Context-aware image inpainting with application to virtual restoration of old paintings

Tijana Ružić and Aleksandra Pižurica
Department of Telecommunications and Information Processing TELIN-IPI-iMINDS
Ghent University, Belgium
Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium
Email: truzic@telin.ugent.be

***Abstract*—In this paper, we explore how to use spatial context for image inpainting and we test it in two applications: photo-editing and crack removal in digitized old paintings. Context is determined based on the texture and color features. We use these contextual features to guide the search for image patches that can fill in the missing/damaged regions in a visually plausible way. A priori knowledge about spatial consistencies (similarities) among neighbouring image patches is encoded via a Markov Random Field (MRF) model. With this prior, the process of image inpainting is formulated as an optimization problem. We define an efficient inference engine as a valuable alternative to the so-called belief propagation methods. As a case study, we focus on crack removal from the *Adoration of the Mystic Lamb* (brothers Van Eyck, 1432), demonstrating potentials for virtual restoration of old paintings.**

I. Introduction

Image inpainting, or image completion, is an image processing task of filling in the missing region in an image in a visually plausible way. Applications include image restoration (e.g. scratch or text removal), image coding and transmission (e.g. recovery of missing blocks), photo-editing (e.g. object removal), virtual restoration of digitized paintings (crack removal), etc. In this paper, we will consider two applications: photo-editing and crack removal.

Specifically, for crack removal we will use the *Adoration of the Mystic Lamb*, also known as the *Ghent Altarpiece* as a case study (see Fig. 1). The polyptych consisting of 12 panels, dated by inscription 1432, was painted by Jan and Hubert van Eyck and is considered as one of the greatest masterpieces. It is still located in the Saint Bavo Cathedral in Ghent, its original destination. As in most 15^{th} century Flemish paintings on Baltic oak, fluctuations in relative humidity, acting over time on the wooden support, caused age cracks, which is one of the most common deteriorations in old paintings. These cracks form an undesired pattern that is, however, inherent to our appreciation of these paintings as old and valuable. Yet, for specialists in visual perception for example, it is of interest how our perception of the painting is affected when observing it before the ageing process. Moreover, the crack patterns not only make art historical analysis more difficult, but also in the example of the inscribed text (see Fig. 1c), the palaeographical deciphering. Therefore, an important task for the restoration of digitized paintings is the automatic detection and removal of cracks, which is a very difficult and delicate problem.

In literature, two categories of image inpainting approaches can be distinguished: diffusion-based and patch-based. Diffusion-based methods [3], [4] fill in the missing region by smoothly propagating information from the boundary to the interior of the missing region by diffusion process. Important is to propagate linear structures, e.g. object lines and boundaries, by continuing the lines that arrive at the border of the missing region inside the hole. This propagation includes solving partial differential equations (PDE). These methods only use information in the immediate surrounding of the missing pixel to infer its value, thus they can also be referred to as pixel-based. Pixel-based approaches yield good results when inpainting long thin regions, such as cracks, so they were often used for crack inpainting. Crack inpainting methods considered in literature so far include order statistics filtering [5], [6], controlled anisotropic diffusion [5] and interpolation [7]. However, diffusion-based methods in general experience difficulties in replicating texture and introduce blur when filling in larger holes.

Patch-based methods [8] fill in the missing region patch-by-patch by searching for well-matching replacement patches in the available (undamaged) part of the image and copying them to corresponding locations. Additionally, emphasis is put on structure propagation by, for example, defining the filling order [8]–[10]. Compared to the diffusion-based methods, patch-based methods produce better results, especially when inpainting large missing regions.

Patch-based methods can be categorized into "greedy" [1], [8], [11], [12], non-local [10], [13] and global [9], [14], [15]. The "greedy" ones choose only one best match at a time, which may be quite limiting and cause visually inconsistent results, while non-local methods choose multiple candidate patches and the final patch represents their weighted average [10] or their sparse combination [13]. Finally, global methods define inpainting as a global optimization problem. This, in addition to the choice of multiple candidates (called *labels*), allows for one label to be chosen eventually for each position so that the whole set of patches (at all positions) minimizes a global optimization function. An efficient optimization approach for this problem, using belief propagation, was proposed in [9]. Major problem with this type of approaches is complexity (stemming from a huge number of candidate patches that can cover certain position). Various solutions on how to reduce the search for candidate patches in a meaningful way were proposed in [10], [11], [14], [16].

