A Variational Framework for Non-local Image Inpainting

Pablo Arias[†], Vicent Caselles[†], and Guillermo Sapiro[‡]

†Universitat Pompeu Fabra, DTIC, 08018 Barcelona, Spain
‡University of Minnesota, ECE, Minneapolis, MN 55455, USA
{pablo.arias,vicent.caselles}@upf.edu,guille@umn.edu

Abstract. Non-local methods for image denoising and inpainting have gained considerable attention in recent years. This is in part due to their superior performance in textured images, a known weakness of purely local methods. Local methods on the other hand have demonstrated to be very appropriate for the recovering of geometric structure such as image edges. The synthesis of both types of methods is a trend in current research. Variational analysis in particular is an appropriate tool for a unified treatment of local and non-local methods. In this work we propose a general variational framework for the problem of non-local image inpainting, from which several previous inpainting schemes can be derived, in addition to leading to novel ones. We explicitly study some of these, relating them to previous work and showing results on synthetic and real images.

1 Introduction

Image inpainting, also known as *image completion* or *disocclusion*, is an active research area in image processing. The purpose of inpainting is to obtain a visually plausible image interpolation in a region in which data are missing due to damage or occlusion. Usually, to solve this problem, the only available data is the image outside of the region to be inpainted. In addition to its theoretical importance, image inpainting is a very important problem due to its applications to image and video editing and restoration.

Inpainting methods found in the literature can be classified into two groups: *geometry-* and *texture-oriented* methods.

Geometry-oriented methods. Images are modeled as functions with some degree of smoothness (expressed for instance in terms of the curvature of the level lines or the total variation of the image), and the interpolation is performed by continuing and imposing this model inside the inpainting domain. This has been performed either using variational techniques, as for instance in [3, 11, 12, 19, 28, 29], or with PDEs [4, 7, 36]. These methods show a good performance in propagating smooth level lines or gradients. However they fail in the presence of texture. This is often referred to as *structure* or *cartoon* inpainting.

Texture-oriented methods. Texture-oriented inpainting was born as an application of texture synthesis, e.g., [18, 21]. Its recent development was triggered in part by [18, 37] using non-parametric sampling techniques. In these works texture is modeled as a two dimensional probabilistic *graphical model*, in which the

value of each pixel is conditioned by its neighborhood. These approaches rely directly on a sample of the desired texture to perform the synthesis.

In practice these methods work progressively by expanding a region of synthesized texture. The value for a target pixel x is copied from the center of a square patch in the sample image, chosen among those that best match the available portion of the patch centered at x. Levina and Bickel [26] recently provided a probabilistic theoretical justification for this strategy.

This method (as explained above or with some modifications) has been extensively used for inpainting [5, 6, 14, 17, 18, 31]. As opposed to geometry-oriented inpainting, these so-called *exemplar-based* approaches, are *non-local*: To determine the value at x, the whole image may be scanned in the search of a matching patch.

Since these texture approaches are *greedy* procedures (each hole pixel is visited only once), the results are very sensitive to the order in which pixels are processed [14]. This issue was addressed in [24, 38] where the inpainting problem is stated as the optimization of an energy derived from probabilistic graphical models (see also [25]).

A variational justification for texture-based methods was presented in [16], where the inpainting problem is reformulated as that of finding a *correspondence map* $\Gamma : O \to O^c$, O being the inpainting domain and O^c its complement w.r.t. the image domain. Denoting the image by u, the inpainted value at position $x \in O$ is then given by $u(x) = u(\Gamma(x))$, $\Gamma(\cdot)$ being the correspondence map. The authors proposed a continuous energy functional in which the unknown is the correspondence map itself:

$$E(\Gamma) = \int_O \int_{\Omega_p} (u(\Gamma(x - y)) - u(\Gamma(x) - y))^2 \mathrm{d}y \mathrm{d}x,$$

where Ω_p is the patch domain (centered at $(0,0)$). Thus Γ should map a pixel x and its neighbors in such a way that the resulting patch is close to the one centered at $\Gamma(x)$. This model has been the subject of further (theoretical) analysis by Aujol *et al.*[1].

A different variational model was presented in [32]. Images are modeled as ensembles of patches on a given patch manifold. For inpainting, the patch manifold can be learned from the set of patches in the hole's complement. The method is iterative, with each iteration having two steps. First, the patches in the hole are projected onto the manifold. Since this is done for each patch independently, the projected patches are not necessarily coherent with each other, *i.e.*overlapping patches may have different values in the overlap region. Therefore, in the second step, an image is computed by averaging the patches in the ensemble.

Exemplar-based methods provide impressive results in recovering textures and repetitive structures. However, their ability to recreate the geometry without any example is limited and not well understood. Therefore, different strategies have been proposed which combine geometry and texture inpainting [5, 10, 17, 23]. These methods usually decompose the image in some sort of structure and texture components. Structure is reconstructed using some geometry-oriented scheme, and this is used to guide the texture inpainting.

Contributions of this work. Despite these combined methods, geometry and texture inpainting are still quite separate fields, each one with its own analy-

sis and implementation tools. Variational models as the one introduced in this paper can provide common tools allowing a unified treatment of both trends. We therefore propose a variational framework for non-local image inpainting as a contribution to the modeling and analysis of texture-oriented methods. Our formulation is rather general and different inpainting schemes can be derived naturally from it, via the selection of the appropriate patch metric.

