Hierarchical Super-Resolution-Based Inpainting

Olivier Le Meur, Mounira Ebdelli, and Christine Guillemot

***Abstract*— This paper introduces a novel framework for examplar-based inpainting. It consists in performing first the inpainting on a coarse version of the input image. A hierarchical super-resolution algorithm is then used to recover details on the missing areas. The advantage of this approach is that it is easier to inpaint low-resolution pictures than high-resolution ones. The gain is both in terms of computational complexity and visual quality. However, to be less sensitive to the parameter setting of the inpainting method, the low-resolution input picture is inpainted several times with different configurations. Results are efficiently combined with a loopy belief propagation and details are recovered by a single-image super-resolution algorithm. Experimental results in a context of image editing and texture synthesis demonstrate the effectiveness of the proposed method. Results are compared to five state-of-the-art inpainting methods.**

***Index Terms*— Examplar-based inpainting, single-image super-resolution.**

I. Introduction

IMAGE inpainting refers to methods which consist in filling in missing regions (holes) in an image [1]. Existing methods can be classified into two main categories.

The first category concerns diffusion-based approaches which propagate linear structures or level lines (so-called isophotes) via diffusion based on partial differential equations [1], [2] and variational methods [3]. The diffusion-based methods tend to introduce some blur when the hole to be filled in is large. The second family of approaches concerns examplar-based methods which sample and copy best matching texture patches from the known image neighbourhood [4]–[7]. These methods have been inspired from texture synthesis techniques [8] and are known to work well in cases of regular or repeatable textures. The first attempt to use examplar-based techniques for object removal has been reported in [6]. The authors in [5] improve the search for similar patches by introducing an a priori rough estimate of the inpainted values using a multi-scale approach which then results in an iterative approximation of the missing regions from coarse-to-fine levels. The two types of methods (diffusion- and examplar-based) can be efficiently combined, e.g. by using structure tensors to compute the priority of the patches to be filled in as in [9]. A recent approach [10] combines an examplar-based approach with super-resolution. It is a two-steps algorithm. First a coarse version of the input picture is inpainted. The second step consists in creating an enhanced resolution picture from the coarse inpainted image.

Although tremendous progress has been made in the past years on examplar-based inpainting, there still exists a number of problems. We believe that the most important one is related to the parameter settings such as the filling order and the patch size.

This problem is here addressed by considering multiple inpainted versions of the input image. To generate this set of inpainted pictures, different settings are used. The inpainted pictures are then combined yielding the final inpainted result. Notice that the inpainting algorithm is preferably applied on a coarse version of the input image; this is particularly interesting when the hole to be filled in is large. This provides the advantage to be less demanding in terms of computational resources and less sensitive to noise and local singularities. In this case the final full resolution inpainted image is recovered by using a super-resolution (SR) method similarly to [10]. SR methods refer to the process of creating one enhanced resolution image from one or multiple input low resolution images. These problems are then referred to as single or multiple images SR, respectively. In both cases, the problem is of estimating high frequency details which are missing in the input image(s). The proposed SR-aided inpainting method falls within the context of single-image SR.

The SR problem is ill-posed since multiple high-resolution images can produce the same low-resolution image. Solving the problem hence requires introducing some prior information. The prior information can be an energy functional defined on a class of images which is then used as a regularization term together with interpolation techniques [11]. This prior information can also take the form of example images or of corresponding LR-HR (Low Resolution - High Resolution) pairs of patches learned from a set of un-related training images [12] or from the input low resolution image itself [13]. This latter family of approaches is known as examplar-based SR methods [12]. An examplar-based super-resolution method embedding K nearest neighbours found in an external patch database has also been described in [14]. Instead of constructing the LR-HR pairs of patches from a set of un-related training images, the authors in [13] extract these correspondences by searching for matches across different scales of a multi-resolution pyramid constructed from the input low-resolution image.

The proposed method builds upon the super-resolution-based inpainting method proposed in [10] which is based on examplar-based inpainting (in particular Criminisi-like

Manuscript received September 14, 2012; revised December 20, 2013 and February 22, 2013; accepted April 15, 2013. Date of publication May 2, 2013; date of current version August 28, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wai-Kuen Cham.

O. Le Meur is with the Institut de Recherche en Informatique et Systèmes Aléatoires, University of Rennes 1, Rennes 35042, France (e-mail: olemeur@irisa.fr).

M. Ebdelli and C. Guillemot are with INRIA, SIROCCO, Rennes 35042, France (e-mail: mounira.ebdelli@inria.fr; christine.guillemot@inria.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2261308

approach [4]) and single-image examplar-based super-resolution [13]. The main novelty of the proposed algorithm is the combination of multiple inpainted versions of the input picture. The rationale behind this approach is to cope with the sensitivity of examplar-based algorithms to parameters such as the patch size and the filling order. Different combinations have been tested and compared. Besides this major point, different adjustments regarding examplar-based inpainting and SR methods are described such as the use of the coherence measure to constrain the candidate search [15].

In summary, the proposed method improves on the state-of-the-art examplar-based inpainting methods by proposing a new framework involving a combination of multiple inpainting versions of the input picture followed by a single-image examplar-based SR method. Notice that the SR method is used only when the inpainting method is applied on a low resolution of the input picture.

The paper is organized as follows. In Section II, an overview of the proposed examplar-based inpainting algorithm is presented. In Section III, the details of the inpainting algorithm as well as the combination of the inpainted pictures are given. Section IV presents the super-resolution method. Experiments and comparisons with state-of-the-art algorithms are performed in Section V. Finally we conclude this work in Section VI.

II. Notations and Algorithm Overview

A. Notations

The following notations are used throughout this paper:

- $\mathbf{I}$ is a color input 2D image:

$$\mathbf{I}: \begin{vmatrix} \Omega \subset \mathcal{R}^n \rightarrow \mathcal{R}^m \\ \mathbf{x} \rightarrow \mathbf{I}(\mathbf{x}) \end{vmatrix} \quad (1)$$

  where $n = 2$ and in this case $\mathbf{x} = (x, y)$ represents a vector indicating spatial coordinates of a pixel $p_{\mathbf{x}}$. Ω is the generic definition domain of images. $\mathbf{I}$ is a color image composed of three components ($m = 3$). Here, we consider the (R, G, B) color space.
- $\mathbf{I}_i : \Omega \rightarrow \mathcal{R}$ represents the i^{th} image channel of $\mathbf{I}$.
- The definition domain Ω is here composed of two parts: $\Omega = S \cup T$, S being the known part of $\mathbf{I}$ (source region) and T the target one or unknown parts of $\mathbf{I}$.
- $\widehat{\mathbf{I}}^{(c)}$ is the inpainted picture obtained by using the c^{th} setting configuration. The setting refers to the patch's size, the size of the search window, the picture's orientation, to name a few. We consider that c is in the range 1 to M.
- $\widehat{\mathbf{I}}^{(*)}$ is the final inpainted picture.
- We denote by $p_{\mathbf{x}}^{(k)}$ the pixel located at $\mathbf{x}$ in the image $\widehat{\mathbf{I}}^{(k)}$.
- $\psi_{p_{\mathbf{x}}}$ is the patch centered on the pixel $p_{\mathbf{x}}$.
- $\psi_{p_{\mathbf{x}}}^{uk}$ denotes the unknown pixels of the patch whereas $\psi_{p_{\mathbf{x}}}^{k}$ denotes its known pixels.
- LR and HR denote Low Resolution and High Resolution respectively. $\psi_{p_{\mathbf{x}}}^{HR}$ is a patch of high resolution.
- $\mathcal{D}^{HR}$ is a set of HR patches (a dictionary) which is used by the super-resolution algorithm.

B. Algorithm Overview

Image completion of large missing regions is a challenging task. As presented in the previous section, there are a number of solutions to tackle the inpainting problem. In this paper, we propose a new inpainting framework relying on both the combination of low-resolution inpainting pictures method and a single-image super-resolution algorithm. In the following sections, we briefly present the main ideas of this paper and the reasons why the proposed method is new and innovative.

The proposed method is composed of two main and sequential operations. The first one is a non-parametric patch sampling method used to fill in missing regions. The inpainting algorithm is preferably applied on a coarse version of the input picture. Indeed a low-resolution picture is mainly represented by its dominant and important structures of the scene. We believe that performing the inpainting of such a low-resolution image is much easier than performing it on the full resolution. A low-resolution image is less contaminated by noise and is composed by the main scene structures. In other words, in this kind of picture, local orientation singularities which could affect the filling order computation are strongly reduced. Second, as the picture to inpaint is smaller than the original one, the computational time is significantly reduced compared to the one necessary to inpaint the full resolution image. To give more robustness, we inpaint the low-resolution picture with different settings (patch's size, filling order, etc). By combining these results, a final low-resolution inpainted picture is obtained. Results will show that the robustness and the visual relevance of inpainting is improved. The second operation is run on the output of the first step. Its goal is to enhance the resolution and the subjective quality of the inpainted areas. Given a low-resolution input image, which is the result of the first inpainting step, we recover its high-resolution using a single-image super-resolution approach. Fig. 1 illustrates the main concept underlying the proposed method namely:

1) a low-resolution image is first built from the original picture;
2) an inpainting algorithm is applied to fill in the holes of the low-resolution picture. Different settings are used and inpainted pictures are combined;
3) the quality of the inpainted regions is improved by using a single-image super-resolution method.

III. Multiple Examplar-Based Inpainting

This section aims at presenting the proposed inpainting method and the combination of the different inpainted images.

A. Examplar-Based Inpainting

The proposed examplar-based method follows the two classical steps as described in [4]: the filling order computation and the texture synthesis. These are described in the next sections.

1) Patch Priority: The filling order computation defines a measure of priority for each patch in order to distinguish the structures from the textures. Classically, a high priority indicates the presence of structure. The priority of a patch centered on $p_{\mathbf{x}}$ is just given by a data and confidence term. The latter is exactly the one defined in [4]. Regarding the data

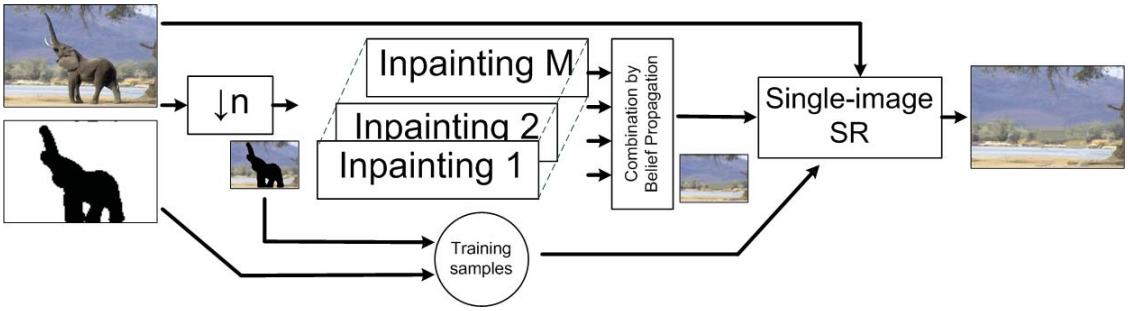


Fig. 1. The framework of the proposed method.

term, a tensor-based [9] and a sparsity-based [16] data terms have been used. These terms are briefly described below.

The tensor-based priority term is based on a structure tensor also called Di Zenzo matrix [17]; this is given by:

$$\mathbf{J} = \sum_{i=1}^{m} \nabla I_i \nabla I_i^T \tag{2}$$

$\mathbf{J}$ is the sum of the scalar structure tensors $\nabla I_i \nabla I_i^T$ of each image channel I_i (R,G,B). The tensor can be smoothed without cancellation effects: $\mathbf{J}_\sigma = \mathbf{J} * G_\sigma$ where $G_\sigma = 1/2\pi\sigma^2 \exp(-(x^2+y^2)/2\sigma^2)$, with standard deviation σ. One of the main advantages of a structure tensor is that a structure coherence indicator can be deduced from its eigenvalues. Based on the discrepancy of the eigenvalues, the degree of anisotropy of a local region can be evaluated. The local vector geometry is computed from the structure tensor $\mathbf{J}_\sigma$. Its eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$ ($\mathbf{v}_i \in R^n$) define an oriented orthogonal basis and its eigenvalues $\lambda_{1,2}$ define the amount of structure variation. The vector $\mathbf{v}_1$ is the orientation with the highest fluctuations (orthogonal to the image contours), whereas $\mathbf{v}_2$ gives the preferred local orientation. This eigenvector (having the smallest eigenvalue) indicates the isophote orientation. A data term D is then defined as [18]:

$$D(p_\mathbf{x}) = \alpha + (1-\alpha)\exp\left(-\frac{\eta}{(\lambda_1-\lambda_2)^2}\right) \tag{3}$$

where η is a positive value and $\alpha \in [0,1]$ ($\eta = 8$ and $\alpha = 0.01$). On flat regions ($\lambda_1 \approx \lambda_2$), any direction is favoured for the propagation (isotropic filling order). When $\lambda_1 >> \lambda_2$ indicating the presence of a structure, the data term is important.

