inscribed circle center. Then, to take into account the spatial extent of the fragment domain, binary maps are dilated (in an isotropic way in the absence of orientation prior knowledge) to provide so-called *keypoint area maps*. The radius of the dilation depends on the radii of the inner and circumscribed circles as well as their respective centers. Empirically, we set the dilation radius equal to $\frac{\alpha}{2}(R_j - r_j + \|C_j - c_j\|_2)$, where $\|.\|_2$ is the L^2 norm in $\mathbb{R}^2$ and $\alpha \in [0, 1]$ is a prior parameter aiming to cope with the fact that, when the centers c_j and C_j are distinct, this can unnecessarily enlarge the fragment at some places.

2.2. Handling fresco and fragment borders

In the considered application, the fragments may have any geometric shape so that when considering their bounding box, some pixel values are missing.

Two solutions could be envisaged when faced to the previous issue: the modification of the keypoint detector-descriptor, and the data modification. For the first solution, keypoint detector adaptation goes through the modification of the used linear filters for the detector based on detection of important gradient values in two perpendicular directions (such as for those inspired from Harris' one [12], including SIFT [13]) or through the adaptation of the considered set of pixels for detectors based on values along a chain such as those inspired by FAST [14]. For the descriptor however, the situation is more complex. Indeed, if the aim of keypoint description is the matching based on a distance (or similarity) between descriptors, it appears rather tricky to modify one descriptor (due to missing pixels) independently of the second one (having a different local context).

Thus, despite the attractiveness of the first solution, we turn towards the second solution, which presents in addition the benefit of processing in the same way the fresco borders. Specifically, pixel values are extrapolated in order to fill the missing parts. Note that as illustrated on Fig. 1, the extrapolation prevents the detection of keypoints on these parts while allowing keypoint descriptor computation.

2.3. Handling non-overlapping between fragments

The fragment non-overlapping prior is used as a validity test for fragment transformation (translation and rotation). Indeed, for any $j \in \{1, \ldots, \sharp\mathcal{F}\}$, denoting by $\mathcal{T}_{\mathbf{t}_j,\theta_j}(X_j) \in \mathbb{R}^2$ the result of the geometric transformation $\left(\mathbf{t}_j, \theta_j\right)$ of its domain X_j, we can check whether $\mathcal{T}_{\mathbf{t}_j,\theta_j}(X_j)$ is included in Y (the fresco domain), and whether it does not overlap the other domains $\mathcal{T}_{\mathbf{t}_k,\theta_k}(X_k), \forall k \neq j$. Practically, to take into account some imprecision in the geometrical transformation estimation as well as its reliability, the previous tests are implemented as follows. The relative overlapping area between $\mathcal{T}_{\mathbf{t}_j,\theta_j}(X_j)$ and Y

$$\frac{\mathcal{A}\left(\mathcal{T}_{\mathbf{t}_j,\theta_j}(X_j) \cap Y\right)}{\mathcal{A}\left(X_j\right)} \in [0, 1], \tag{1}$$

($\mathcal{A}$ denotes the area of a set) shall be greater than a given threshold (denoted by $\omega_{in} \in [0, 1]$). Similarly, we relax the overlapping constraint so that the relative overlapping area between $\mathcal{T}_{\mathbf{t}_j,\theta_j}(X_j)$ and $\mathcal{T}_{\mathbf{t}_k,\theta_k}(X_k)$, for any $(k, j) \in \{1, \ldots, \sharp\mathcal{F}\}^2$, s.t. $\forall k \neq j$

$$\frac{\mathcal{A}\left(\mathcal{T}_{\mathbf{t}_j,\theta_j}(X_j) \cap \mathcal{T}_{\mathbf{t}_k,\theta_k}(X_k)\right)}{\min\left\{\mathcal{A}\left(\mathcal{T}_{\mathbf{t}_j,\theta_j}(X_j)\right), \mathcal{A}\left(\mathcal{T}_{\mathbf{t}_k,\theta_k}(X_k)\right)\right\}} \in [0, 1], \tag{2}$$

shall be lower than a given threshold (denoted by $\omega_{ovl} \in [0, 1]$).