In the present work we study three of them, *patch NL-means*, *-medians*, and *-Poisson*. The former is related to the method of [38] and can be interpreted in terms of the *mean shift* [13] and the manifold models of [32]. The other schemes are, to the best of our knowledge novel. The latter imposes coherence of the gradients, in addition to that of the gray levels, which implies a smoother continuation of the information across the boundary and inside the inpainting domain, thus acting as a basic local regularization.

Our work is related to recent variational formulations of non-local denoising ([2, 9]) by Gilboa and Osher [20]. The image redundancy and self-similarity (measured as patch similarity) is encoded by a non-local weight function $w : O \times O^c \to \mathbb{R}$. This function serves as a fuzzy correspondence, and differs from the works [1, 16], although a (eventually multivalued) correspondence map can be approximated as a limit of our model.

Notation. Images are denoted as functions $u : \Omega \to \mathbb{R}$, where Ω denotes the image domain, usually a rectangle in $\mathbb{R}^2$. Pixel positions are denoted by x, x', z, z' or y, the latter for positions inside the patch. A patch of u centered at x is denoted by $p_u(x) = p_u(x, \cdot) : \Omega_p \to \mathbb{R}$, where Ω_p is a rectangle centered at $(0, 0)$. The patch is defined by $p_u(x, y) = u(x + y)$, with $y \in \Omega_p$. $O \subset \Omega$ is the hole or inpainting domain, and $O^c = \Omega \setminus O$. We still denote by u the part of the image u inside the hole, while $\hat{u}$ is the part of u in O^c: $\hat{u} = u|_{O^c}$. Additional notation will be introduced in the text.

2 Variational framework

Our variational framework is inspired by the following non-local functional

$$F_w(u) = \int_O \int_{O^c} w(x, x')(u(x) - \hat{u}(x'))^2 \mathrm{d}x' \mathrm{d}x. \tag{1}$$

The weight function $w : O \times O^c \to \mathbb{R}^+$ measures the similarity between patches centered in the inpainting domain and in its complement. *Gaussian* weights are commonly used, given by $w(x, x') = \exp\left(-\frac{1}{h}\|p_u(x) - p_{\hat{u}}(x')\|^2\right)$, where $\|\cdot\|$ is a weighted L_2 norm in the space of patches and h is the scale. A similar functional was proposed in [20] as a non-local regularization energy in the context of image denoising which models the non-local means filter [2, 9] (see [35] for a different model of non-local means). An extension to super-resolution is presented in [34].

In [20] the weights are considered known and remain fixed through all the iterations. While this might be appropriate in applications where the weights can be estimated from the noisy image, in the image inpainting scenario here addressed, weights are not available and have to be inferred together with the image ([33, 34]). One of the novelties of the proposed framework is the inclusion of adaptive weights in a variational setting.

For this reason, we will consider the weight function w as an additional unkown. Instead of prescribing explicitly the Gaussian functional dependence of w w.r.t. u, we will do it implicitly, as a component of the optimization process. In doing so, we obtain a simpler functional, avoiding to deal with the complex, non-linear dependence between w and u. In our formulation, $w(x, \cdot)$ is a probability density function, $\int_{O^c} w(x, x')\mathrm{d}x' = 1$, and can be seen as a relaxation of the one-to-one correspondence map of [1, 16], providing a fuzzy correspondence between each $x \in O$ and the complement of the inpainting domain.

In this setting, we propose an energy which contains two terms, one of them is inspired by (1) and measures the coherence between the pixels in O and those in O^c, for a given similarity weight function w. This permits the estimation of the image u from the weights w. The second term allows us to compute the weights given the image. The complete proposed functional is

$$E(u, w) = \frac{1}{h}\widetilde{F}_w(u) - \int_{\widetilde{O}} H_w(x)\mathrm{d}x, \tag{2}$$

where

$$\widetilde{F}_w(u) = \int_{\widetilde{O}} \int_{\widetilde{O}^c} w(x, x')\|p_u(x) - p_{\hat{u}}(x')\|_{a,\varphi}\mathrm{d}x'\mathrm{d}x, \tag{3}$$

for a given norm-like function $\|\cdot\|_{a,\varphi}$ between patches, and

$$H_w(x) = -\int_{\widetilde{O}^c} w(x, x') \log w(x, x')\mathrm{d}x',$$

is the entropy of the probability $w(x, \cdot)$.

We take $\widetilde{O}$, the *extended inpainting domain*, as the set of centers of patches that intersect the hole, *i.e.* $\widetilde{O} = O + \Omega_p = \{x \in \Omega : (x + \Omega_p) \cap O \neq \emptyset\}$. Thus, patches $p_{\hat{u}}(x')$ centered in $x' \in \widetilde{O}^c$ are entirely outside O (Figure 1), simplifying the Euler-Lagrange equation for the minimizer. Accordingly, we consider that the weight function w is defined over $\widetilde{O} \times \widetilde{O}^c$ and $\int_{\widetilde{O}^c} w(x, x')\mathrm{d}x' = 1$.

For a simplified presentation, we assume that $\widetilde{O} + \Omega_p \subseteq \Omega$, *i.e.* every pixel in $\widetilde{O}$ supports a patch centered on it and contained in Ω. This is not true if the inpainting domain reaches the boundary of the image, and details on the treatment of this situation are given in Section 5. Analogously, we also shrink $\widetilde{O}^c$ to have $\widetilde{O}^c + \Omega_p \subseteq \Omega$.