The sparsity-based priority has been proposed recently by Xu et al. [16]. In a search window, a template matching is performed between the current patch $\psi_{p_\mathbf{x}}$ and neighbouring patches $\psi_{p_\mathbf{j}}$ that belong to the known part of the image. By using a non-local means approach [15], a similarity weight $w_{p_\mathbf{x},p_\mathbf{j}}$ (i.e. proportional to the similarity between the two patches centered on $p_\mathbf{x}$ and $p_\mathbf{j}$) is computed for each pair of patches. The sparsity term is defined as:

$$D(p_\mathbf{x}) = \left\|\mathbf{w}_{p_\mathbf{x}}\right\|_2 \times \sqrt{\frac{|N_s(p_\mathbf{x})|}{|N(p_\mathbf{x})|}} \tag{4}$$

where N_s and N represent the number of valid patches (having all its pixels known) and the total number of candidates in the search window. When $\left\|\mathbf{w}_{p_\mathbf{x}}\right\|_2$ is high, it means larger sparseness whereas a small value indicates that the current input patch is highly predictable by many candidates.

2) Texture Synthesis: The filling process starts with the patch having the highest priority. To fill in the unknown part of the current patch $\psi^{uk}_{p_\mathbf{x}}$, the most similar patch located in a local neighbourhood $\mathcal{W}$ centered on the current patch is sought. A similarity metric is used for this purpose. The chosen patch $\psi^*_{p_\mathbf{x}}$ maximizes the similarity between the known pixel values of the current patch to be filled in $\psi^k_{p_\mathbf{x}}$ and co-located pixel values of patches belonging to $\mathcal{W}$:

$$\psi^*_{p_\mathbf{x}} = \arg\min_{\psi_{p_\mathbf{j}} \in \mathcal{W}} d\left(\psi^k_{p_\mathbf{x}}, \psi^k_{p_\mathbf{j}}\right) \quad \text{s.t } Coh\left(\psi^{uk}_{p_\mathbf{x}}\right) < \lambda_{coh} \tag{5}$$

where $d(.)$ is the weighted Bhattacharya used in [10]. $Coh(.)$ is the coherence measure initially proposed by Wexler *et al.* [15]:

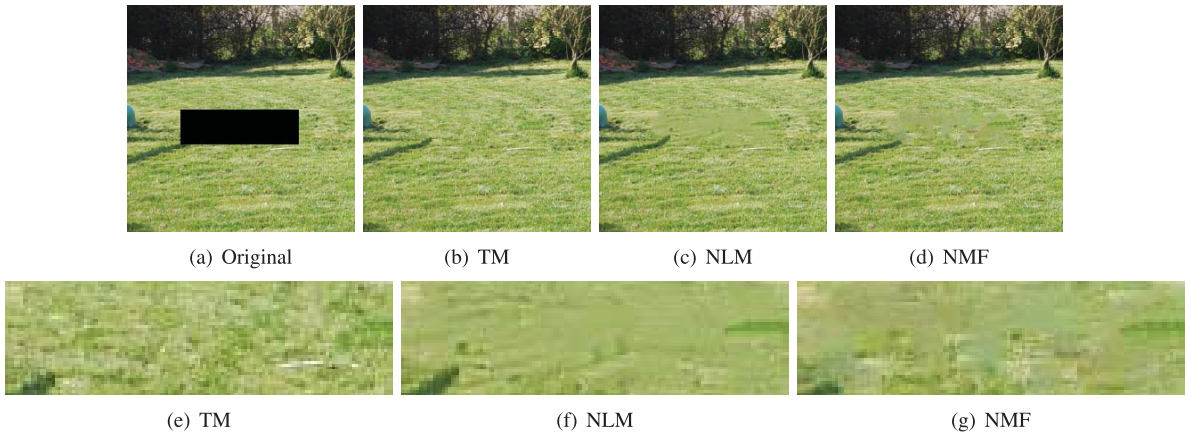
$$Coh(\psi^{uk}_{p_\mathbf{x}}) = \min_{p_\mathbf{j} \in S}\left(d_{SSD}(\psi^{uk}_{p_\mathbf{x}}, \psi^{uk}_{p_\mathbf{j}})\right) \tag{6}$$

where d_{SSD} is the sum of square differences. The coherence measure Coh simply indicates the degree of similarity between the synthesized patch $\psi^{uk}_{p_\mathbf{x}}$ and original patches. Therefore, the constraint in equation (5) prevents pasting in the unknown regions a texture that would be too different from original textures. If none of the candidates fulfil the constraint (5), the filling process is stopped and the priority of the current patch is decreased. The process restarts by seeking the patch having the highest priority. It is interesting to note that a recent study [19] uses a similar term to predict the quality of the inpainting.

Compared to our previous work [10], there is another substantial difference: we only use the best match to fill in the hole whereas a linear combination of the K most similar patches is generally performed to compute the patch $\psi^*_{p_\mathbf{x}}$ in [10], [15], [16], [20]. In these cases, the estimated patch is then given by:

$$\psi^*_{p_\mathbf{x}} = \sum_{i}^{K} w_{p_\mathbf{x},p_\mathbf{i}} \times \psi^k_{p_\mathbf{i}} \tag{7}$$

where K is the number of candidates which is often adapted locally so that the similarity of chosen neighbours lies within a range $(1+\alpha) \times d_{\min}$, where $d_{\min}$ is the distance between the current patch and its closest neighbours. Different methods can be used to compute the weights. It could be based on either a non-negative matrix factorization (NMF) [21] or a non local



(a) Original (b) TM (c) NLM (d) NMF

(e) TM (f) NLM (g) NMF

Fig. 2. (a) Original picture with a black hole to be filled in, (b) result obtained with a simple template matching (TM), with the linear combination of the K most similar neighbors in which weights are computed by using a non local means method (c), and by a non negative matrix factorization (d). For this example, the maximum number of neighbors is 10. (e)–(g) zoom in the inpainted part for TM, NLM, and NMF, respectively.

