Spectral analysis of masked signals in the context of image inpainting

Sylvie Le Hégarat-Mascle[1] and Emanuel Aldea[1]

SATIE Laboratory, Paris-Saclay University, France
{sylvie.le-hegarat,emanuel.aldea}@u-psud.fr

Abstract. This paper proposes a computationally efficient algorithm for evaluating a sum of squared differences in the image domain in the presence of arbitrary mask configurations. Among the many potential applications of this algorithm, we consider for illustration an image inpainting task. The results show that on a diverse sample of hundreds of simulated holes in the tested images, the proposed technique is more effective than the baseline normalized cross-correlation, even when the masks are properly dealt with by the baseline.

Keywords: Spectral analysis · Fourier domain · image inpainting.

1 Introduction

In image processing, a lot of tasks have been performed in the Fourier domain, given the capability of the Discrete Fourier Transform (DFT) to provide a discriminative spectral representation of a uniformly spaced discrete signal. Computationally wise, the DFT has been appealing as well due to the Fast Fourier Transform algorithm which allowed for addressing a wide range of applications in a practical manner. Specifically, image registration is a particularly suited task, since Fourier domain representations allow for recovering the translation offset, as well as for inferring the relative rotation and scale up to some extent [17, 16, 15, 10]. The classification task may also be addressed with great success, as the spectra of images with rich content featuring real-world environments often contain diverse frequencies that may be exploited for robust discrimination [2, 21, 9].

One important but often irrelevant drawback of Fourier domain analysis is the requirement for the uniform sampling of the signal domain (in our case, the image domain). Although this does not raise particular concerns for most applications due to the widespread use of standard image sensors which sample uniformly the surrounding environment, in some cases either the support for the signal of interest might not satisfy the above constraint, or it is otherwise detrimental to process the entire domain indiscriminately. The most common situations are when 1) only some categories of objects in a scene should be considered for Fourier analysis and/or some categories of objects should be entirely ignored, 2) some context-specific masks underline the sub-part of interest in the field of view of the sensor, and 3) a part of multiple disjoint parts of the object of interest are visible. In all these circumstances, the absent area or the area containing irrelevant data would in fact contribute as well to the spectrum of the signal, and thus bias its representation and directly impact the subsequent analysis.

The community proposed various strategies for accounting for masked areas in a manner consistent with the Fourier transform, however these algorithms usually fail to preserve faithfully the spectrum of the regions of interest [20, 8, 13]. In contrast, Padfield [14] introduced an algorithm which directly and explicitly integrates the masking into the FFT algorithm steps. A normalized crossed-correlation (NCC) is computed in the Fourier domain, which is thereafter used for solving image registration. We propose to generalize this idea by extending the initial approach and demonstrate that 1) it may be used to compute other similarity criteria related to norms and 2) it may be applied to other tasks beyond registration.

2 Use of the masked Fourier transform

To measure the consistency between two areas of two different images, several criteria have been proposed, among them the NCC and the sum of squared differences (SSD). Both are computed on image subparts defined as spatial domains. To be able to compute a map giving the value of the considered consistency measure in one pass (in contrast with using a sliding window), we aim at formulating the consistency measure based on the Fast Fourier Transforms (FFT). Besides, we assume that the considered measure has to be evaluated on a domain which is not necessarily rectangular.

Let us consider f_1 and f_2 as the two considered images, with f_2 being the moving image. Let us denote by D_1 the f_1 2D domain and by $D_2(u, v)$ the f_2 translated/shifted by the 2D vector of coordinates (u, v). Then, $D_{u,v} = D_1 \cap D_2(u, v)$ denote the intersection (or overlap) between the two domains, so that the sums involved in the consistency measures are computed on $D_{u,v}$.

In [14], Padfield expressed the sums $\sum_{(x,y)\in D_{u,v}} 1$, $\sum_{(x,y)\in D_{u,v}} f_1(x, y)$ and $\sum_{(x,y)\in D_{u,v}} f_1(x, y)^2$ (and the corresponding expressions for f_2) after introducing the masks m_1 and m_2 corresponding to the indicator functions of the D_1 and D_2 domains, and their Fourier transforms $M_i = \mathcal{F}(m_i), i \in \{1, 2\}$:

$$|D_{u,v}| = \sum_{(x,y)\in D_{u,v}} 1 = \mathcal{F}^{-1}(M_1 \cdot M_2^*)(u, v), \tag{1}$$