Let us now make some additional comments on the functional. The term $(u(x) - \hat{u}(x'))^2$ in F_w, penalizing differences between pixels, is substituted by $\|p_u(x) - p_{\hat{u}}(x')\|_{a,\varphi}$. This has to be understood together with the inclusion of the second term, which integrates the entropy of each probability $w(x, \cdot)$ over $\widetilde{O}$. For a given completion u, and for each $x \in \widetilde{O}$, the optimum weights minimize the mean patch error for $p_u(x)$, given by $\int_{\widetilde{O}^c} w(x, x')\|p_u(x) - p_{\hat{u}}(x')\|_{a,\varphi}\mathrm{d}x'$, while maximizing the entropy. The resulting weights are Gaussian, as can be confirmed easily by derivating the energy. This can be related to the *principle of maximum entropy* [22], widely used for inference of probability distributions. The parameter h controls the trade-off between both terms and is also the scale parameter of the Gaussian weights. Since $w(x, \cdot)$ is a probability, we discard trivial minima of E with $w(x, x') = 0$ everywhere.

The patch norm-like function. Patches are functions defined on Ω_p, and are compared using $\|\cdot\|_{a,\varphi}$. We consider a non-decreasing and continuously differentiable function $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ with $\varphi(0) = 0$ and define $\|\cdot\|_{a,\varphi}$ by

$$\|p\|_{a,\varphi} = \int_{\Omega_p} g_a(y)\varphi(|p(y)|)\mathrm{d}y,$$

where g_a is an *intra-patch* weight function, a Gaussian centered at the origin with standard deviation a. The L_1 and the squared L_2 norms are particular cases of $\|\cdot\|_{a,\varphi}$ when $\varphi(t) = t$ and $\varphi(t) = t^2$, respectively. In Section 3 we consider another norm involving derivatives of the patch. As will be described below, the patch norm determines not only the similarity criterion but also the image synthesis, and thus is a key element in the framework.

2.1 Probabilistic-geometric model interpretation

The proposed model can be written in terms of the generalized Kullback-Leibler divergence [15]. Given two positive and integrable functions p, q defined over a certain measure space $\mathcal{X}$, the generalized Kullback-Leibler divergence is given by: $\mathrm{KL}(p, q) = \int_{\mathcal{X}} p(s) \log\left(\frac{p(s)}{q(s)}\right) \mathrm{d}s - \int_{\mathcal{X}} p(s)\mathrm{d}s + \int_{\mathcal{X}} q(s)\mathrm{d}s$, assuming that the integrals exist. With this notation (and taking into account that $w(x, \cdot)$ is a probability) the functional E can be written as

$$E(u, w) = \int_{\widetilde{O}} \mathrm{KL}\left(w(x, \cdot), r(x, \cdot)\right) \mathrm{d}x - \int_{\widetilde{O}} \int_{\widetilde{O}^c} r(x, x')\mathrm{d}x'\mathrm{d}x,$$

where r is the Gaussian weight function $r(x, x') = \exp\left(-\frac{1}{h}\|p_u(x) - p_{\hat{u}}(x')\|_{a,\varphi}\right)$. The first term integrates the divergence between the functions $w(x, \cdot)$ and $r(x, \cdot)$, for each $x \in \widetilde{O}$. The second term can be interpreted by noticing that

$$\tilde{q}(x) = \int_{\widetilde{O}^c} r(x, x')\mathrm{d}x' \tag{4}$$

is a density estimate (in the patch space) of the set of patches in O^c: The higher the amount of patches in $\widetilde{O}^c$ close to $p_u(x)$ (according to the scale parameter h), the higher the value of $\tilde{q}$.

The minimizers (u^*, w^*) are obtained when $w^*(x, x') = r^*(x, x')/\tilde{q}^*(x)$, (Gaussian weights normalized by (4)), and the patches of the inpainted image are in regions of high density in the patch space. This provides a geometric intuitive interpretation of our variational formulation. The image is considered as an ensemble of overlapping patches. Known patches in $\widetilde{O}^c$ are fixed, forming a patch density model used to estimate the patches in $\widetilde{O}$. The richness of the framework is given in part by the fact that different norms in the patch space induce inpainting schemes of different nature, as we are going to see next.

2.2 Minimization of E

We have formulated the inpainting problem as the constrained optimization

$$(u^*, w^*) = \underset{u,w}{\arg\min}\, E(u, w) \quad \text{subject to} \int_{\widetilde{O}^c} w(x, x')\mathrm{d}x' = 1 \quad \forall x \in \widetilde{O}. \tag{5}$$

To minimize the energy E, we use an alternate coordinate descent algorithm. At each iteration, two optimization steps are solved: The constrained minimization of E with respect to w while keeping u fixed; and the minimization of E with respect to u with w fixed. This procedure yields the following iterative scheme

1. [*Initial Condition*]   Given $u_0(x)$ with $x \in O$.

2. [*Weights Update*] $w_k = \arg\min_w E(u_k, w)$, subject to $\int_{\widetilde{O}} w(x, x')dx' = 1$.

3. [*Image Update*] $u_{k+1} = \arg\min_u E(u, w_k)$.