In this paper, we describe a novel global Markov Random Field (MRF) based inpainting method where contextual features are used both to improve the inpainting result and to

(a) (b) (c)

Fig. 1. Some details from the *Adoration of the Mystic Lamb*: (a) the piece of jewellery in the God the Father panel, (b) Adam's face in the Adam panel, (c) the book in the Annunciation to Mary panel.

accelerate the search for candidate patches. The main novelty is context-aware label selection, which limits the search for labels to the areas of interest based on contextual information. We employ Gabor-based texture descriptors similar to those in [17], [18] and extend them with color information. While the related context descriptors were used in other domains like scene recognition [17] and scene completion using millions of photographs [18], we do not know of any works where such descriptors were used for patch-based image inpainting, except in our previous work [12], [15]. We demonstrate that the inpainting process can largely benefit from such a context aware label search and selection, both in terms of speed and quality. We also discuss a novel optimization approach, which builds upon our recent inference method [19] to make it suitable for global inpainting problem with large number of labels.

Preliminary results of this work were presented in [15], evaluated only on photo-editing with artificial masks. In this paper, we apply the method also to virtual restoration of old paintings, i.e. for crack removal. We show that the proposed patch-based method outperforms pixel-based one, and that the use of context can be beneficial also in this application. Additionally, we briefly discuss alternative use of context via segmentation, which we applied for crack removal in [1], [2].

The paper is organized as follows. The proposed context-aware global inpainting method is explained in Sec. II, while the experiments and results on natural images and artwork are shown in Sec. III. The paper is concluded with Sec. IV.

II. Context-Aware Global Image Inpainting

A. Inpainting as a global optimization problem

Consider the input image I, with Ω the region to be filled, called the *target* region, and Φ the known part of the image, called the *source* region. We define a Markov Random Field (MRF) $G = (\nu, \varepsilon)$ over the target region Ω as a lattice of overlapping $w \times w$ patches which intersect Ω. These patches then represent MRF nodes $p \in \nu$ whose labels are all possible patches $x_p \in \Lambda$ taken from Φ, while edges ε make a four neighbourhood system around each central node. Furthermore, data cost $V_p(x_p)$ of assigning a label x_p to node p is defined as the sum of squared differences (SSD) between the known pixels at the node and the corresponding label pixels (note that it is zero if the node is completely inside Ω). Finally, pairwise potential $V_{pq}(x_p, x_q)$, where p and q are neighbouring nodes, is similarly defined as SSD between labels x_p and x_q in their region of overlap. The global inpainting problem can now be formulated as minimizing the energy

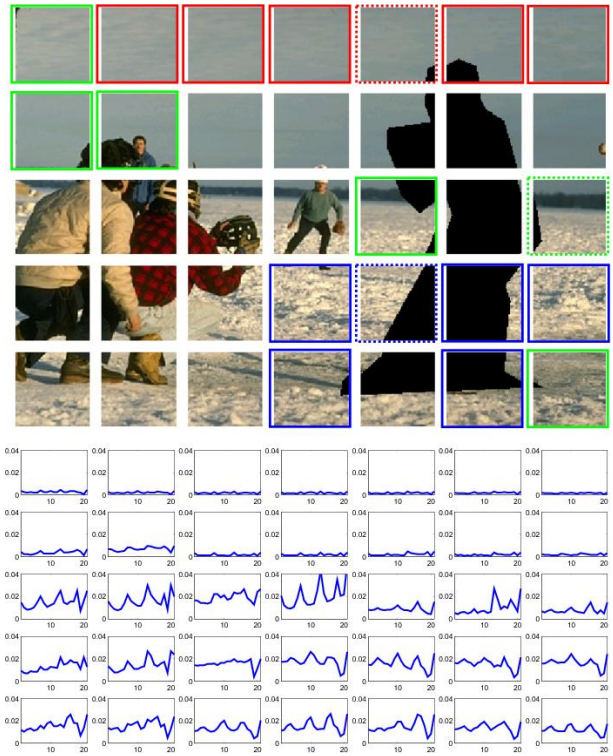

Fig. 2. Top: Division of the image into 5×7 non-overlapping blocks. Block matches of blocks in dashed squares in the top image are shown in squares of matching color. Bottom: Corresponding contextual descriptors plotted over 21 components and with values ranging between 0 and 0.04 (see text for details).