$$\overline{f_1} = \frac{1}{|D_{u,v}|} \sum_{(x,y)\in D_{u,v}} f_1(x, y) = \frac{\mathcal{F}^{-1}(F_1 \cdot M_2^*)(u, v)}{\mathcal{F}^{-1}(M_1 \cdot M_2^*)(u, v)}, \tag{2}$$

$$\sum_{(x,y)\in D_{u,v}} f_1(x, y)^2 = \mathcal{F}^{-1}(\mathcal{F}(f_1 \cdot f_1) \cdot M_2^*)(u, v), \tag{3}$$

with $F_i = \mathcal{F}(f_i), i \in \{1, 2\}$, and all the multiplications in these equations being elementwise (Hadamard product), and M_2^* denotes the complex conjuguate of the Fourier transform of M_2 (that is also the Fourier transform of the transposed of m_2).

2.1 Average of Squared Differences with masks

In this work, we derive the formulation of the Average of Squared Differences either centered or not centered, using the Fourier transform, i.e. without resorting to a sliding window. Indeed, conversely to the NCC measure, the SSD has been very popular due to its convenient summation properties, which are valuable for considering multichannel images such as color ones. Note that in the standard case, this measure is not normalized since the support has a set size and the normalization would be useless. In our case however, we will explicitly focus on the normalized (i.e., averaged with respect to the number of valid pixels) version since the masks will modify the support size.

Using the previous notations, we express the sum of the squared differences as follows:

$$\begin{aligned} SSD(u,v) &= \sum_{(x,y)\in D_{u,v}} \left(\left(f_1(x,y)-\overline{f_1}\right)-\left(f_2(x-u,y-v)-\overline{f_{2,u,v}}\right)\right)^2 &(4)\\ &= \sum_{(x,y)\in D_{u,v}} f_1(x,y)^2 + \sum_{(x,y)\in D_{u,v}} f_2(x-u,y-v)^2 &(5)\\ &\quad -2\sum_{(x,y)\in D_{u,v}} f_1(x,y)f_2(x-u,y-v) - |D_{u,v}|\left(\overline{f_1}-\overline{f_{2,u,v}}\right)^2 \\ &= \Big[\mathcal{F}^{-1}(\mathcal{F}(f_1\cdot f_1)\cdot M_2^*)) + \mathcal{F}^{-1}(M_1\cdot\mathcal{F}(f_2^{'}\cdot f_2^{'}))- &(6)\\ &\quad 2\mathcal{F}^{-1}(F_1\cdot F_2^*) - \left(\mathcal{F}^{-1}(F_1\cdot M_2^*) - \mathcal{F}^{-1}(F_2\cdot M_2^*)\right)^2\Big](u,v), \end{aligned}$$

with $f_2^{'}$ being the f_2 image flipped (central symmetry).

The previous equation stands for the SSD with centered differences. In some cases, it might be advisable to consider the uncentered differences, and this version of the SSD is even simpler to express:

$$SSD(f_1,f_2) = \mathcal{F}^{-1}(\mathcal{F}(f_1\cdot f_1)\cdot M_2^*)) + \mathcal{F}^{-1}(M_1\cdot\mathcal{F}(f_2^{'}\cdot f_2^{'})) - 2\mathcal{F}^{-1}(F_1\cdot F_2^*). \quad (7)$$

Note that the computation of the uncentered SSD involves two less inverse Fourier transforms than that of the centered SSD. Then, the map of the average of squared differences (ASD) is provided by dividing pixel per pixel SSD map values (Eq. (6) or (7)) by $\mathcal{F}^{-1}(M_1\cdot M_2^*)$ map values:

$$ASD(f_1,f_2) = \frac{SSD(f_1,f_2)}{\mathcal{F}^{-1}(M_1\cdot M_2^*)}. \quad (8)$$