4. [*Stopping Criterion*] If $\|u_{k+1} - u_k\| > \tau$, go back to step 2.

In the weights update step, the minimization of E w.r.t. w yields:

$$w_k(x, x') = \frac{1}{\tilde{q}(x)} \exp\left(-\frac{1}{h}\|p_{u_k}(x) - p_{\hat{u}}(x')\|_{a,\varphi}\right).$$

The normalizing factor $\tilde{q}(x)$ is the density estimate given by (4), for patch $p_{u_k}(x)$.

The parameter h determines the selectivity of the similarity. If h is large, maximizing the entropy becomes more relevant, yielding weights which are less selective. In the limit, when $h \to \infty$, $w(x, \cdot)$ becomes a uniform distribution over $\widetilde{O}^c$. On the other hand, a small h yields weights which concentrate on the patches close to $p_u(x)$. In fact, as we will mention later on, in the limit as $h \to 0$, $w(x.\cdot)$ can be considered as an approximation to an (eventually multivalued) correspondence.

The image update step deserves more attention and is described next.

Image update step. We now detail the derivation of the image update step for the cases $\varphi(t) = t^2$ and $\varphi(t) = t$. We refer to the resulting algorithms as *patch-wise non-local means* (patch NL-means), and *medians* (patch NL-medians), respectively.

Patch-wise non-local means. If $\varphi(t) = t^2$ the image energy term is quadratic on u, and its minimum for fixed weights w can be computed explicitly leading to a non-local average:

$$u(x) = \frac{1}{C(x)} \int_{\Omega_p} g_a(y) \int_{\widetilde{O}^c} w(x - y, z')\hat{u}(z' + y)\mathrm{d}z'\mathrm{d}y, \qquad (6)$$

for each $x \in O$, where the normalization constant $C(x)$ is given by $C(x) = \int_{\Omega_p} g_a(y)\mathrm{d}y = \mathcal{A}(\Omega_p)$, the area of the patch (measured according to g_a).

Figure 1 explains this equation. The value at x considers all patches containing x. For instance the patch $p_u(x - y)$ covers x, $p_u(x - y, y) = u(x)$. This patch is compared to all patches in the complement, $p_{\hat{u}}(z')$, yielding the weights $w(x - y, z')$. Each of these patches contributes the term $w(x - y, z')\hat{u}(z' + y)$ to the average, *i.e.*its value at position y weighted by $w(x - y, z')$.

Patch-wise non-local medians. We now consider the L_1 norm in the energy E, corresponding to $\varphi(t) = t$. The Euler equation for u, given the weights w, is

$$\delta_u E(u, w)(z) = \int_{\Omega_p} g_a(y) \int_{\widetilde{O}^c} w(z - y, x')\mathrm{sign}[u(z) - \hat{u}(x' + y)]\mathrm{d}x'\mathrm{d}y = 0. \quad (7)$$
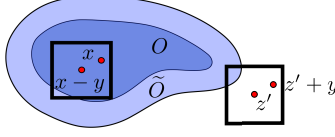
Fig. 1. Patch-wise non-local means inpainting. The value at $x \in O$ is computed using all the patches that overlap x. The patch centered at $x - y$ contributes with the term $w(x - y, z')\hat{u}(z' + y)$ to the average for each $z' \in \widetilde{O}^c$.

The solution of this equation is given by a weighted median of the values outside the hole. We can see this easily by defining $z' = x' + y$ and rewriting Eq. (7) as

$$\delta_u E(u, w)(z) = \int_{O^c} \text{sign}[u(z) - \hat{u}(z')]\rho_z(z')\mathrm{d}z',$$

where

$$\rho_z(z') := \int_{\Omega_p} \chi_{\widetilde{O}^c}(z' - y)g_a(y)w(z - y, z' - y)\mathrm{d}y. \tag{8}$$

For a given $z \in O$, the function $\rho_z : O^c \to \mathbb{R}^+$ weights the contribution of each location z' to the median. The quantity $\rho_z(z')$ is computed by integrating the similarity $w(z - y, z' - y)$ between all patches that overlap z' and those that overlap z in the *same relative position*. It tells us how much evidence there is supporting $u(z')$ as the intensity value for z. The function $\chi_{\widetilde{O}^c}$ takes the value 1 on $\widetilde{O}^c$ and 0 on $\widetilde{O}$.

2.3 Revisiting related work

We conclude this section by further connecting our work with previous art. The method in [38] is closely related to the patch NL-means scheme of Eq. (6). The key difference lies in the underlying theoretical model. The problem is addressed as a MRF, where pixels outside the hole are observable variables, missing pixels in the hole are the parameters, and the hidden variables are given by the correspondence $\Gamma : O \to O^c$, which assigns a patch outside the hole to each x in O. The method can be seen as an approximate EM algorithm for maximizing the log-likelihood w.r.t. the pixels in O, and some approximations have to be taken to make the optimization tractable. Based on heuristics, the authors also propose to use more robust estimators than the mean for the synthesis of pixels. With the framework here proposed, robust estimators (as the median) naturally result from particular choices of the patch norm $\| \cdot \|_{a,\varphi}$.