2.4. Overall algorithm

The steps of the proposed algorithm are illustrated and summarized in Fig. 1 for the fragment marked by a cyan cross.

Firstly, *keypoint area maps* providing rough locations of fragments (without regards to their rotation) are computed in parallel for all fragment images $\{f_j\}_{j=1}^{\sharp\mathcal{F}}$ with respect to the fresco image I (see Section 2.1).

Secondly, the set of keypoints $\mathcal{K}$ and their descriptors are extracted on the extrapolated grayscale fresco image; keypoints and descriptors are then extracted for all extrapolated fragment images $\{f_j\}_{j=1}^{\sharp\mathcal{F}}$ in parallel (see Section 2.2). FAST [14] and BRISK [15] were empirically evaluated as best candidates for feature detector and descriptor, respectively. For FAST, the parameter controlling the minimum accepted quality of corners is denoted by $\eta_{mqc} \in [0, 1]$ while the parameter controlling the minimum intensity difference between corner and surrounding region is denoted by $\eta_{mc} \in]0, 1[$. For BRISK, default values of the parameters are used. For each fragment f_j, let $\mathcal{K}_j \subset \mathcal{K}$ be the subset of keypoints lying in the j^{th} *keypoint area map*. Matches are then searched between fragment keypoints and $\mathcal{K}_j$ elements (i.e. color filtered fresco keypoints). The strongest matches are selected with a threshold denoted by $\rho_{mt} \in]0, 100]$. Ambiguous matches are rejected with a parameter denoted by $\rho_{mr} \in]0, 1]$. The geometric transformation matrix is estimated based matched keypoints pairs using SVD and RANSAC algorithms with default parameters to remove remaining outliers.

Finally, once geometric transforms of all fragments are gathered, they are sequentially processed in decreasing order of their number of matched keypoints pairs and greedily placed to form the recomposed fresco $\widetilde{I}$ if they lie inside the fresco and do not significantly overlap with previously placed fragments (see Eq. (1) and Eq. (2) in Section 2.3).

Relying on pattern matching, fresco reconstruction involves the following stages: for each fragment, (i) determine its location in the whole fresco domain, (ii) evaluate if the derived solution (possibly empty) is reliable enough. With respect to the proposed algorithm steps, the fragment location estimation is carried by constraining keypoint matching based on *keypoint area maps* (derived from confidence maps), and solution relevance is both controlled during keypoint matching (ρ_m parameter, RANSAC criterion) and taken into account during sequential placement of segments.

The results presented in next section have been derived by setting the parameters as follows: $\alpha = 0.5$, $N = 3$, $B = 16$, $\tau_{cm} = 0.85$ (confidence maps), $\eta_{mqc} = 0.03$, $\eta_{mc} = 0.01$ (features detection thresholds), $\rho_{mr} = 0.6$, $\rho_{mt} = 50$ (matching), and $\omega_{in} = 0.9$, $\omega_{ovl} = 0.1$ (constraints). Bilinear interpolation is also used for generating the recomposed fresco.

3. Experimental results

3.1. Dataset

The experiments use the DB1 dataset designed for the DAFNE challenge [2], composed of 55 frescoes. For each fresco, its high resolution image along with a set of fragment images is provided. The set of fragments results from a random plane tessellation with suitable statistics followed by an erosion process applied to each fragment. For each fresco, 18 distinct sets of fragments are available, therefore providing a total of 990 instances. This dataset is designed to be realistic, natural and challenging for the cultural heritage domain in terms of resolution, diversity in scenes and pictorial assets. The characteristics of this dataset are summarized in Table 1. For each instance, the list of fragments of the ideal reconstructed fresco is also available as well as the "ground truth" translation and the angle of rotation of any fragment composing it.