(a) Mask (b) S_1 (c) S_2 (d) S_3 (e) S_4 (f) S_5 (g) S_6 (h) S_7 (i) S_8

Fig. 3. The low-resolution picture is inpainted with different settings S_x (see Table I for the setting details). (a) Original picture with the hole to be filled in. Pictures S_1 up to S_8 (from (b) to (i)) are the inpainting results (note the setting sensibility of the inpainting algorithm).

means filter [15], [22], to name a few. Combining several candidates increases the algorithm robustness. However, it tends to introduce blur on fine textures as illustrated by Fig. 2. In our method, only the best candidate is chosen. Its unknown parts are pasted into the missing areas. A Poisson fusion [23] is applied to hide the seams between known and unknown parts. An α-blending is also applied to combine known parts of the candidate and the source patch ($\psi^k_{p_{\mathbf{x}}} \leftarrow \alpha\psi^k_{p_{\mathbf{x}}} + (1-\alpha)\psi^{k,*}_{p_{\mathbf{x}}}$, with $\alpha = 0.75$). It gives more robustness by locally regularizing the results.

Although the proposed method is able to fill in holes in a visually pleasant fashion (as illustrated by Figs. 2 and 9), it still suffers from problems of one-pass greedy algorithms. Indeed for most of existing approaches, the setting such as the patch size and the filling order, to name the most important factors, may dramatically impact the quality of results. To overcome this issue, we combine inpainted pictures obtained when different settings are used. In this study, we consider $M = 13$, meaning that the low-resolution picture is inpainted 13 times. Parameters are given in Table I: the patch size is chosen between 5×5, 7×7, 9×9 and 11×11. The filling order is computed by either the sparsity-based or the tensor-based method. The input picture can also be rotated by 180 degrees. This allows changing the filling order.

TABLE I

THIRTEEN CONFIGURATIONS USED TO FILL IN THE UNKNOWN PARTS OF THE PICTURES

Setting	Parameters
1 (default)	Patch's size 5×5 Decimation factor $n = 3$ Search window 80×80 Sparsity-based filling order
2	default + rotation by 180 degrees
3	default + patch's size 7×7
4	default + rotation by 180 degrees + patch's size 7×7
5	default + patch's size 11×11
6	default + rotation by 180 degrees + patch's size 11×11
7	default + patch's size 9×9
8	default + rotation by 180 degrees + patch's size 9×9
9	default + patch's size 9×9 + Tensor-based filling order
10	default + patch's size 7×7 + Tensor-based filling order
11	default + patch's size 5×5 + Tensor-based filling order
12	default + patch's size 11×11 + Tensor-based filling order
13	default + rotation by 180 degrees + patch's size 9×9 + Tensor-based filling order

B. Combining Multiple Inpainted Images

The combination aims at producing a final inpainted picture from the M inpainted pictures. Before delving into this subject in details, Fig. 3 illustrates some inpainted results obtained for a given setting. We notice again that the setting plays an important role. To obtain the final inpainted picture, three kinds of combination have been considered. The first two methods are very simple since every pixel value in the final picture is achieved by either the average or the median operator

as given below:

$$\widehat{\mathbf{I}}^{(*)}(p_{\mathbf{x}}) = \frac{1}{M}\sum_{i=1}^{M}\widehat{\mathbf{I}}^{(i)}(p_{\mathbf{x}}) \quad (8)$$

$$\widehat{\mathbf{I}}^{(*)}(p_{\mathbf{x}}) = MED_{i=1}^{M}\widehat{\mathbf{I}}^{(i)}(p_{\mathbf{x}}) \quad (9)$$

The advantage of these operators is their simplicity. However they suffer from at least two main drawbacks. The average operator as well as the median one do not consider the neighbours of the current pixel to take a decision. Results might be more spatially coherent by considering the local neighbourhood. In addition, the average operator inevitably introduces blur as illustrated by Fig. 5.

To cope with these problems, namely blur and spatial consistency of the final result, the combination is achieved by minimizing an objective function. Rather than using a global minimization that would solve exactly the problem, we use a Loopy Belief Propagation which in practice provides a good approximation of the problem to be solved. This approach is described in the next section.

1) Loopy Belief Propagation: As in [24], the problem is to assign a label to each pixel $p_{\mathbf{x}}$ of the unknown regions T of the picture $\widehat{\mathbf{I}}^{(*)}$. The major drawback of the belief propagation is that it is slow especially when the number of labels is high. Komodakis and Tziritas [24] have designed a priority Belief Propagation in order to deal with this complexity bottleneck. Indeed, the number of labels in [24] is equal to the number of patches in the source region. Here the approach is simpler since the number of labels is rather small; a label is simply the index of the inpainted picture from which the patch is extracted. A finite set of labels $\mathcal{L}$ is then composed of M values ($M = 13$ here), going from 1 to M. This problem can be formalized with a Markov Random Field (MRF) $G = (\upsilon, \epsilon)$ defined over the target region T. The MRF nodes υ are the lattice composed of pixels inside T. Edges ϵ are the four-connected image grid graph centered around each node. We denote $\mathcal{N}_4$ this neighbourhood system. The labelling assigns a label l ($l \in \mathcal{L}$) to each node/pixel $p_{\mathbf{x}}$ ($p_{\mathbf{x}} \in T$) so that the total energy E of the MRF is minimized (we denote by l_p the label of pixel $p_{\mathbf{x}}$) [25], [26]:

$$E(l) = \sum_{\mathbf{p}\in\upsilon} V_d(l_p) + \sum_{(n,m)\in\mathcal{N}_4} V_s(l_n, l_m) \quad (10)$$

where,

- $V_d(l_p)$ is called the label cost (or the data cost) [24]. This represents the cost of assigning a label l_p to a pixel p_x. This is given by:

$$V_d(l_p) = \sum_{n\in\mathcal{L}}\sum_{\mathbf{u}\in\upsilon}\left\{\widehat{\mathbf{I}}^{(l)}(\mathbf{x}+\mathbf{u}) - \widehat{\mathbf{I}}^{(n)}(\mathbf{x}+\mathbf{u})\right\}^2 \quad (11)$$

  where υ is a square neighbourhood (3×3) centered on the current pixel. The cost increases when the dissimilarity between the current patch and colocated patches in other inpainted pictures is high.
- The pairwise potential or the discontinuity cost $V_s(l_n, l_m)$ is a quadratic cost function given by:

$$V_s(l_n, l_m) = \lambda \times (l_n - l_m)^2 \quad (12)$$

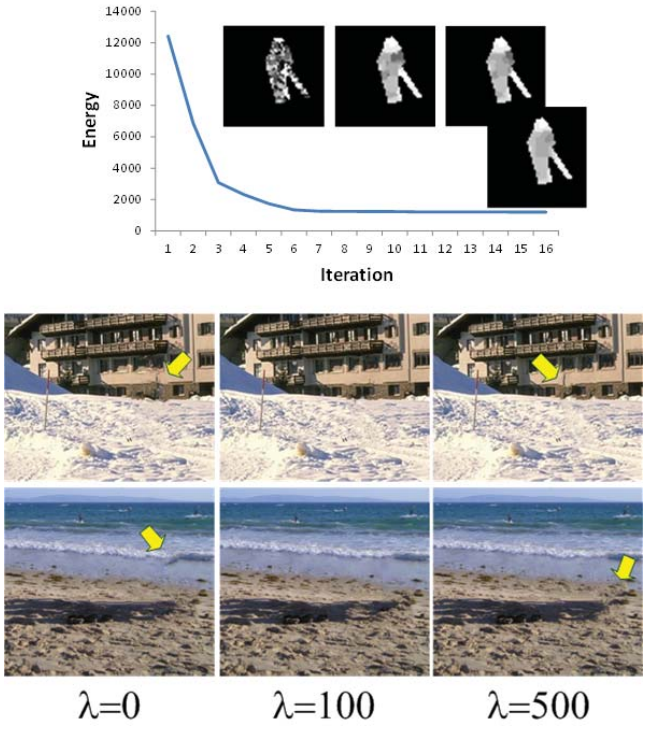


Fig. 4. Illustration of the energy minimization. On the top, the convergence process of the LBP is illustrated. The labeling obtained for four different iterations are also shown. The bottom images show the influence of the smoothness term for three values of λ (see Equation (12)).

where λ is a weighting factor and is set to 100 (Fig. 4 illustrates the influence of this factor). The discontinuity cost is here based on a difference between labels rather than the difference between pixel values.

The minimization of the energy E over the target region T can be achieved using loopy belief propagation (LBP) [27] and corresponds to the maximum a posteriori (MAP) estimation problem for an appropriately defined MRF [28]. Fig. 4 illustrates the minimization of the total energy as well as the influence of the smoothness term. When $\lambda = 0$, there is no smoothness term. Some artefacts indicated by arrows are visible. When $\lambda = 500$, artefacts are visible. A good trade-off is obtained by setting the value λ to 100. On the same Fig. 4, the labelling convergence is given for four iterations (pictures on top of this figure represent label values, not pixel values). This illustrates the fact that the label choice is not greedy.

2) Comparison of the Combination Methods: Fig. 5 illustrates the performance of the combination methods. As expected, when the different inpainted pictures are averaged, the reconstructed areas are blurred. The blur is less striking when the median operator is used to combine pictures. The LBP method provides the best result. The texture is well retrieved and thanks to the global energy minimization results are spatially consistent. In the following, we use the LBP method to combine low-resolution inpainted pictures.

3) Possible Extensions: Two extensions of the proposed method could be also envisioned. First, rather than combining the low-resolution inpainted pictures, the best impainting picture could be selected either by a user (semi-supervised

Fig. 5. Comparison of combination methods. (a) Input picture, (b) results obtained by averaging all inpainted pictures, (c) by taking the median pixel values, and (d) by using a Loopy belief propagation.

inpainting) or by automatically evaluating the inpainting quality using a coherence measure as proposed in [19].

IV. Super-Resolution Algorithm

Once the combination of the low-resolution inpainted pictures is completed, a hierarchical single-image super-resolution approach is used to reconstruct the high resolution details of the image. We stress the point that the single-image SR method is applied only when the input picture has been downsampled for the inpainting purpose. Otherwise the SR method is not required. As in [10], [12], the problem is to find a patch of higher-resolution from a database of examples. The main steps, illustrated in Fig. 6 are described below:

1) Dictionary building: it consists of the correspondences between low and high resolution image patches. The unique constraint is that the high-resolution patches have to be valid, i.e. entirely composed of known pixels. In the proposed approach, high-resolution and valid patches are evenly extracted from the known parts of the image. The size of the dictionary is a user-parameter which might influence the overall speed/quality trade-off. An array is used to store the spatial coordinates of HR patches ($\mathcal{D}^{HR}$). Those of LR patches are simply deduced by using the decimation factor equal to 2;
2) Filling order of the HR picture: The computation of the filling order is similar to the one described in Section III. It is computed on the HR picture with the sparsity-based method. The filling process starts with the patch $\psi_{p_{\mathbf{x}}}^{HR}$ having the highest priority and which is composed of known and unknown parts. Compared to a raster-scan filling order, it allows us to start with the structures and then to preserve them;
3) For the LR patch corresponding to the HR patch having the highest priority, its best neighbour in the inpainted images of lower resolution is sought. This search is performed in the dictionary and within a local neighbourhood. Only the best candidate is kept. From this LR candidate, a HR patch is simply deduced. Its pixel values are then copied into the unknown parts of the current HR patch $\psi_{p_{\mathbf{x}}}^{HR}$.

After the filling of the current patch, the priority value is propagated and the aforementioned steps are iterated while there exist unknown areas. A Poisson and alpha-blending are again used to hide seams between known and unknown parts and to improve robustness.

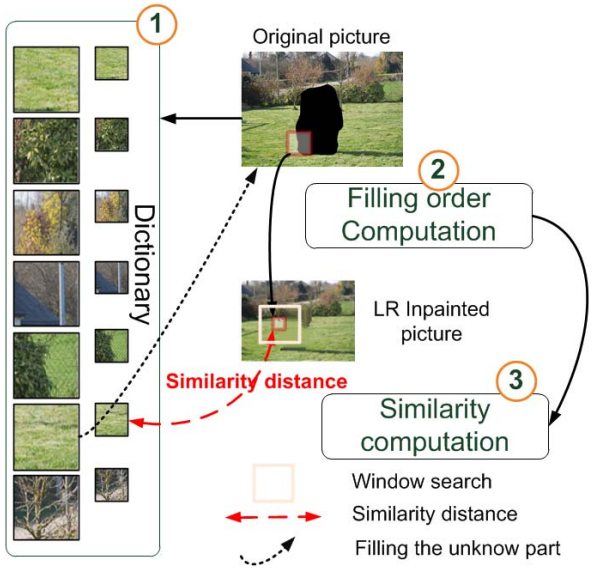


Fig. 6. Flowchart of the super-resolution algorithm. The missing parts of the red block are filled in by the best candidate stemming either from the dictionary or from the local neighborhood. The top image represents the original image with the missing areas whereas the bottom one is the result of the low-resolution inpainting.

Compared to [10], the SR method is applied in a hierarchical manner. For instance, if the input picture of resolution (X, Y) has been down-sampled by four in both directions, the SR algorithm is applied twice: a first time to recover the resolution $(\frac{X}{2}, \frac{Y}{2})$ and a second time to recover the native resolution.

V. Experimental Results

In this section, the proposed approach is tested on a variety of natural images and compared to five state-of-the-art inpainting methods.