2.2 Normalized Cross Correlation with masks

In [14], following this reformulation of the intermediate terms which accounts for the impact of the masks m_1 and m_2 on the Fourier transform, the final evaluation of the NCC becomes:

$$NCC(f_1, f_2) = \frac{\mathcal{F}^{-1}(F_1 \cdot F_2^*) - \frac{\mathcal{F}^{-1}(F_1 \cdot M_2^*) \cdot \mathcal{F}^{-1}(M_1 \cdot F_2^*)}{\mathcal{F}^{-1}(M_1 \cdot M_2^*)}}{\sqrt{\mathcal{F}^{-1}(\mathcal{F}(f_1 \cdot f_1) \cdot M_2^*) - \frac{(\mathcal{F}^{-1}(F_1 \cdot M_2^*))^2}{\mathcal{F}^{-1}(M_1 \cdot M_2^*)}} \sqrt{\mathcal{F}^{-1}(M_1 \cdot \mathcal{F}(f_2' \cdot f_2')) - \frac{(\mathcal{F}^{-1}(M_1 \cdot F_2^*))^2}{\mathcal{F}^{-1}(M_1 \cdot M_2^*)}}} \tag{9}$$

3 Inpainting with the masked Fourier transform

Inpainting is a radical form of image restoration in which pixels inside a missing region of the image are filled with information provided by the surrounding areas, and potentially by the entire domain. The concept of self-similarity is central to this task, since the main underlying assumption is that repeating patterns from other parts of the image will be relevant for the missing area, and will thus be imported there as well. In the spatial domain, the inpainting task has been traditionally addressed, as an ill-posed problem, by total variation regularization, by dictionary based approaches [5, 6], by diffusion with partial differential equations [11, 22] or by some hybrid strategy [3, 1]. Compared to these commonly encountered approaches, methods which use explicitly the Fourier transform of the deteriorated input image are less common. Some works focus on some specific cases of inpainting which are particularly suited for spectral analysis (e.g. removal of text overlay [18]). In [23], the FFT is used as an accelerating step for patch matching in exemplar based image inpainting, while in [12] the inpainting task is performed in the Fourier space of the image representation (i.e. some coefficients are assumed to be missing). In [19], the image is decomposed in decomposed in texture and non-texture components, then texture inpainting and denoising is performed in a subsequent step. In [7], the authors employ an alternative to the FFT, namely nonharmonic analysis, in order to minimize the impact of the side lobes appearing on truncated data. None of these works try to cope explicitly with the impact of the mask on the extracted frequencies, nor they propose a fully spectral, computationally efficient approach.

In this work, we propose to take advantage of the fact that we are able to compute a consistency criterion only considering the available pixels (i.e., fully disregarding the missing ones) thanks to the use of Fourier transform based expressions introduced in Section 2. Based on these latter, we are then able to propose an efficient strategy for solving the inpainting problem in which the mask represents the area to be filled. The proposed algorithm searches, for each area including missing pixels, another area which presents similar structures and colors than the ones near the missing pixels. This search can be performed globally or locally depending on the assumptions. We then expect such an approach to be all the more performing that the image presents repetitive structures or textures.

The global outline is given by Algorithm 1. In the considered algorithm, as well as in the conducted experiments in Section 4, we consider RGB images. However, it is straightforward to apply the proposed approach in another color space, or even to multispectral (or hyperspectral) images. The size l of the search

Algorithm 1 Inpainting by research of similar areas; input: RGB image I, search area side length l, boolean b_u indicating if ASD is centered; output: RGB image $\tilde{I}$. FFT stands for the Fast Fourier Transform and IFFT for the Inverse Fast Fourier Transform.

$L \leftarrow$ list of I areas with missing pixels
Initialize $\tilde{I}$ to I
for each element A_j of L **do**
    $(x, y) \leftarrow A_j$ center coordinates
    $d \leftarrow A_j$ min square bounding box side length
    $f_1 \leftarrow$ rectangular tile centered on (x, y) and having side length equal to l
    Cut a box of side length d at the center of f_1
    $n \leftarrow \lceil log_2(d) \rceil$
    $f_2 \leftarrow$ rectangular tile centered on (x, y) and having side length equal to 2^n
    $m_1 \leftarrow$ binary mask of f_1 valid pixels
    $M_1 \leftarrow \text{FFT}(m_1)$
    $m_2 \leftarrow$ binary mask of f_2 valid pixels
    $M_2 \leftarrow \text{FFT}(m_2)$
    $\check{M}_{12} \leftarrow \text{IFFT}(M_1 \cdot M_2^*)$
    Initialize map S to 0 in every pixel
    for each channel k of f_1 **do**
        for $i \in 1, 2$ **do**
            $f_{i,k} \leftarrow$ channel k of f_i
            $F_i \leftarrow \text{FFT}(f_{i,k})$
            $f_{i,k}^2 \leftarrow f_{i,k} \cdot f_{i,k}$
            $G_i \leftarrow \text{FFT}(f_{i,k}^2)$
        end for
        $\check{F}_{12} \leftarrow \text{IFFT}(F_1 \cdot F_2^*)$
        $\check{H}_{12} \leftarrow \text{IFFT}(G_1 \cdot M_2^*)$
        $\check{H}_{21} \leftarrow \text{IFFT}(M_1 \cdot G_2^*)$
        if b_u **then**
            $\check{J}_{12} \leftarrow \text{IFFT}(F_1 \cdot M_2^*)$
            $\check{J}_{21} \leftarrow \text{IFFT}(M_1 \cdot F_2^*)$
        end if
        $S \leftarrow S + \check{H}_{12} + \check{H}_{21} - 2\check{F}_{12} + b_u \left(\frac{\check{J}_{12} - \check{J}_{21}}{\check{M}_{12}} \right)$
    end for
    $S \leftarrow \frac{S}{\check{M}_{12}}$
    $(\hat{u}, \hat{v}) \leftarrow \arg\min_{(u,v)} S(u, v)$
    Fill the missing values in $\tilde{I}$ by pasting the values of I around $(x + \hat{u}, y + \hat{v})$
end for