$$E(\mathbf{x}) = \sum_{p \in \nu} V_p(x_p) + \sum_{(p,q) \in \varepsilon} V_{pq}(x_p, x_q). \qquad (1)$$

This type of optimization problems can be solved using loopy belief propagation (LBP) algorithms [20]. The inpainting method of [9] introduced an improved version of belief propagation called priority BP (p-BP), to deal more efficiently with problems where each node has a huge number of labels, as is the case here. A problem with [9] is that at the beginning of the algorithm it involves exhaustive search for labels, which makes it very slow, especially for bigger images. We introduce contextual information to perform *constrained* search making the algorithm thereby much faster. Therefore, our proposed method consists of two parts: (1) context-aware label selection and (2) efficient energy optimization.

B. Context-Aware Label Selection

Our idea is to guide the label selection by contextual information, where context is characterized by texture and color descriptors within a fixed block around each node. Texture descriptors contain a set of low-level image features that describe the texture in an image or image area. We will use similar texture descriptors as [17], [18], which are obtained by

Algorithm 1: Algorithm for context-aware label selection

for $p \leftarrow 1$ **to** P **do** // P is the number of nodes
  find the block B_i to which p belongs
  compute block reliability
  $\rho_i = \begin{cases} 1 & \text{if } \#\{B_i \cap \Phi\} > \frac{\#B_i}{2} \\ 0 & \text{otherwise} \end{cases}$
  if $\rho_i = 1$ **then**
    $e(j) = \sum_{n=1}^{N_f+C}(g_i(n) - g_j(n))^2, \forall j = \{1, \ldots, MN\}$
    choose $K = MN/r$ blocks $\hat{B}_i^j$ whose g_j yield K smallest $e(j)$
    define new source region $\Phi_p = \cup\{\hat{B}_i^1, \ldots \hat{B}_i^K\}$
  else
    $\Phi_p = \emptyset$
    foreach *neighbouring block* B_n **do**
      repeat steps 4-6 and $\Phi_p = \cup\{\Phi_p, \hat{B}_n^1, \ldots \hat{B}_n^K\}$
    end
  end
end

filtering the image with a bank of multi-scale oriented filters and then averaging the outputs within square non-overlapping blocks [17]. Such a representation is called a *gist* and it gives coarse description of textures in the image and their spatial organization.

We divide the image into $M \times N$ square non-overlapping blocks (see Fig. 2) and for each block B_i we compute its texture descriptor g_i as:

$$g_i(n) = \frac{1}{\#\{B_i \cap \Phi\}} \sum_{y \in B_i \cap \Phi} |I(y) \otimes h_n(y)|^2, \forall n \in \{1, \ldots, N_f\}. \quad (2)$$

$\otimes$ is a convolution operator, $\#\{B_i \cap \Phi\}$ represents the number of *known* pixels y in a corresponding block B_i and N_f is the number of filters in a chosen filter bank. We use Gabor filters of six orientations and across three scales, total of 18 filters. Then g_i is a 18-dimensional vector whose components are ordered by orientation per each scale, from high to low scales, i.e. high to low spatial frequencies.

In addition to texture, it is also beneficial to include color as a feature for contextual description. Therefore, we add three more components into the feature vector $\mathbf{g_i}$ which represent the average color within the block per each HSV color channel:

$$g_i(N_f + n) = \frac{1}{\#\{B_i \cap \Phi\}} \sum_{y \in B_i \cap \Phi} I_n(y), \forall n \in \{1, \ldots, C\}, \quad (3)$$

where $C = 3$ is the number of color channels. The averaged color values per channel are typically higher than averaged filter responses. Hence, we normalize the color components by the factor f, $g_i(N_f + n) = g_i(N_f + n)/f$, which is the ratio between maximum value of three color components and maximum value of the averaged filter responses on first N_f components:

$$f = \frac{\max_{k \in \{N_f+1, N_f+2, N_f+C\}} g_i(k)}{\max_{l \in \{1, \ldots, N_f\}} g_i(l)}. \quad (4)$$