The patch NL-means algorithm is also related to the interesting manifold image models of [32]. Eq. (6) can be split into two steps which are analog to Peyré's manifold and image projection steps. First, for each patch centered in $\widetilde{O}$ we compute a new patch as a weighted average of all patches in the complement, according to the patch similarity weights $p_u^{MS}(z) := \int_{\widetilde{O}^c} w(z, z')p_{\hat{u}}(z')\mathrm{d}z'$ with $z \in \widetilde{O}$. Doing this for each hole position yields an incoherent ensemble of patches. The image is obtained by averaging these patches: $u(z) = \frac{1}{\mathcal{A}(\Omega_p)} \int_{\Omega_p} p_u^{MS}(z - y, y)\mathrm{d}y$.

We use a density model, instead of the manifold model of [32]. Indeed, $p_u^{MS}(x)$ is the *mean shift* operator applied to $p_u(x)$. It is known that the iteration of this operator corresponds to an adaptive gradient ascent of the Parzen estimate of a PDF [13], which in this case is generated by the set of patches in the complement of the hole. The use of a density model entails some advantages, mainly from the computational point of view, learning a manifold model is computationally costly. Furthermore, the assumption that patches lie on a manifold is questionable (one could think for instance in a *stratification* as a more realistic model), and its dimension is hard to determine for real images.

3   Higher order variational models

The proposed variational framework allows the introduction of derivatives of the image, by considering them in the patch norm used in (3). In this section we study a functional using the L_2 norm of the gradients of the patches,

$$\|p(y)\|_{a,\nabla}^2 = \int_{\Omega_p} g_a(y) \|\nabla p(y)\|_2^2 \mathrm{d}y,$$

where $\| \cdot \|_2$ is the Euclidean norm in $\mathbb{R}^2$. Firstly, the similarity weights are now based on patch gradients, and secondly, the image update step is given by a *non-local* Poisson equation, *i.e.* a Poisson equation with non-local coefficients. The functional is obtained by substituting in (2) the image energy term $\widetilde{F}_w(u) = \int_{\widetilde{O}} \int_{\widetilde{O}^c} w(x,x') \|p_u(x) - p_{\hat{u}}(x')\|_{a,\nabla}^2 \mathrm{d}x' \mathrm{d}x$ (we assume that $u|_{O^c} = \hat{u}$).

The Euler equation w.r.t. u of the resulting functional is

$$\nabla \cdot [C(z)\nabla u(z)] = \nabla \cdot \boldsymbol{v}(z), \tag{9}$$

for all $z \in O$, where $u|_{O^c} = \hat{u}$ and the field $\boldsymbol{v} : O \to \mathbb{R}^2$ is given by

$$\boldsymbol{v}(z) = \int_{\Omega_p} g_a(y) \int_{\widetilde{O}^c} w(z-y,x') \nabla \hat{u}(x'+y) \mathrm{d}x' \mathrm{d}y.$$

The solutions are minimizers of $\int_{\widetilde{O}} C(z) \|\nabla u(z) - \boldsymbol{v}(z)\|_2^2 \mathrm{d}z$ (as before, $C(z) = \mathcal{A}(\Omega_p)$). Therefore, u is computed as the image with the closest gradient (in the L_2 sense) to the *guiding* vector field $\boldsymbol{v}$, which corresponds to a non-local weighted average of the gradients in the complement. The coefficients in the average have exactly the same form as in (6). The only difference is that the patch similarity weights used here are Gaussian weights of the L_2 norm of the gradients. See [30] for further uses of the Poisson equation in image editing.

This energy can be combined with the patch NL-means energy by considering a linear combination of the corresponding image energy terms. The resulting scheme computes the weights based on the image together with its gradient, and updates the image by solving a linear combination of Eqs. (6) and (9).

4   Confidence mask

For large inpainting domains, it is useful to introduce a mask $\kappa : \Omega \to (0,1]$ which assigns a confidence value to each pixel, depending on the certainty of its

information (see also [14, 24]). This will help in guiding the flow of information from the boundary towards the interior of the hole, eliminating some local minima and reducing the effect of the initial condition. The resulting image energy term takes the form

$$\widetilde{F}_w(u) = \int_{\widetilde{O}} \int_{\widetilde{O}^c} \kappa(x) w(x, x') \|p_u(x) - p_{\hat{u}}(x')\|_{a,\varphi} \mathrm{d}x' \mathrm{d}x,$$

where κ modulates the penalization of the incoherences between w and the φ-norm between patches.

The effect of κ on the image update step is easier to visualize on the evidence function ρ_z, Eq. (8). Recall that this function gathers all evidence supporting $u(z')$ as a value for $u(z)$, for each $z' \in O^c$. As in Eq. (8), now taking κ into account, we obtain $\rho_{\kappa,z}(z') = \int_{\Omega_p} \chi_{\widetilde{O}^c}(z' - y) g_a(y) \kappa(z - y) w(z - y, z' - y) \mathrm{d}y$. Thus, the contribution of the patch $p_u(z - y)$ to the evidence function is now weighted by its confidence. Patches with higher confidence will support stronger evidence. In this case the weights are given by $w(x, x') = \frac{1}{\widetilde{q}(x)} \exp\left(-\frac{\kappa(x)}{h} \|p_u(x) - p_{\hat{u}}(x')\|_{a,\varphi}\right)$.

The inclusion of the confidence mask modifies the patch space scale parameter h. If the confidence is high, the effective scale $h/\kappa(x)$ will be lower, thus increasing the selectivity of the similarity measure. If the information at x is uncertain, more patches are considered similar. The same reasoning applies to the patch NL-Poisson energy, with similar modifications to Eq. (9).