3.2. Evaluation protocol

To evaluate the performance of the proposed approach and the alternative ones, five metrics are considered: accuracy (ACC), F-measure (FM), Mean Translation Error (MTE), Mean Orientation Error (MOE) and Relative Cover Rate (RCR). Given the set of fragment images $\mathcal{F}$, let us define $\mathcal{S} = (\mathbf{t}_i, \theta_i, l_i)_{i=1}^{\sharp\mathcal{F}}$ and $\mathcal{S}' = (\mathbf{t}'_i, \theta'_i, l'_i)_{i=1}^{\sharp\mathcal{F}}$ as the solution obtained by some approach and the ground truth, respectively. Any fragment $i \in \{1, \ldots, \sharp\mathcal{F}\}$ is part of $\mathcal{S}$ (resp. $\mathcal{S}'$) if and only if $l_i = 1$ (resp. $l'_i = 1$). For convenience, for any fragment $i \in \{1, \ldots, \sharp\mathcal{F}\}$ such that $l_i = 0$ or $l'_i = 0$, a dummy translation (i.e. $\mathbf{t}_i = \mathbf{t}'_i = (0, 0)^T$) and a dummy orientation (i.e. $\theta_i = \theta'_i = 0$) is introduced.

For some non-negative tolerance thresholds $\tau_t \in \mathbb{R}_{\geq 0}$ and $\tau_r \in \mathbb{R}_{\geq 0}$ on translation and orientation respectively, accuracy and F-measure are respectively denoted by $ACC_{\tau_t}^{\tau_r}(\mathcal{S}, \mathcal{S}') \in [0, 100]$ and $FM_{\tau_t}^{\tau_r}(\mathcal{S}, \mathcal{S}') \in [0, 100]$. In the latter metrics, any fragment $i \in \{1, \ldots, \sharp\mathcal{F}\}$ in $\mathcal{S}$ is considered to be a true positive with respect to $\mathcal{S}'$ if and only if the distance between $\mathbf{t}_i$ and $\mathbf{t}'_i$ is smaller than τ_t and the distance between θ_i and θ'_i is smaller than τ_r. False positives, true negatives and false negatives can then be deduced. Other (parameter-free) metrics are defined as

$$MTE(\mathcal{S}, \mathcal{S}') = \frac{\sum_{i=1}^{\sharp\mathcal{F}} l_i l'_i \|\mathbf{t}_k - \mathbf{t}'_k\|_2}{\sum_{i=1}^{\sharp\mathcal{F}} l_i l'_i} \in \mathbb{R}_{\geq 0},$$

$$MOE(\mathcal{S}, \mathcal{S}') = \frac{\sum_{i=1}^{\sharp\mathcal{F}} l_i l'_i \min\left(|\theta_i - \theta'_i|, 2\pi - |\theta_i - \theta'_i|\right)}{\sum_{i=1}^{\sharp\mathcal{F}} l_i l'_i} \in \mathbb{R}_{\geq 0},$$

$$RCR(\mathcal{S}, \mathcal{S}') = \frac{\left|\mathcal{A}(\cup_{\substack{i\in\mathcal{F}\\ l_i=1}} \mathcal{T}_{\mathbf{t}_i,\theta_i}(X_i)) - \mathcal{A}(\cup_{\substack{i\in\mathcal{F}\\ l'_i=1}} \mathcal{T}_{\mathbf{t}'_i,\theta'_i}(X_i))\right|}{\mathcal{A}(\cup_{\substack{i\in\mathcal{F}\\ l'_i=1}} \mathcal{T}_{\mathbf{t}'_i,\theta'_i}(X_i))} \in \mathbb{R}_{\geq 0}.$$

3.3. Comparison against alternative approaches

Beyond the comparison with various approaches proposed by different authors which has been performed through the DAFNE challenge, here we focus on the two following alternatives to the proposed method that we implemented for testing.