The resulting $(N_f + C)$-dimensional feature vector $\mathbf{g_i}$ ($N_f + C = 21$) shows dominant orientations and scales within the block B_i and the average color of that block. Fig. 2 illustrates these feature vectors corresponding to different blocks of an image. We can see that texture features (the first N_f components) are small for nearly flat blocks (most of the blocks in the first two rows). For the blocks with dominant edges the peaks appear at positions corresponding to a particular orientation and tend to increase when the scale increases. Textured blocks containing snow for example, have smaller descriptor values and smaller peaks at multiple orientations.

Now we can use the feature vectors defined above to find blocks with similar content and we will limit the label set only to those blocks. The idea is to constrain the source region for node $p \in B_i$ to $\Phi_p \subset \Phi$, as shown in pseudo code in Algorithm 1. The block B_i itself is always included in this limited source region ($e(i) = 0$). For examples of block matches see marked blocks at the top of Fig. 2. Note that we also introduce binary variable ρ_i that represents the reliability of the block because some of the blocks that intersect the target region can have too little or even no known pixels based on which context information can be obtained, in which case we use the neighbouring information. Note that the presented context-aware approach can be used to guide and limit candidate patch selection in any inpainting process, i.e. it is not restricted only to global inpainting.

C. Efficient Energy Minimization

Although we use contextual descriptors to limit the labels to areas of interest and, therefore, substantially reduce their number ($\Lambda_p \in \Phi_p$, $|\Lambda_p| \approx |\Lambda|/r$), we are still dealing with thousands of labels per MRF node, which is too complex for subsequent optimization. Here we introduce an efficient inference method, by extending our recent approach from [19].

We propose to first prune the labels of each node by visiting them in the order of priority, where both pruning and priority are determined based on only one term for each label of a node $D_p(x_p)$ that we call *dissimilarity*. This allows us to define a computationally tractable MRF and perform simple and fast inference method to obtain the final inpainting result. Label pruning and priority scheduling are also present in p-BP. However, our approach is simpler, faster, more memory efficient and allows the application on bigger images.

A pseudo-code of this algorithm is given under Algorithm 2. We can see that dissimilarity $D_p(x_p)$ is initially computed based on the agreement of node's labels with the known part of the patch $V_p(x_p)$ and subsequently updated with the neighbouring influence expressed with pairwise potential $V_{pq}(x_p, x_q)$, where both $V_p(x_p)$ and $V_{pq}(x_p, x_q)$ are defined in Sec. II-A. Based on this similarity measure, both priorities are computed (as defined in step 4 of Algorithm 2) and label pruning is performed: L labels with lowest dissimilarity are kept as node's labels, while others are discarded.

Fig. 3. Inpainting results. From left to right: image with missing region in black, results of [10], results of [9], result of the complete proposed method.

Algorithm 2: Algorithm for efficient energy minimization

initialization:
for $p \leftarrow 1$ **to** P **do**
    compute $D_p(x_p) = V_p(x_p), \forall x_p \in \Lambda_p$
    compute priority $Pr_p = 1/\#\{x_p | D_p(x_p) < T_{sim}\}$
end
label pruning:
for $t \leftarrow 1$ **to** P **do**
    p = unvisited node of highest priority
    apply label pruning: $x_p \in \{x_p^1, \dots x_p^L\}, L \ll |\Lambda_p|$
    for *any unvisited neighbour q of node p* **do**
        $D_q(x_q) = D_q(x_q) + \min_{x_p} V_{pq}(x_p, x_q), \forall x_q \in \Lambda_q$
        update priority of node q, Pr_q
    end
end
inference method: $\hat{\mathbf{x}} = \arg\min E(\mathbf{x})$

At this point, we have a completely defined 4-connected MRF, where each node p has a set of L possible labels, where $L \ll |\Lambda_p|$, and where the potential functions are defined like in Sec. II-A. Now we employ our recent inference method called neighbourhood-consensus message passing (NCMP) [19] to determine one label per node, where the set of labels $\hat{\mathbf{x}}$ over all nodes minimizes the energy in 1. This method is simpler and faster than belief propagation and was proved to give good results in other patch-based MRF models. Finally, chosen patches are stitched together in the region of overlap using minimum error boundary cut [21], as suggested in [9].