A. Intrinsic Performance of the Proposed Method

Figs. 7 and 8 illustrate the results of the proposed approach. A down sampling factor of 4 in both directions is used. Pictures have a resolution varying in the range 420×380 to 720×512. Fig. 7(b) gives the inpainted low-resolution picture whereas the column (c) illustrates the final result. The proposed approach faithfully recovers the texture such as the grass, the sand and the snow. Structures are also well recovered. Fig. 8 presents more results which are visually plausible and pleasing in most of the cases. The less-favourable results are obtained when the hole to be filled in is rather small. In this case it might be better to reduce the down sampling factor or even to perform the inpainting at the full resolution.

B. Parameters Analysis

When the proposed method, called the baseline method in the following, has been defined, we have fixed several

Fig. 7. Results of the proposed method. (a) Input images with unknown regions, (b) inpainted low-resolution images, and (c) final inpainted images (for the sake of visibility, we do not respect the down sampling factor of 4 between pictures (b) and (c)).

Fig. 8. Results of the proposed approach for a variety of pictures of natural content. The unknown parts in black visible on pictures of odd rows are more or less large and complex to fill in. Pictures of even rows are our results.

parameters such as the number of K-NN for the inpainting ($K = 1$), the similarity metric (weighted Bhattacharya) and the color space (RGB). We evaluate subjectively the relevance of these choices on two pictures. Results are illustrated in Fig. 9.

1) K-NN Patches for Inpainting: In Section III-A2, we set the number of K-NN for the examplar-based inpainting to 1 preventing blur apparition as illustrated by Fig. 2. However, as the inpainting is applied on a low-resolution picture, it might

(a) (b) (c) (d) (e) (f)

Fig. 9. Parameters analysis. (a) Baseline method, (b) baseline + $K = 4$, (c) baseline + $K = 8$, (d) Y space, (e) SSD, and (f) with a decimation factor of 2 in both directions (instead of 4).

(a) Original (b) Criminisi [4] (c) PatchMatch [7] (d) Ours

Fig. 10. Comparison with state-of-the-art methods: (a) original pictures with unknown regions, (b) Criminisi's algorithm, (c) PatchMatch method, and (d) proposed method.

make sense to use more than one candidate. We compare results obtained by the baseline approach to those obtained for three different K values: 1 (baseline algorithm), 4 and 8. Results are illustrated in Fig. 9. Results are similar for $K = 4$. For $K = 8$, the inpainted quality is not as visually pleasing as the one obtained by the two previous settings.

2) RGB Space Versus Y Component: The color space, used to perform the candidate search for instance, might play an important role. By using the luminance component only, results are visually less good than the baseline one, especially for the first picture.

3) SSD Versus Weighted Bhattacharya: Qualitatively speaking, inpainted results obtained with the weighted Bhattacharya metric are slightly better than those obtained with the SSD metric.

4) Downsampling Factor Equal to 2: By performing the inpainting at a higher resolution, results are not as good as those obtained by the baseline approach, especially on the first picture (first row). This could be explained by the size of the search window which should be increased for this condition.

C. Comparison With State-of-the-Art Methods

State-of-the-art inpainting methods are used for comparison purposes. The first one is the well known method proposed by Criminisi *et al.* [4]. We have used a third part implementation available on http://www.cc.gatech.edu/~sooraj/inpainting/ (we use a patch's size of 9×9 and a local windows search of 80×80). The second one is the PatchMatch method [7] which is a fast algorithm for computing dense approximate nearest neighbour correspondences between patches of two image regions. This algorithm is available in Adobe Photoshop CS5. The shift-map method [29] is the third algorithm considered in the test. This approach is treating inpainting as a global optimization in a multi-scale

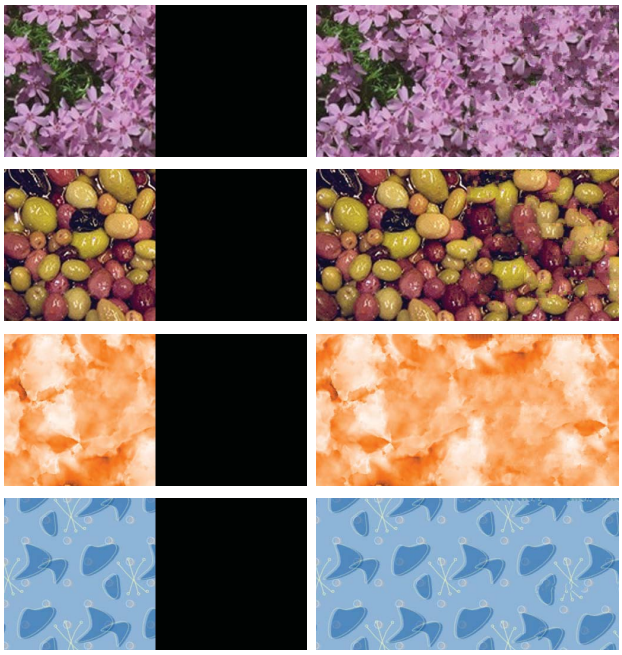
(a) Original (b) He [30] (c) Ours

(a) Original (b) He [30] (c) Shift-Map [29] (d) BP [24] (e) Ours

Fig. 11. Comparison with state-of-the-art results. On the three first rows: (a) original image, (b) He's results (extracted from [30]), and (c) proposed approach. On the three last rows: (a) original image, (b) He's results, (c) Shift-map results [29], (d) BP results [24], and (d) proposed approach.

scheme. The goal is to shift pixel values from the known part of the picture to the missing part. A graph cut optimization allows finding the best mapping. Inspired by [29], He and Sun [30] recently proposed a new method. The idea is to find the most important 2D offsets (displacement) in the picture. These offsets are then used to shift the input picture. The shifted images are finally combined. Method of Komodakis and Tziritas [24] relying on priority belief propagation is also tested.

Fig. 10 illustrates results of the proposed approach, Criminisi *et al.* [4] and PatchMatch method [7]. Our method outperforms Criminisi *et al.* [4]. Compared to PatchMatch method [7], results are similar for the two first pictures and better for the two last pictures. Comparison between the



(a) (b)

Fig. 12. Examples of texture synthesis. (a) Chunk of texture (256×256) with the unknown parts in black (256×256). (b) Results obtained by the proposed approach.