5   Experimental results

We tested the proposed methods with gray scale and color images. The energy for the latter can be obtained by considering a norm for color patches that adds the norms of the three scalar components: $\|p_{\overrightarrow{u}}(x)\|_{a,\varphi} = \sum_{i=1}^{3} \|p_{u_i}(x)\|_{a,\varphi}$, where $\overrightarrow{u} : \Omega \to \mathbb{R}^3$ is the color image, and u_i, with $i = 1, 2, 3$, its components (analogously for $\| \cdot \|_{a,\nabla}$). Thus, the weights will take into account the three channels. Given the weights, each channel is updated using the corresponding scheme for scalar images. All channels are updated using the same weights. This scheme can be applied to any Euclidean color space. We show results with RGB and CIE La*b* color spaces.

In our implementation we use a square patch domain Ω_p of side $s \in \mathbb{N}$, with the Gaussian intra-patch weights g_a centered on it. For all experiments we set $s = 3a$ (s should be chosen such that most of the effective support of the Gaussian fits in the patch, we used a smaller s to lower the computational cost). This leaves only two independent parameters, namely, the intra-patch Gaussian width a, and the patch similarity scale h. The former determines the size of the patch, a parameter inherent to all patch-based techniques. It should be large enough so as to allow the identification of the image patterns.

In the limit when $h \to 0$, we compute the weights as $\lim_{h \to 0} w_h(x, x') = \frac{1}{\#n(x)} \delta(x' - n(x))$, where $n(x) \subseteq O^c$ is the set of nearest neighbors of x, $i.e. n(x) = \{x' \in O^c : \|p_u(x) - p_{\hat{u}}(x')\|_{a,\varphi} = \delta_n\}$, where δ_n represents the nearest neighbor

distance. In practice, we assume that $\#n(x) = 1$, *i.e.* the nearest neighbor is unique. The choice of this parameter will be addressed later.

The confidence mask, when used, adds another parameter. We found good results using the following function:

$$\kappa(x) = \begin{cases} (1 - \kappa_0)e^{-\frac{d(x, \partial O)}{\tau_\kappa}} + \kappa_0 & \text{if } x \in O, \\ 1 & \text{if } x \in O^c, \end{cases}$$

which shows an exponential decay w.r.t. the distance to the boundary inside the hole $d(\cdot, \partial O)$, where $\tau_\kappa > 0$ is the decay time and $\kappa_0 > 0$ determines the asymptotic value reached far away from the boundary.

If a patch centered in the inpainting domain does not fit in the image, we mirror the image w.r.t. the boundary to complete the patch. Whenever the hole reaches $\partial \Omega$, the Poisson equation requires a different boundary condition. We have considered Neuman boundary conditions, *i.e.* $\nabla u \cdot \mathbf{n}(x) = 0$, for $x \in O \cap \partial \Omega$, where $\mathbf{n}(x)$ is the normal direction at the boundary. This amounts again to a reflection of u w.r.t. $\partial \Omega$.

The computational cost of each iteration is $\mathcal{O}(\mathcal{A}(O) \times \mathcal{A}(O^c) \times s^2)$. This is typical of non-local methods, and several strategies can be used for speed-up [8, 27].

Figure 2 compares the results of the three methods on a texture with two different mean intensities, darker on the right half of the image. The inpainting domain hides all patches on the boundary between the dark and bright textures. With this we can test the ability of each method to *create* an interface between both regions. Situations like these are common in real inpainting problems due for instance to shadows. Moreover, when inpainting non-regular textures, a good completion may not be possible just by copying, and creating new patterns becomes necessary (see Figure 3).

We have also added Gaussian noise with standard deviation $\sigma = 10$ to show the influence of the patch space scale parameter h. Figure 2 shows two results for each scheme, one with $h = 0$, and the other with a higher h, chosen empirically for each method.
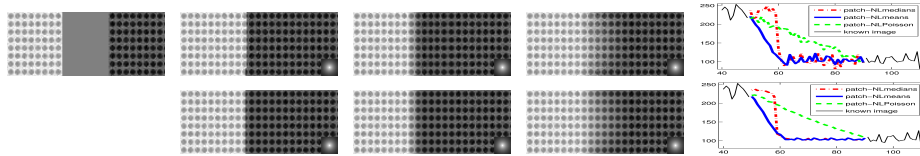


Fig. 2. Results with $s = 15$ and $a = 5$. The first four columns correspond to the initial condition, result of path NL-medians, -means and -Poisson. Top row, $h = 0$, bottom row $h = 0.01$, $h = 0.05$ and $h = 0.04$, respectively. The used intra-patch weight kernel g_a is shown in each figure on the bottom right. The fifth column shows the value of the images for a horizontal line going between the circles.

The rightmost column in Figure 2 plots the image values for a horizontal line between the circles. The interpolation done by the patch NL-Poisson method is

linear, since this is a solution of the homogeneous Poisson equation. The profile shown by patch NL-means shows a smooth transition when both regions meet, whereas the use of the L_1 norm yields a sharp edge. The results using a higher h show some denoising, since for larger h, more patches are regarded as similar to each patch in $\widetilde{O}$ and each pixel value is synthesized from more complement pixels. For inpainting of noiseless images, we use $h = 0$.