The first one is based on DeepMatch [7]. As explained in Section 1, it is a powerful approach widely used in pattern matching such as for optical flow estimation [16]. Thanks to the consideration of independent patches, DeepMatch allows for correspondences between non rigid objects. In our case, since fragments are rigid, this feature will be used instead for transformation validation. Indeed, the output of DeepMatch is

[2] https://vision.unipv.it/DAFchallenge/DAFNE_dataset/

a set of matched patches so that their geometrical consistency can be evaluated along with the global transformation estimation. Practically, the set of the matched patches contains numerous outliers so that we cannot reconstruct the fresco without filtering them. For this, firstly the tried-and-tested method that is RANSAC [17] has been applied for affine transformation estimation. Secondly, we perform a filtering based on the scale parameter constraint after estimating this latter assuming the conservation of distances (true in our case of affine transformation applied to rigid objects).

The second alternative is based on Normalized Cross-Correlation (NCC) maps derived using Fast Fourier Transforms (FFT). [18] proposes a NCC computation that takes into account the masked feature of some areas in the images. It allows us to relax the constraint of tile shape for the fragments, and also to mask the areas of the fresco where previously processed fragments have been placed. However, if NCC is a powerful criterion to estimate the translation, it is very sensitive to orientation, so that this latter has to be estimated beforehand. We turned to a solution based on the comparison of histograms of gradient orientations, using a histogram distance robust to occlusion, namely [19] distance.

The comparison has been done on a subset of the DB1 dataset consisting of 5 frescoes having varying features (see Fig. 2 caption). The results are presented in Fig. 2 and Tab. 2. The frescoes are presented by rough order of complexity in terms of total number of fragments and number of spurious fragments.

First of all, we note that the proposed approach is systematically superior according to ACC and FM indicators. We also notice that, irrespective of the considered approach, ACC (or FM) values achieved for varying thresholds (τ_t and τ_r) are very close, meaning that when fragments are placed, the estimated transformation is very precise. Comparing the approaches, one can notice that the "DeepMatch approach" is more cautious since it places a lower number of fragments (lower ACC and FM values), allowing for lower MTE and MOE in some cases. It generally entails a lower RCR value except for fresco ♯19, which means that most false negative fragments are very small fragments. Finally, note that the "Dense approach" achieves lower performance than the proposed one for every selected fresco. It exhibits a clearly lower performance than DeepMatch for fresco ♯20 but provides a better result for fresco ♯38.

In summary, this comparison confirms the challenge results, namely the performance of the proposed approach with respect to alternative ones (here based on different principles). Finally, let us mention that the alternative approaches are far slower in terms of execution times.

3.4. Performance analysis

Statistics of performance metrics for reconstructed frescoes from the entire DB1 dataset are summarized in Tab. 3 with tolerance thresholds in translation and rotation of $\tau_t = 10$ pixels and $\tau_r = 5$ degrees. Globally, the proposed approach offers good performance for 34 instances out of 990 for which $ACC = FM = 100\%$ and $RCR = 0\%$ (20 instances among them contain spurious fragments).

However, since one can observe a high variability from one fresco to another, we specify the performance per fresco. Figure 3 shows the ACC and MTE box plots when varying the set of fragments. It allows us to visualize the impact of the image content not only on the reconstruction performance but also on its robustness with respect to the fragment sets. Regardless of the fragmentation of the fresco, this shows the poor performance achieved on frescoes ♯13, ♯19 and ♯36. Indeed, these frescoes exhibit repetitive patterns, hence making difficult matching of feature descriptors.