III. Experiments and Results

A. Photo-editing of natural images

We tested our method on a number of different natural images for photo-editing application. The parameters of the algorithm are the following: number of labels per node $L = 10$, number of iterations of NCMP $T = 10$, number of chosen blocks $K = MN/r$, where $r = 6$, and $T_{sim} = SSD_0/2$, where SSD_0 is a predefined median value of SSDs between $w \times w$ patches. Patch size w and number of blocks $M \times N$ were varied and the optimal ones were chosen. Results on some of the images are shown on Fig. 3. For images from top to bottom, we used the same patch sizes for all global methods, $w = 15$ and $w = 13$, respectively, while block divisions for the proposed method are 5×7 and 3×4, respectively. We can see that the proposed method gives the best results, i.e. more accurate and more visually pleasing, compared to the global method from [9] and state-of-the-art method from [10]. Moreover, our method is almost 5 times faster than [9]. For example, for the image in the top row of Fig. 3, the reference method from [9] takes 1152.8s, while the proposed method takes 284.3s (with our MatLab implementation with $w = 15$). For detailed comparison, see [15]).

B. Virtual Restoration of Old Paintings

Virtual restoration of digitized old paintings involves the detection and removal of cracks. Cracks, once detected, can be treated as missing regions that need to be filled in. We focus here on the difficult problem of crack inpainting in the *Ghent Altarpiece*, from which three small parts are shown in Fig. 1. The cracks in this painting are particular in a number of ways. Their width ranges from very narrow and barely visible to larger areas of missing paint. Furthermore, depending on the painting's content, they appear as dark lines on a bright background or vice versa. Also, since this masterpiece is rich in delicate details, it is sometimes difficult to make a distinction between the cracks and image content. For other particular details, see [1], [2]. We use here the crack detection method from [2].

Crack inpainting methods considered in literature so far are mostly pixel-based, among which controlled anisotropic diffusion [5] is the best performing one. In [22] a patch-based texture synthesis method was used. However, in our case, these pixel-based methods perform insufficiently well due to the width of the cracks, whitish borders around them and the quality of the scans (noise and scanning artefacts). As a consequence, the results are blurry, as illustrated in the middle column of Fig. 4. The results of our context-aware global method are shown in the right column of Fig. 4. By