area is an input parameter left to the user. Note that using a global search, the Fourier transforms associated to the search area (the whole image in this case) can be computed once at the beginning of the algorithm. However, in this case, we lose the support of a locality constraint. Note also that, to benefit from fast Fourier transform algorithms, l has to be a power of 2. In the extraction of f_1 and f_2 the original image can be padded with 0 (masked pixels) if necessary. Finally, since the patch that is researched in the image (around the area to fill) is extracted from the image itself, it is necessary to mask the pixels of the patch (to avoid to reselect the original patch location), which is simply done by "cutting" a box of patch size in original image (and filling it with black pixels).

4 Experimental results

We performed experiments related to the inpainting task on images selected from the publicly available DAFNE challenge dataset [4]. As previously stated, the proposed algorithm is generally applicable to grayscale or color images, but the samples which are present in DAFNE offer a good diversity in terms of appearance and style of the content, with a good balance between repeating patterns and more singular structures. The hole creation process is performed by randomly and uniformly removing disk patches of content from the selected images.
Evaluation metrics Two widely used metrics are considered for the numerical evaluation. First, the Root Mean Squared Error (RMSE) is considered, then we also compute a metric which is more specific for benchmarking image reconstruction tasks, namely the peak signal-to-noise ration (PSNR). Although PSNR is partly related to the RMSE, it highlights better the method performance independently of the numerical range of the studied signal values.
Algorithm variants Based on the general idea for inpainting with the masked Fourier transform, we consider three variants depending on the similarity measure which is employed in searching for the most visually close data: (1) the normalized cross correlation (NCC, cf. Eq. (9)) computed on the intensity image, (2) a weighted sum of the three intensity-based criteria that may be potentially used: NCC, centered ASD (cf. Eq. (6) and Eq. (8)) and uncentered ASD (cf. Eq. (7) and Eq. (8)) called uASD, (3) the uASD computed on the three color channels (benefiting from good summation property), called 3D uASD.

Tables 1 and 2 show the obtained results in terms of evaluation metric statistics computed on 100 holes per fresco. We clearly see that the NCC is not sufficient to find a good image patch to fill a given hole, mainly since it is "only' based on intensity relative variations. Adding the two criteria based on squared differences significantly improves the results (between 2 and 8 dB depending on the considered fresco) while significantly reducing the standard deviation. Finally, the benefit of color information is also visible, varying however with the content of the fresco: the Lanzani fresco presents rather homogeneous colors (cf. Figure 1, second line) whereas the Dellafrancesca fresco has a variety of colors (cf. Figure 1, third line) allowing for a more significant improvement of performance.

Table 1. RMSE statistics obtained on three frescoes for three algorithms based on different consistency measures; statistics (mean, median and standard deviation after ± symbol) are derived from 100 simulated holes. Best results are highlighted in green.