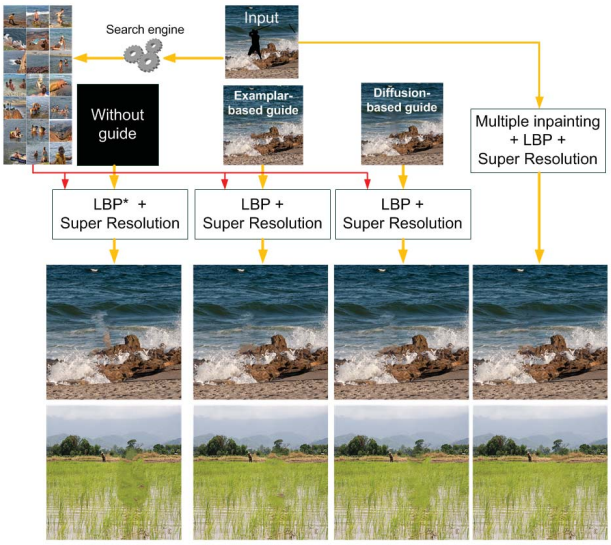


Fig. 13. Extension of the proposed approach. From a set of similar images, an inpainting is performed on a low-resolution version of the input picture. The LBP uses a set of similar images (21 images here (top-left)) and a guide (if any). The SR is applied to recover the details of the missing pixels. Different guides (examplar-based or diffusion-based inpainting) can be used to drive the LBP. (*) (in the first box on the left) means that the term V_g of Equation (13) is null. Bottom-up line gives other results.

proposed approach and He's method [30], shift-map [29] and priority belief propagation [24] is given on Fig. 11. The tested pictures are extracted from [30]. Concerning the comparison between our method and He's method, results are quite similar except for the first and last pictures for which He's approach gives more visually pleasing results, although there is no obvious artefacts in our results. Regarding shift-map and BP methods (the three last rows of Fig. 11), artefacts are visible such as on the stone wall (fourth row) and on the waterfall (BP method on the fifth row). On these pictures, our method is more robust than shift-map and Belief Propagation methods.

D. Extension

1) Application to Texture Synthesis: Fig. 12 presents additional results for texture synthesis. A small chunk of texture was placed into the upper left corner of an empty image. The goal is to fill in the missing parts composed here of 256×256 pixels in a visually plausible way. This application is interesting in itself, but it also allows showing a limitation of the proposed approach. When the texture is stochastic as illustrated by Fig. 12, there exist many ways to fill in the holes in a visually plausible fashion. Performing M different inpaintings and combining them would lead to a poor quality. To overcome this limitation, the input texture is inpainted just once at a low resolution. The SR method is then applied. For texture synthesis, the patch size is arbitrarily set to 9×9 and we down-sample the input chunk by 4 in both directions. In addition, two modifications are also brought to the algorithm. First we limit the search for the closest candidate to the input picture (here it is the original texture chunk). Second the filling order is the raster scan order.

The two first texture chunks illustrated by Fig. 12 are complex to synthesize. Some artefacts are visible on the first two results whereas for the two last results the synthesised texture has a very good quality.

2) Using Similar Images: Another extension would be to use similar images in a similar fashion to Hays and Efros [31] and Whyte *et al.* [32]. We have adapted the proposed method to take as input the image to inpaint, similar images to the image source and an image which is used as a guide for the LBP. For that the energy of the LBP has been modified as follows:

$$E(l) = \sum_{\mathbf{p} \in \nu} V_d(l_p) + \sum_{(\mathbf{p},\mathbf{q}) \in \nu} V_s(l_p, l_q) + \sum_{\mathbf{p} \in \nu} V_g(l_p) \quad (13)$$

where, the first term is the same as previously defined (see equation (11)). The second term is here different since it involves pixel values rather than label values as in [30]. This term is given by:

$$V_s(m,n) = \|\widehat{\mathbf{I}}^{(m)}(\mathbf{p}) - \widehat{\mathbf{I}}^{(m)}(\mathbf{q})\|^2 + \|\widehat{\mathbf{I}}^{(n)}(\mathbf{p}) - \widehat{\mathbf{I}}^{(n)}(\mathbf{q})\|^2 \quad (14)$$

with $(\mathbf{p}, \mathbf{q})$ are neighbouring pixels and m, n represent their labels. The third term V_g is used to guide the labelling. It is given by:

$$V_g(n) = \|\widehat{\mathbf{I}}^{(n)}(\mathbf{p}) - \widehat{\mathbf{I}}_g(\mathbf{p})\|^2 \quad (15)$$

where $\widehat{\mathbf{I}}_g$ is a prediction image of the unknown pixels. The prediction can be obtained by a classical inpainting such as our baseline method presented in Section III-A or by an diffusion-based method as illustrated in Fig. 13. This term is useful to select the most appropriate pixel values in the set of similar pictures: for a given pixel $\mathbf{p}$, if the candidate pixel value at that location is close to the predicted one $\widehat{\mathbf{I}}_g(\mathbf{p})$, the associate cost is low. Fig. 13 illustrates this method. From an image to inpaint

TABLE II
RUNNING TIME IN FUNCTION OF THE NUMBER OF MISSING PIXELS. THE RUNNING TIME IS GIVEN FOR THE LOW-RESOLUTION INPAINTINGS AND THE SUPER-RESOLUTION

Picture	Resolution	Missing Areas	Inpainting and LBP	SR	Total
Elephant	480 × 320	17%	32s	1m58s	2m30s
SnowBall	480 × 320	13%	22s	1m36s	1m58s
Tiger	480 × 320	28%	59s	2m36s	3m35s
Soldier	320 × 480	30%	51s	2m39s	3m30s
Cow	600 × 400	13%	40s	2m11s	2m51s

(top-middle), 21 similar images were found by using *Google Image search*. To guide the LBP, we use the inpainted image from either our baseline inpainting method or the Navier-Stoke diffusion method [33]. These two guides coupled with the similar images already provide quite good results. Result of the method involving multiple inpainting (bottom-right of Fig. 13) is also shown and further improves the quality.

Note that rather than using similar images, it would be possible to use state-of-the-art inpainted pictures and to combine them. This kind of solution might improve performance because of the better complementary of state-of-the-art inpainting methods. It would be possible to mix results from examplar-based, diffusion-based and shift-map inpainting methods. Finally, a semi-supervised approach for which the user is able to choose the best low-resolution inpainted picture could be easily designed. In this case, it is no longer necessary to combine them.

E. Running Time

Table II gives the running time of the proposed approach. Simulation has been performed on a laptop with an Intel Core i7 2.40GHz and 4Go RAM. As the proposed approach is not multithreaded, it just uses one core. In addition, no optimization was made. Table II indicates that the super-resolution process is the most time-consuming step (in average it represents 80% of the total time). It is due to the template matching which is not parallelized. To improve the performance, the template matching could be replaced by an approximate nearest neighbour search [7].