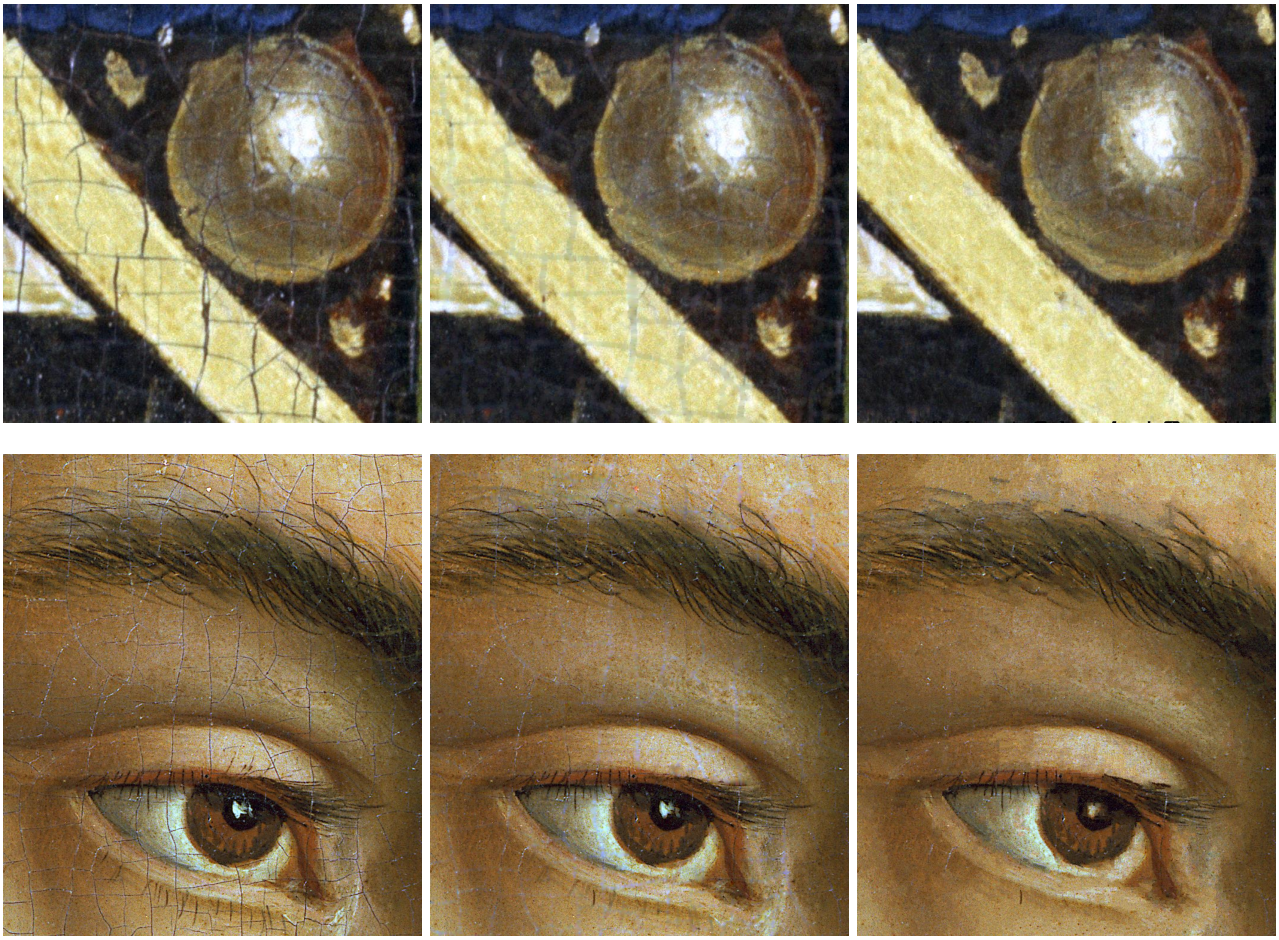

Fig. 4. Inpainting results for the small part of Fig. 1a (top) and Fig. 1b (bottom). From left to right: original image, result of controlled anisotropic diffusion and result of the proposed context-aware global approach.

visual comparison, we can see that our patch-based method outperforms the pixel-based one.

A very challenging problem for inpainting is the book in the *Annunciation to Mary* panel (see Fig. 1c and the top image in Fig. 5). In this case, accurate inpainting is very important because of paleographical deciphering of this text [2]. Many cracks in this panel are difficult to distinguish from the brush strokes, and hence remain undetected. This can misguide the inpainting process: small parts of letters appear erroneously in the background and the other way around, parts of letters get "deleted", i.e. replaced by background. Moreover, positions of cracks remain partly visible. We treated these specific problems in detail in [1], [2], where we also proposed alternative use of contextual information, based on image segmentation into background and foreground. A result of this approach is illustrated in Fig. 5.

IV. Conclusion

In this paper, we introduced a novel MRF based inpainting method that uses contextual descriptors to improve the quality and the efficiency of the inpainting process. We also proposed a simple and efficient way to perform optimization by first pruning the labels to some small number and then separately employing the inference method to obtain final inpainting result. Results in photo-editing application demonstrated the benefits of such an approach in comparison with the state-of-the-art methods, both in terms of quality and speed. The proposed method also yields very encouraging results for crack removal in digitized old paintings, as it was demonstrated on examples from the Ghent Altarpiece.

Acknowledgment

We thank Saint Bavo cathedral, Lukas Art in Flanders and the Dierickfonds for permission to use the Van Eyck images based on photographic negatives (b45, g09, 39-19) from the Dierickfonds in this research report.