Algorithm	**Tiepolo**		**Lanzani**		**DellaFrancesca**	
	mean	median	mean	median	mean	median
NCC	35.64 ±27.67	31.76	36.32 ±17.15	33.92	35.85 ±21.73	30.01
NCC+ASD +uASD	25.86 ±19.20	20.56	29.09 ±11.52	28.14	29.74 ±14.41	27.28
3D uASD	24.74 ±19.02	20.13	28.69 ±12.03	26.99	27.40 ±13.81	25.28

Table 2. PSNR statistics obtained on three frescoes for three algorithms based on different consistency measures; statistics (mean, median and standard deviation after ± symbol) are derived from 100 simulated holes. Best results are highlighted in green.

Algorithm	**Tiepolo**		**Lanzani**		**DellaFrancesca**	
	mean	median	mean	median	mean	median
NCC	47.61 ±21.87	41.66	41.58 ±11.92	40.35	42.76 ±13.14	42.79
NCC+ASD +uASD	52.85 ±19.79	50.36	45.41 ±10.66	44.08	45.79 ±12.16	44.70
3D ASD	54.0 ±19.92	50.78	45.88 ±11.01	44.91	47.66 ±12.56	46.23

Finally, let us have a qualitative look at the obtained results. Figure 1 shows the whole frescoes with simulated holes and the inpainting results from 3D uASD, while Figure 2 shows some selected subareas. Indeed, Figure 1 allows us to check quickly that holes have been filled with consistent values since at first glance it is difficult to see the location of the simulated holes (without the help of the left image) whereas Figure 2 allows us to evaluate the visual quality of the local reconstruction (fresco subareas of size 128×128 pixels), as well as to point out some remaining imperfections. Specifically, in Figure 2, for each line considering details from a different fresco, left side shows a subarea that is rather well reconstructed (it is difficult to guess where the holes were even focusing on this subarea), while the right side shows a subarea with at least one partially flawed reconstruction: one hole filled with an incorrect shade of white (first line), a discontinuity in the leg of the soldier (second line) and appearing inconsistencies in the drapery of tunics (third line). However, all these imperfections are explainable and we hope to reduce them by searching in future works a patch candidate for the filling not only in translation but also in rotation.