To further investigate the features impacting performance, we analyzed the fragment characteristics for the four clusters of TP, FP, TN or FN fragments. Tested characteristics include for each fragment, the number of keypoints, the number of matched keypoints, the size and the 'contrast' (measured by mean standard deviation over color channels). Unsurprisingly, some features are at least roughly correlated like the number of keypoints and the fragment size. Also unsurprisingly, most FN present very low contrast, size and therefore keypoint numbers. Figure 4 shows a typical example of obtained plots. It has been drawn for fresco ♯45 and for independent features 'contrast' and 'size'. We note that the four clusters partly overlap (the overlap being dependent on the considered fresco), but the impact of fragment features on the cluster membership is unquestionable.

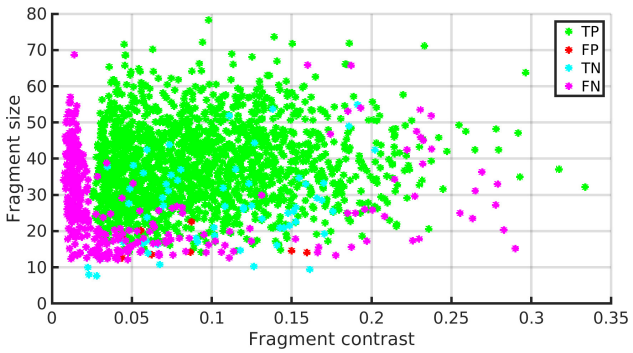

Fig. 4. Example of clusters TP, FP, TN and FN of fragments according to their (contrast, size) features concerning the fresco ♯45.

Finally, the global complexity of the whole algorithm is difficult to estimate theoretically since it depends not only on the pixel number (that is known) but also on the keypoint number (that is unknown). However, thanks to the representativeness of the DAFNE database, we are able to derive empirical values and main tendencies of the proposed algorithm. Then, Fig. 5 depicts how the execution time is affected by the number of fragments and their size for a 16-cores Intel i7-6900K CPU @ 3.2GHz with 64GB of RAM running MATLAB R2017b. Execution times range from about 9 secs to 18 mins with a median value of lower than a minute.