The top row of Figure 3 shows results with the three schemes for a non-regular texture. The result with patch NL-medians shows image regions copied without any modification. The boundaries between these regions are determined so that each patch on the boundary is close to some patch outside the hole. This does not always yields a smooth transition. Copied patterns can also be seen in the result of the patch NL-means, but the copies are less sharp, and the discontinuities less noticeable. The patch NL-Poisson shows a better continuation of the color at the boundary of the hole. However the inpainted structure looks too blurry (zoom on the pdf file for details).
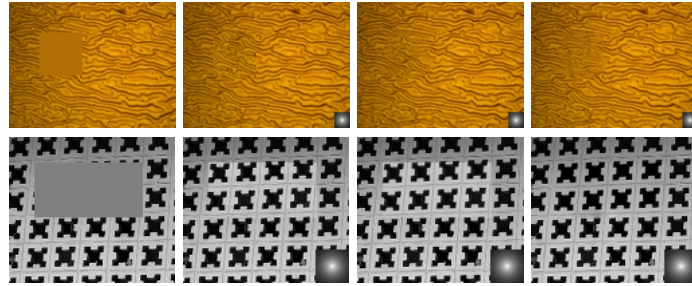
Fig. 3. From left to right: Initial condition, result of path NL-medians, -means and -Poisson. Top: Results with $s = 15$, $a = 5$ and $h = 0$, using the CIE La*b* color space. Bottom: Results with $s = 25$, $a = 8$ and $h = 0$. Gray scale image.

The bottom row of Figure 3 depicts results on a regular texture. The regularity of the texture hides the blurring effects of the L_2 metrics (both on the image and gradients). At a stable state, all patches overlapping on a pixel will agree on its value. Notice that in this case, the patch NL-Poisson is able to reconstruct the illumination gradient of the image. This is imposed to the solution of the Poisson equation by the boundary conditions. In addition to the non-local inpainting, this scheme performs also a local interpolation based on the hole's boundary. Instead, the other methods copied the information from the bottom of the image, generating a discontinuity at the top.

The results shown in Figure 4, were computed using a confidence mask shown at the leftmost column. In both cases the patch NL-medians scheme yields the best results, comparable to state of the art (see results in [17, 24] for results on the same images). The images look as a composition of copied regions (although some parts in the *elephant* image seem new). Again the patch NL-means shows blurred results, most noticeable for *elephant* due to the non-regularity of the textures. The patch NL-Poisson fails with this image. In this case the gradient

is not a good feature for computing the weights. However, the results for Figure 4 are still reasonable, it did recover the structure of the image. Due to the averaging of gradients, when overlapping patches do not agree on the value of a pixel, lower gradients may appear. These generate phantom edges surrounding the cylinders. Presumably a more robust estimation of the gradient would not have this problem. We are currently developing a scheme using the L_1 norm between the gradients of patches.

The initial condition for *elephant* is the original image, whereas for *cylinders* the hole was filled with 128 as constant gray level. A confidence mask κ with low confidence inside the hole helps in diminishing the influence of the initialization. Further results are available at: `http://gpi.upf.edu/static/vnli`.

Fig. 4. From left to right, inpainting domain with confidence mask, result of path NL-medians, -means and -Poisson (the latter only for the first row). Top: *cylinders*- Results with $s = 27$, $a = 9$ and $h = 0$ for patch NL-medians and -means and $s = 33$, $a = 1$ and $h = 0$ patch NL-Poisson. Bottom: *elepahnt*- Results with $s = 19$, $a = 6$ and $h = 0$. The parameters of the confidence mask are $\tau_0 = 5$ and $\kappa_0 = 0.4$ in all cases except for the patch NL-medians with the bottom image, in which we set $\kappa_0 = 0.1$. Results using RGB. Please refer to [17, 24] for other results on the same images.

6   Conclusions and future work

In this work we present a variational framework for non-local image inpainting. The proposed energy lends itself to intuitive interpretations, and contrary to previous variational models, allows a straightforward minimization using a coordinate descent scheme. Beyond the specific application of inpainting, this framework provides also a sound variational modelling of non-local regularizers with *adaptive* weights, extending previous work in which the weights are considered known and fixed.

Starting from this model, we derived three different inpainting schemes, each one corresponding to a different norm measuring the distance between patches. We showed results on synthetic and natural images comparing their properties.

The derived *patch NL-means* provides a variational interpretation of the methods proposed by [32, 38]. The *patch NL-medians* is the one that showed

the best overall performance, comparable with the state of the art. The results obtained suggest a possible relation with the piece-wise traslation models of [1, 16]. The *patch NL-Poisson* presents two interesting features. First, the similarity weights are computed based on the gradients, allowing the transference of information from areas with different intensity level. Second, the image completion is the result of a Poisson equation, thus incorporating some basic local regularization, meaning the completion must be differentiable and its gradient squared integrable. This traduces in a local interpolation based on the image values at the boundary of the inpainting domain. This method performs well for structured textures, but fails for non-regular textures.

We are currently exploring several additional aspects of this framework, including the use of robust norms in the general φ setting and the L_1 norm between patch gradients.