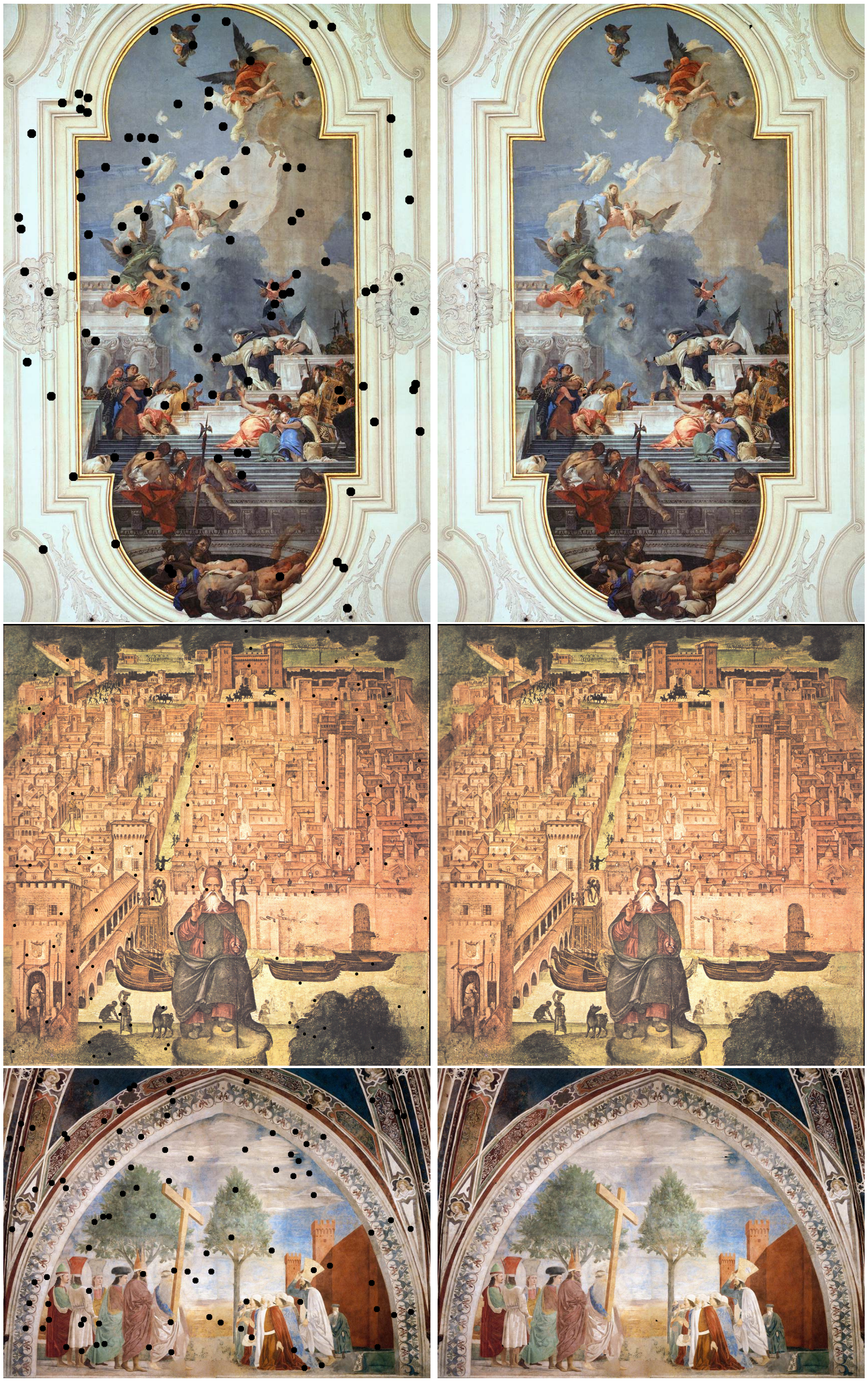

Fig. 1. Reconstructed frescoes: original image with simulated holes (left) and inpainting result (right); first line: Tiepolo fresco ("The Institution of the Rosary"), second line: Lanzani fresco ("Sant Antonio protegge Pavia"), and third line: Della Francesca fresco ("Exaltation of the Cross").

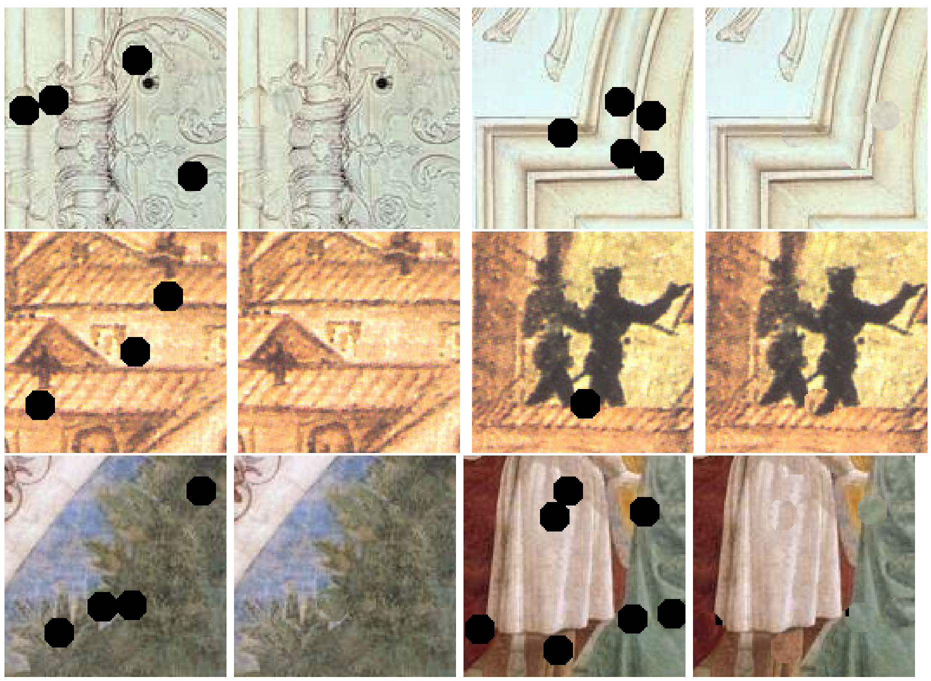

Fig. 2. Details of the reconstructed frescoes: simulated holes (left) and inpainting result (right); first line: Tiepolo fresco, second line: Lanzani fresco, and third line: Della Francesca fresco.

5 Conclusion

In this work, we extended the strategy initially proposed by [14] beyond NCC, to more consistency measures between masked image subparts. We have also addressed with success a novel task, namely inpainting, which has not been considered until now with these techniques, despite its suitability with the method assumptions. For future work, we intend to fully exploit the other properties of the Fourier transform (e.g. following Fourier–Mellin approach) which would allow us to perform more complex rotation and scaling invariant queries, and to characterize more in detail the computational advantage of the proposed algorithm.