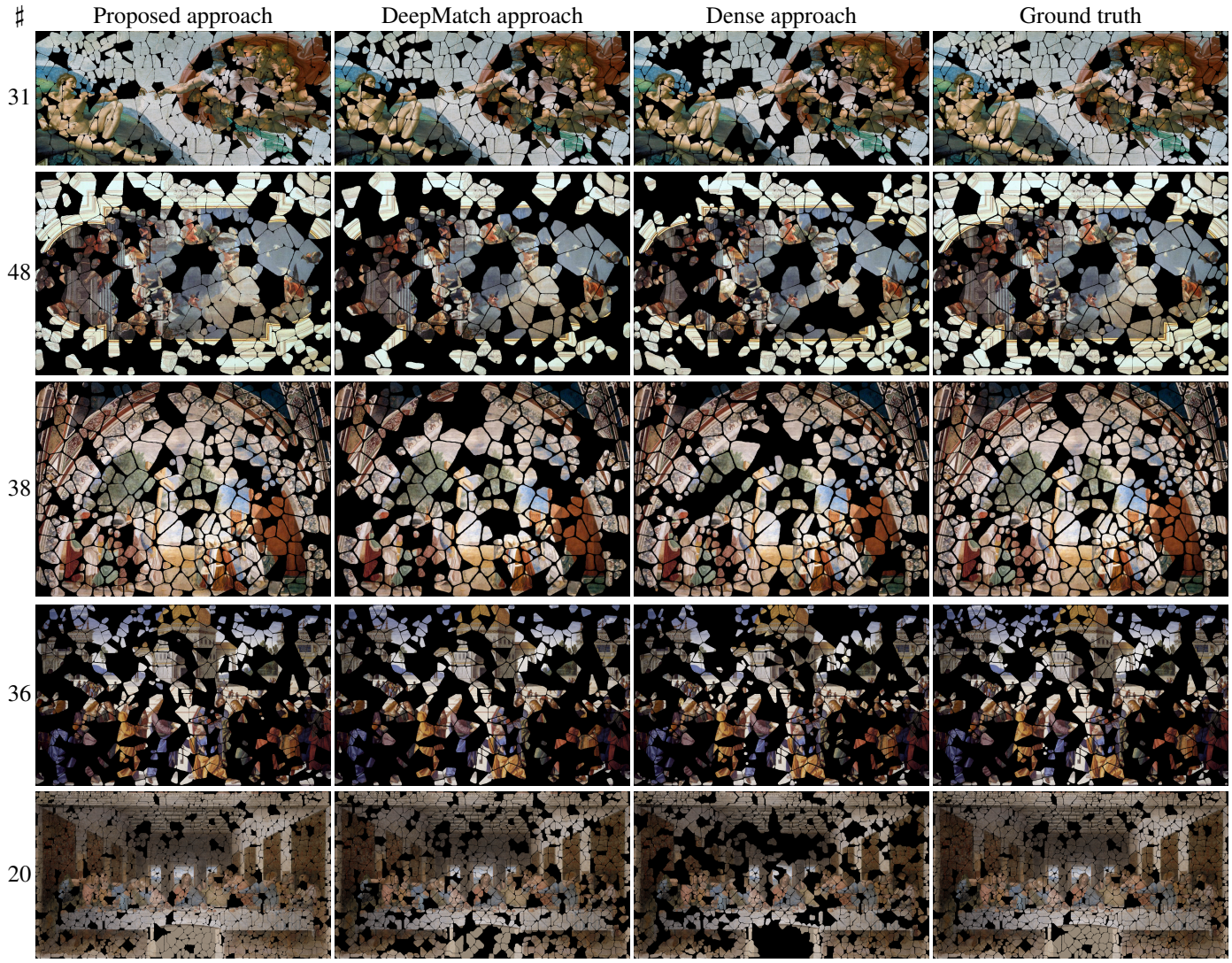

Fig. 2. Qualitative comparison between the proposed approach and the alternative ones. Fresco size varies from 850×1231 **pixels (fresco** ♯48**) to** 5193×2926 **pixels (**♯20**), number of fragments varies from** 299 (♯48) **to** 856 (♯20)**, percentage of spurious fragments varies from** 0% (♯20, ♯31) **to** 15.8% (♯36)**, and area of missing fragments varies from** 17.4% (♯20) **to** 55.5% (♯36)**.**

Table 1. Summary of the characteristics of the entire DB1 dataset.

	Fresco size (in pixels)	♯ fragments	Ratio of true fragments (%)	Ratio of spurious fragments (%)	Cover of true fragments (%)	Cover of spurious fragments (%)
Min.	730×826	101	60.74	0	26.38	0
Max.	4933×3856	2196	100	39.25	88.03	8.01
Median	2123×821	418	93.45	6.55	64.54	1.32

Table 2. Quantitative comparison between proposed approach and alternative ones on a subset of frescoes. ACC and FM measurements are given with and without brackets for $(\tau_t = 10, \tau_r = 5)$ **and** $(\tau_t = 20, \tau_r = 10)$**, respectively. Best measurements are shown in bold for each metric and each instance.**

	Proposed approach					DeepMatch approach					Dense approach				
	ACC	FM	MTE	MOE	RCR	ACC	FM	MTE	MOE	RCR	ACC	FM	MTE	MOE	RCR
31	**91.75 (91.75)**	**95.7 (95.7)**	2.03	**0.56**	**1.48**	74.00 (74.30)	85.00 (85.20)	**0.92**	2.13	7.80	82.22 (84.44)	90.24 (91.57)	1.86	2.46	15.70
48	**84.28 (84.28)**	**89.72 (89.72)**	5.16	1.64	**6.31**	64.50 (64.50)	73.50 (73.50)	**1.64**	**1.05**	24.14	60.87 (64.55)	71.53 (74.88)	40.34	12.70	20.33
38	**92.22 (92.22)**	**95.29 (95.29)**	**3.83**	**1.25**	**1.99**	69.70 (70.90)	78.80 (79.80)	9.05	2.17	17.10	82.13 (83.86)	88.81 (90.00)	5.30	3.31	11.32
36	**90.24 (90.5)**	**93.78 (93.96)**	21.5	3.16	**2.03**	76.50 (76.70)	83.70 (83.90)	**1.00**	**0.53**	9.72	72.49 (74.34)	82.31 (83.70)	23.86	6.03	7.36
20	**100 (100)**	**100 (100)**	**1.93**	**0.05**	**0**	77.50 (77.70)	87.40 (87.40)	5.39	0.82	11.00	73.71 (75.12)	84.87 (85.79)	21.91	2.76	24.29