VI. CONCLUSION

A novel inpainting approach has been presented in this paper. The input picture is first down sampled and several inpaintings are performed. The low-resolution inpainted pictures are combined by globally minimizing an energy term. Once the combination is completed, a hierarchical single image super resolution method is applied to recover details at the native resolution. Experimental results on a wide variety of images have demonstrated the effectiveness of the proposed method.

One interesting avenue of future work would be to extend this approach to the temporal dimension. Also, we plan to test other SR methods to bring more robustness to the method. But the main important improvement is likely the use of geometric constraint and higher-level information such as scene semantics in order to improve the visual relevance.

Windows executable is available on authors webpage: http://people.irisa.fr/Olivier.Le_Meur/.

REFERENCES

[1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. 27th Annu. Conf. Comput. Graph. Interact. Tech.*, Jul. 2000, pp. 417–424.
[2] D. Tschumperlé and R. Deriche, "Vector-valued image regularization with PDEs: A common framework for different applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 506–517, Apr. 2005.
[3] T. Chan and J. Shen, "Variational restoration of non-flat image features: Models and algorithms," *SIAM J. Appl. Math.*, vol. 61, no. 4, pp. 1338–1361, 2001.
[4] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by examplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.
[5] I. Drori, D. Cohen-Or, and H. Yeshurun, "Fragment-based image completion," *ACM Trans. Graph.*, vol. 22, no. 2003, pp. 303–312, 2003.
[6] P. Harrison, "A non-hierarchical procedure for re-synthesis of complex texture," in *Proc. Int. Conf. Central Eur. Comput. Graph., Vis. Comput. Vis.*, 2001, pp. 1–8.
[7] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, p. 24, Aug. 2009.
[8] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. 7th IEEE Comput. Vis. Pattern Recognit.*, Sep. 1999, pp. 1033–1038.
[9] O. Le Meur, J. Gautier, and C. Guillemot, "Examplar-based inpainting based on local geometry," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 3401–3404.
[10] O. Le Meur and C. Guillemot, "Super-resolution-based inpainting," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 554–567.
[11] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. Katsaggelos, "SoftCuts: A soft edge smoothness prior for color image super-resolution," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 969–981, May 2009.
[12] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar.–Apr. 2002.
[13] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 10. Oct. 2009, pp. 349–356.
[14] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, vol. 1. Jun.–Jul. 2004, pp. 275–282.
[15] Y. Wexler, E. Shechtman, and M. Irani, "Space-time video completion," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun.–Jul. 2004, pp. I-120–I-127.
[16] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1153–1165, May 2010.
[17] S. Di Zenzo, "A note on the gradient of a multi-image," *Comput. Vis., Graph., Image Process.*, vol. 33, no. 1, pp. 116–125, 1986.
[18] J. Weickert, "Coherence-enhancing diffusion filtering," *Int. J. Comput. Vis.*, vol. 32, nos. 2–3, pp. 111–127, 1999.
[19] J. Kopf, W. Kienzle, S. Drucker, and S. B. Kang, "Quality prediction for image completion," *ACM Trans. Graph.*, vol. 31, no. 6, p. 131, 2012.
[20] A. Bugeau, M. Bertalmío, V. Caselles, and G. Sapiro, "A comprehensive framework for image inpainting," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2634–2644, Oct. 2010.
[21] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing System*. Cambridge, MA, USA: MIT Press, 2000.
[22] A. Buades, B. Coll, and J. Morel, "A non local algorithm for image denoising," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2005, pp. 60–65.
[23] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," in *Proc. SIGGGRAPH*, 2003, pp. 313–318.
[24] N. Komodakis and G. Tziritas, "Image completion using efficient belief propagation via priority scheduling and dynamic pruning," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2649–2661, Nov. 2007.

[25] Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1222–1239, Nov. 2001.
[26] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
[27] J. Yedidia, W. Freeman, and Y. Weiss, "Constructing free energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, Jul. 2005.
[28] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA, USA: MIT Press, 1987.
[29] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. Int. Conf. Comput. Vis.*, Sep. 2009, pp. 151–158.
[30] K. He and J. Sun, "Statistics of patch offsets for image completion," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 16–29.
[31] J. Hays and A. A. Efros, "Scene completion using millions of photographs," *ACM Trans. Graph.*, vol. 26, no. 3, p. 4, 2007.
[32] O. Whyte, J. Sivic, and A. Zisserman, "Get out of my picture! Internet-based inpainting," in *Proc. 20th Brit. Mach. Vis. Conf.*, 2009, pp. 1–11.
[33] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2001, pp. 355–362.

Olivier Le Meur received the Ph.D. degree from the University of Nantes, Nantes, France, in 2005. From 1999 to 2009, he has worked in the media and broadcasting industry. In 2003, he joined the Research Center of Thomson-Technicolor, Rennes, France, where he supervised a research project concerning the modeling of the human visual attention. Since 2009, he has been an Associate Professor for image processing with the University of Rennes 1, Rennes. His current research interests include the understanding of the human visual attention, computational modeling of the visual attention, and saliency-based applications (video compression, objective assessment of video quality, retargeting).

Mounira Ebdelli received the Engineering degree from the National School of Computer Science, Tunisia, France, in 2005, and the M.Sc. degree in perception and digital communication from the National School of Engineering, Tunisia and Paris-Descartes, France, in 2010. She is currently pursuing the Ph.D. degree with the SIROCCO Research Group, INRIA, Rennes, France. From 2005 to October 2008, she was a Developer with STMicroelectronics, Tunisia, and she was a Teaching Assistant with the High Institute of Technological Studies, Zaghouan, Tunisia, from 2008 to November 2010. Her current research interests include signal processing with an emphasis on image and video processing.

Christine Guillemot is a Director of Research with INRIA, the Head of a Research Team with image and video modeling, processing, coding, and communication. She received the Ph.D. degree from Ecole Nationale Supérieure des Télécommunications, Paris, France, and the "Habilitation for Research Direction" degree from the University of Rennes, Rennes, France. From 1985 to 1997, she has been with FRANCE TELECOM, where she has been involved in various projects in the area of image and video coding for TV, HDTV and multimedia. From 1990 to 1991, she was with Bellcore, NJ, USA, as a Visiting Scientist. She has co-authored 15 patents, eight book chapters, 50 journal papers, and 140 conference papers. She has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2000 to 2003, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2004 to 2006, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2007 to 2009. She is currently an Associate Editor for the *Eurasip Journal on Image Communication*. She has been a member of the IEEE IMDSP and MMSP technical committees.