References

[1] T. Ružić, B. Cornelis, L. Platiša, A. Pižurica, A. Dooms, W. Philips, M. Martens, M. De Mey, and I. Daubechies, "Virtual restoration of the ghent altarpiece using crack detection and inpainting," in *ACIVS 2011*, 2011, pp. 417–428.

[2] B. Cornelis, T. Ružić, E. Gezels, A. Dooms, A. Piurica, L. Platiša, J. Cornelis, M. Martens, M. D. Mey, and I. Daubechies, "Crack detection and inpainting for virtual restoration of paintings: The case of the ghent altarpiece," *Signal Process.*, vol. 93, no. 3, pp. 605–619, Mar. 2013.

Fig. 5. Top: small part of Fig. 1c. Bottom: inpainting result of the context-aware method from [2].

[3] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *SIGGRAPH (2000)*, New Orleans, USA, 2000.

[4] D. Tschumperlé, "Fast anisotropic smoothing of multi-valued images using curvature-preserving pde's," *Int. Jour. of Comp. Vision*, vol. 68, no. 1, pp. 65–82, 2006.

[5] I. Giakoumis, N. Nikolaidis, and I. Pitas, "Digital image processing techniques for the detection and removal of cracks in digitized paintings," *IEEE Trans. on Imag. Proc.*, vol. 15, no. 1, pp. 178–188, 2006.

[6] S. V. Solanki and A. R. Mahajan, "Cracks inspection and interpolation in digitized artistic picture using image processing approach," *Int. Jour. of Recent Trends in Eng. (IJRTE)*, vol. 1, no. 2, pp. 97–99, May 2009.

[7] M. Barni, F. Bartolini, and V. Cappellini, "Image processing for virtual restoration of artworks," *IEEE Multimedia*, vol. 7, pp. 34–37, 2000.

[8] A. Criminisi, P. Perez, and K.Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. on Imag. Proc.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.

[9] N. Komodakis and G. Tziritas, "Image completion using efficient belief propagation via priority scheduling and dynamic pruning," *IEEE Trans. on Imag. Proc.*, vol. 16, no. 11, pp. 2649–2661, Nov. 2007.

[10] O. Le Meur, J. Gautier, and C. Guillemot, "Examplar-based inpainting based on local geometry," in *ICIP*, 2011, pp. 3462–3465.

[11] Anupam, P. Goyal, and S. Diwakar, "Fast and enhanced algorithm for exemplar based image inpainting," in *PSIVT '10*, 2010, pp. 325–330.

[12] T. Ružić and A. Pižurica, "Texture and color descriptors as a tool for context-aware patch-based image inpainting," in *SPIE Electronic Imaging (2012)*, vol. 7870, 2012.

[13] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Trans. on Imag. Proc.*, vol. 19, no. 15, pp. 1153–1165, May 2010.

[14] J. Sun, L. Yuan, J. Jia, and H.-Y. Shum, "Image completion with structure propagation," *ACM Trans. Graph.*, vol. 24, pp. 861–868, 2005.

[15] T. Ružić, A. Pižurica, and W. Philips, "Markov random field based image inpainting with context-aware label selection," in *ICIP*, 2012, pp. 1733–1736.

[16] J. Jia and C.-K. Tang, "Image repairing: robust image synthesis by adaptive nd tensor voting," in *CVPR*, 2003, pp. 643–650.

[17] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *Int. Jour. of Comp. Vision*, vol. 42, no. 3, pp. 145–175, 2001.

[18] J. Hays and A. A. Efros, "Scene completion using millions of photographs," *Comm. ACM*, vol. 51, no. 10, pp. 87–94, 2008.

[19] T. Ružić, A. Pižurica, and W. Philips, "Neighbourhood-consensus message passing as a framework for generalized iterated conditional expectations," *Pat. Rec. Letters*, vol. 33, pp. 309–318, Feb 2012.

[20] J. S. Yedidia and W. T. Freeman, "On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs," *IEEE Trans. on Inf. Theory*, vol. 47, no. 2, pp. 736–744, Feb. 2001.

[21] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *SIGGRAPH 2001*, 2001, pp. 341–346.

[22] G. S. Spagnolo and F. Somma, "Virtual restoration of cracks in digitized image of paintings," *Journal of Physics: Conference Series*, vol. 249, no. 1, p. 012059, 2010.