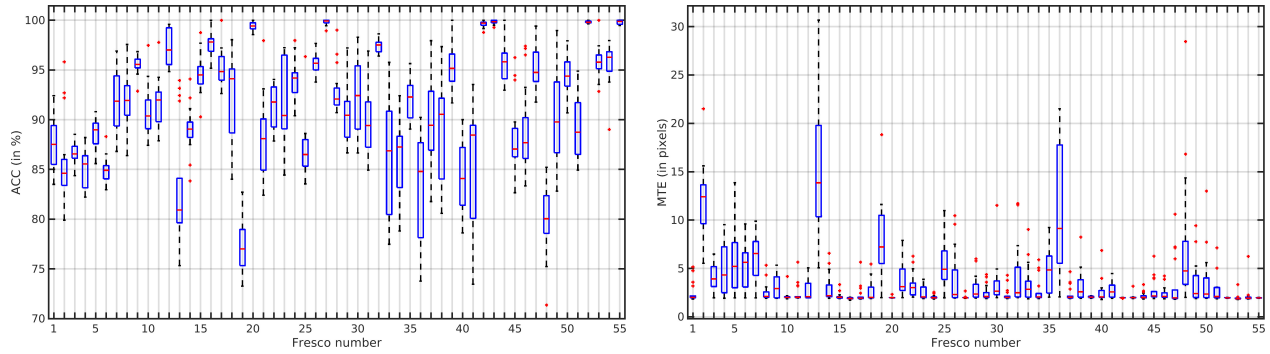

Fig. 3. ACC (%) and MTE (in pixels) box plot varying the set of fragments (18 sets per fresco) versus the considered fresco (among the 55 ones).

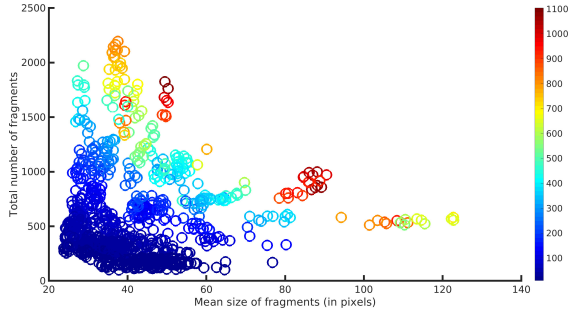

Fig. 5. Overall execution time (in seconds) of the proposed approach per instance of the entire DB1 dataset in function of both the mean size of fragments and the total number of fragments.

Table 3. Reconstruction performance of the proposed approach on the entire DB1 dataset with tolerance thresholds in translation and rotation of $\tau_t = 10$ pixels and $\tau_r = 5$ degrees, respectively.

	ACC	FM	MTE	MOE	RCR
Min.	71.38	81.36	1.70	0.05	0.00
Max.	100.00	100.00	30.66	6.88	22.58
Median	91.96	95.40	2.12	0.81	2.25
Mean	91.28	95.00	3.66	1.04	3.69
Std	6.05	3.55	3.31	0.88	4.05

4. Conclusion

In this paper, we have proposed a fast and efficient procedure for automatically reconstructing frescoes. While benefiting from previous developments in pattern matching, the experiments exhibit excellent results, are fully competitive against other approaches and provide an analysis of failure cases.