References

1. J.-F. Aujol, S. Ladjal, and S. Masnou. Exemplar-based inpainting from a variational point of view. *Submitted*, 2008.
2. S.P. Awate and R.T. Whitaker. Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Trans. on PAMI*, 28(3):364–376, 2006.
3. C. Ballester, M. Bertalmío, V. Caselles, G. Sapiro, and J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. on IP*, 10(8):1200–11, 2001.
4. M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proc. of SIGGRAPH*, 2000.
5. M. Bertalmío, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture inpainting. *IEEE Trans. on Image Processing*, 12(8):882–89, 2003.
6. R. Bornard, E. Lecan, L. Laborelli, and J.-H. Chenot. Missing data correction in still images and image sequences. In *Proc. ACM Int. Conf. on Multimedia*, 2002.
7. F. Bornemann and T. März. Fast image inpainting based on coherence transport. *J. of Math. Imag. and Vis.*, 28(3):259–78, 2007.
8. T. Brox, O. Kleinschmidt, and D. Cremers. Efficient nonlocal means for denoising of textural patterns. *IEEE Trans. on IP*, 17(7):1057–92, 2008.
9. A. Buades, B. Coll, and J.M. Morel. A non local algorithm for image denoising. In *Proc. of the IEEE Conf. on CVPR*, volume 2, pages 60–65, 2005.
10. F. Cao, Y. Gousseau, S. Masnou, and P. Pérez. Geometrically-guided exemplar-based inpainting. *Submitted*, 2008.
11. T. Chan, S. H. Kang, and J. H. Shen. Euler's elastica and curvature based inpaintings. *SIAM J. App. Math.*, 63(2):564–92, 2002.
12. T. Chan and J. H. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. App. Math.*, 62(3):1019–43, 2001.
13. Y. Cheng. Mean shift, mode seeking and clustering. *IEEE Trans. on PAMI*, 17(8):790–99, 1995.
14. A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based inpainting. *IEEE Trans. on IP*, 13(9):1200–1212, 2004.

15. I. Csiszár. Axiomatic characterizations of information measures. *Entropy*, 10(3):261–73, 2008.
16. L. Demanet, B. Song, and T. Chan. Image inpainting by correspondence maps: a deterministic approach. Technical report, UCLA, 2003.
17. I. Drori, D. Cohen-Or, and H. Yeshurun. Fragment-based image completion. *ACM Trans. on Graphics. Special issue: Proc. of ACM SIGGRAPH*, 22(3):303–12, 2003.
18. A.A. Efros and T.K. Leung. Texture synthesis by non-parametric sampling. In *Proc. of the IEEE Conf. on CVPR*, pages 1033–1038, September 1999.
19. S. Esedoglu and J. H. Shen. Digital image inpainting by the Mumford-Shah-Euler image model. *European J. App. Math.*, 13:353–70, 2002.
20. G. Gilboa and S.J. Osher. Nonlocal linear image regularization and supervised segmentation. *SIAM Mult. Mod. and Sim.*, 6(2):595–630, 2007.
21. H. Igehy and L. Pereira. Image replacement through texture synthesis. In *Proc. of the IEEE Conf. on CVPR*, October 1997.
22. E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–30, 1957.
23. J. Jia and C.-K. Tang. Inference of segmented color and texture description by tensor voting. *IEEE Trans. on PAMI*, 26(6):771–86, 2004.
24. N. Komodakis and G. Tziritas. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *IEEE Trans. on IP*, 16(11):2649–61, 2007.
25. A. Levin, A. Zomet, and Y. Weiss. Learning how to inpaint from global image statistics. In *Proc. of ICCV*, 2003.
26. E. Levina and P. Bickel. Texture synthesis and non-parametric resampling of random fields. *Annals of Statistics*, 34(4), 2006.
27. M. Mahmoudi and G. Sapiro. Fast image and video denoising via nonlocal means of similar neighborhoods. *Signal Processing Letters*, 12(12):839–42, 2005.
28. S. Masnou. Disocclusion: a variational approach using level lines. *IEEE Trans. on IP*, 11(2):68–76, 2002.
29. S. Masnou and J.-M. Morel. Level lines based disocclusion. In *Proc. of IEEE ICIP*, 1998.
30. P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. In *Proc. of SIGGRAPH*, 2003.
31. P. Pérez, M. Gangnet, and A. Blake. PatchWorks: Example-based region tiling for image editing. Technical report, Microsoft Research, 2004.
32. G. Peyré. Manifold models for signals and images. *Comp. Vis. and Im. Unders.*, 113(2):249–260, 2009.
33. G. Peyré, S. Bougleux, and L. Cohen. Non-local regularization of inverse problems. In *Proc. of ECCV*, 2008.
34. M. Protter, M. Elad, H. Takeda, and P. Milanfar. Generalizing the non-local-means to super-resolution reconstruction. *IEEE Trans. on IP*, 18(1):36–51, 2009.
35. A. Singer, Y. Shkolnisky, and B. Nadler. Diffusion interpretation of non-local neighborhood filters for signal denoising. *SIAM J. on Imag. Sci.*, 2(1):118–139, 2009.
36. D. Tschumperlé and R. Deriche. Vector-valued image regularization with PDE's: a common framework for different applications. *IEEE Trans. on PAMI*, 27(4), 2005.
37. L.-Y. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In *Proc. of the SIGGRAPH*, 2000.
38. Y. Wexler, E. Shechtman, and M. Irani. Space-time completion of video. *IEEE Trans. on PAMI*, 29(3):463–476, 2007.