References

1. Bugeau, A., Bertalmío, M., Caselles, V., Sapiro, G.: A comprehensive framework for image inpainting. IEEE transactions on image processing **19**(10), 2634–2645 (2010)
2. Chan, C.h., Pang, G.K.: Fabric defect detection by fourier analysis. IEEE transactions on Industry Applications **36**(5), 1267–1276 (2000)
3. Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on image processing **13**(9), 1200–1212 (2004)
4. Dondi, P., Lombardi, L., Setti, A.: Dafne: A dataset of fresco fragments for digital anastlylosis. Pattern Recognition Letters **138**, 631–637 (2020)
5. Elad, M., Starck, J.L., Querre, P., Donoho, D.L.: Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). Applied and computational harmonic analysis **19**(3), 340–358 (2005)
6. Fadili, M.J., Starck, J.L., Murtagh, F.: Inpainting and zooming using sparse representations. The Computer Journal **52**(1), 64–79 (2009)
7. Hasegawa, M., Kako, T., Hirobayashi, S., Misawa, T., Yoshizawa, T., Inazumi, Y.: Image inpainting on the basis of spectral structure from 2-d nonharmonic analysis. IEEE transactions on image processing **22**(8), 3008–3017 (2013)
8. Kaneko, S., Satoh, Y., Igarashi, S.: Using selective correlation coefficient for robust image registration. Pattern Recognition **36**(5), 1165–1173 (2003)
9. Kliangsuwan, T., Heednacram, A.: Fft features and hierarchical classification algorithms for cloud images. Engineering Applications of Artificial Intelligence **76**, 40–54 (2018)
10. Konstantinidis, D., Stathaki, T., Argyriou, V.: Phase amplified correlation for improved sub-pixel motion estimation. IEEE Transactions on Image Processing **28**(6), 3089–3101 (2019)
11. Masnou, S.: Disocclusion: a variational approach using level lines. IEEE Transactions on Image Processing **11**(2), 68–76 (2002)
12. Mousavi, P., Tavakoli, A.: A new algorithm for image inpainting in fourier transform domain. Computational and Applied Mathematics **38**(1), 1–9 (2019)
13. Orchard, J.: Efficient least squares multimodal registration with a globally exhaustive alignment search. IEEE Transactions on Image Processing **16**(10), 2526–2534 (2007)
14. Padfield, D.: Masked object registration in the fourier domain. IEEE Transactions on image processing **21**(5), 2706–2718 (2011)

15. Pan, W., Qin, K., Chen, Y.: An adaptable-multilayer fractional fourier transform approach for image registration. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(3), 400–414 (2008)
16. Reddy, B.S., Chatterji, B.N.: An fft-based technique for translation, rotation, and scale-invariant image registration. IEEE transactions on image processing **5**(8), 1266–1271 (1996)
17. Sheng, Y., Arsenault, H.H.: Experiments on pattern recognition using invariant fourier–mellin descriptors. JOSA A **3**(6), 771–776 (1986)
18. Sridevi, G., Srinivas Kumar, S.: Image inpainting based on fractional-order nonlinear diffusion for image reconstruction. Circuits, Systems, and Signal Processing **38**(8), 3802–3817 (2019)
19. Thai, D.H., Gottschlich, C.: Simultaneous inpainting and denoising by directional global three-part decomposition: connecting variational and fourier domain-based image processing. Royal Society Open Science **5**(7), 171176 (2018)
20. Thevenaz, P., Ruttimann, U.E., Unser, M.: A pyramid approach to subpixel registration based on intensity. IEEE transactions on image processing **7**(1), 27–41 (1998)
21. Tsai, D.M., Wu, S.C., Li, W.C.: Defect detection of solar cells in electroluminescence images using fourier image reconstruction. Solar Energy Materials and Solar Cells **99**, 250–262 (2012)
22. Tschumperlé, D., Deriche, R.: Vector-valued image regularization with pdes: A common framework for different applications. IEEE transactions on pattern analysis and machine intelligence **27**(4), 506–517 (2005)
23. Wang, H., Jiang, L., Liang, R., Li, X.X.: Exemplar-based image inpainting using structure consistent patch matching. Neurocomputing **269**, 90–96 (2017)