For future work, we plan to extend our approach to deal with affine geometrical transformations and then apply it to real data acquired under controlled conditions. But mostly, we plan to address the more difficult case where the fresco image is partially corrupted or even absent. Specifically, we will turn towards Markov Point Processes for the modeling of the priors between neighboring fragments, which is essential in this case.

References

[1] E. D. Demaine, M. L. Demaine, Jigsaw puzzles, edge matching, and polyomino packing: Connections and complexity, Graphs and Combinatorics 23 (1) (2007) 195–208.

[2] D. Goldberg, C. Malon, M. Bern, A global approach to automatic solution of jigsaw puzzles, in: Proc. of the eighteenth annual symposium on Computational geometry, 2002, pp. 82–87.

[3] S. Skakun, J.-C. Roger, E. F. Vermote, J. G. Masek, C. O. Justice, Automatic sub-pixel co-registration of landsat-8 operational land imager and sentinel-2a multi-spectral instrument images using phase correlation and machine learning based mapping, International Journal of Digital Earth 10 (12) (2017) 1253–1269.

[4] P. Loncomilla, J. Ruiz-del Solar, L. Martínez, Object recognition using local invariant features for robotic applications: A survey, Pattern Recognition 60 (2016) 499–514.

[5] F. Zhao, Q. Huang, W. Gao, Image matching by normalized cross-correlation, in: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Vol. 2, IEEE, 2006, pp. II–II.

[6] P. Weinzaepfel, J. Revaud, Z. Harchaoui, C. Schmid, Deepflow: Large displacement optical flow with deep matching, in: Proc. of International Conference on Computer Vision (ICCV), 2013, pp. 1385–1392.

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

[8] F. Porikli, Integral histogram: A fast way to extract histograms in Cartesian spaces, in: Proc. of Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, 2005, pp. 829–836.

[9] W. Hu, N. Xie, R. Hu, H. Ling, Q. Chen, S. Yan, S. Maybank, Bin ratio-based histogram distances and their application to image classification, IEEE Trans. on Pattern Analysis And Machine Intelligence 36 (12) (2014) 2338–2352.

[10] M. Swain, D. Ballard, Indexing via color histograms, in: Proc. of International Conference on Computer Vision (ICCV), 1990, pp. 390–393. doi:10.1109/ICCV.1990.139558.

[11] F. C. Crow, Summed-area tables for texture mapping, ACM SIGGRAPH Computer Graphics 18 (3) (1984) 207–212. doi:10.1145/964965.808600.

[12] C. G. Harris, M. Stephens, A combined corner and edge detector, in: Alvey vision conference, 1988, pp. 147–151.

[13] G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.

[14] E. Rosten, T. Drummond, Machine learning for high-speed corner detection, in: Proc. of European Conference on Computer Vision (ECCV), 2006, pp. 430–443. doi:10.1007/11744023_34.

[15] S. Leutenegger, M. Chli, R. Siegwart, BRISK: Binary Robust Invariant Scalable Keypoints, in: Proc. of International Conference on Computer Vision (ICCV), 2011, pp. 2548–2555. doi:10.1109/ICCV.2011.6126542.

[16] P. Weinzaepfel, J. Revaud, Z. Harchaoui, C. Schmid, Deepflow: Large displacement optical flow with deep matching, in: Proc. of the IEEE international conference on computer vision, 2013, pp. 1385–1392.

[17] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (6) (1981) 381–395.

[18] D. Padfield, Masked object registration in the Fourier domain, IEEE Trans. on image processing 21 (5) (2011) 2706–2718.

[19] W. Hu, N. Xie, R. Hu, H. Ling, Q. Chen, S. Yan, S. Maybank, Bin ratio-based histogram distances and their application to image classification, IEEE Trans. on Pattern Analysis And Machine Intelligence 36 (12) (2014) 2338–2352.