

Exemplar-Based Image Inpainting using an Affine Invariant Similarity Measure

Vadim Fedorov, Pablo Arias, Gabriele Facciolo, Coloma Ballester

▶ **To cite this version:**

Vadim Fedorov, Pablo Arias, Gabriele Facciolo, Coloma Ballester. Exemplar-Based Image Inpainting using an Affine Invariant Similarity Measure. Computer Vision, Imaging and Computer Graphics Theory and Applications 11th International Joint Conference, VISIGRAPP 2016, Rome, Italy, February 27 – 29, 2016, Revised Selected Papers , 2017, Communications in Computer and Information Science book series, 10.1007/978-3-319-64870-5_22 . hal-01829625

HAL Id: hal-01829625

https://enpc.hal.science/hal-01829625

Submitted on 4 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exemplar-Based Image Inpainting using an Affine Invariant Similarity Measure

Vadim Fedorov[1], Pablo Arias[2], Gabriele Facciolo[3], and Coloma Ballester[1]

[1] University Pompeu Fabra, Barcelona, Spain,
{vadim.fedorov, coloma.ballester}@upf.edu;
[2] CMLA, ENS Cachan, Cachan, France,
pablo.arias@cmla.ens-cachan.fr;
[3] LIGM, UMR 8049, École des Ponts, UPE, France,
facciolo@cmla.ens-cachan.fr

Abstract. Patch-based approaches are used in state-of-the-art methods for image inpainting. This paper presents a new method for exemplar-based image inpainting using transformed patches. The transformation is determined for each patch in a fully automatic way from a surrounding texture content. We build upon a recent affine invariant patch similarity measure that performs an appropriate patch comparison by automatically adapting the size and shape of the patches. As a consequence, it intrinsically extends the set of available source patches to copy information from. We incorporate this measure into a variational formulation for inpainting and present a numerical algorithm for optimizing it. We show that our method can be applied to complete a perspectively distorted texture as well as to automatically inpaint one view of a scene using other view of the same scene as a source. We present experimental results both for gray and color images, and a comparison with some exemplar-based image inpainting methods.

1 Introduction

Image inpainting, also known as image completion, disocclusion or object removal, refers to the recovery of occluded, missing or corrupted parts of an image in a given region so that the reconstructed image looks natural. It has become a key tool for digital photography and movie post-production where it is used, for example, to eliminate unwanted objects that may be unavoidable during filming.

Automatic image inpainting is a challenging task that has received significant attention in recent years from the image processing, computer vision, and graphics communities. Remarkable progress has been achieved with the advent of exemplar-based methods, which exploit the self-similarity of natural images by assuming that the missing information can be found elsewhere outside the inpainting domain. Roughly speaking, these methods work by copying patches taken from the known part of the image and pasting them smartly in the inpainting domain. These methods can obtain impressive results but many of them rely on the assumption



Fig. 1. Self-similarity under different distortions. On the left: two views of the same scene related by a projective transformation. On the right: self-similar texture underwent a severe fish-eye lens distortion.

that the required information can be copied as it is, without any transformations. Therefore, applicability of such methods is limited to the scenes in which objects are in a fronto-parallel position with respect to the camera.

In the image formation process, textured objects may appear distorted by some complex transformation (see Fig. 1). This is a pervasive phenomenon in our daily life. In fact, any person can mentally fill-in occluded parts of an image, even if the missing information is available to them under a different perspective. Our brain is able to appropriately transform the available information to match the perspective of the occluded region. For instance, in Fig. 1 one can easily infer what is hidden behind the red rectangle in the graffiti scene on the left, or use the non-trivially distorted context in the right image to fill-in the hole.

In this work we address this issue by transforming known patches before pasting them in the inpainting domain. The transformation is determined for each patch in a fully automatic way. Moreover, instead of searching for an appropriate transformation in a high dimensional space, our approach allows us to determine a single transformation from the surrounding texture content. As opposed to some previous works which only consider rotations and scalings, we can handle full affinities, which in principle extends the applicability of the method to any transformation that can be locally approximated by an affinity, such as perspective distortion.

We follow the approach recently proposed in [16], where affine covariant structure tensor fields computed a priori in each image are used to define an *affine invariant similarity measure* between patches. We incorporate this measure into a variational inpainting formulation. The affine covariant structure tensors determine elliptical patches at each location of the image domain. Due to the affine covariance property of the structure tensors, these patches transform appropriately when computed on an affinely transformed version of the image. Fig. 2 illustrates the patches defined by the affine covariant structure tensors of [16], computed for a set of corresponding points in two images related by a homography. Note

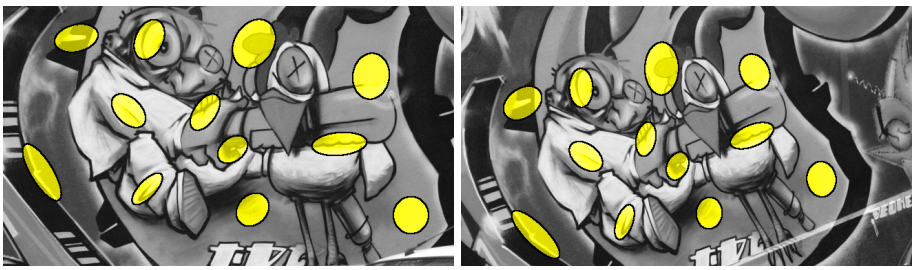


Fig. 2. Affine covariant neighborhoods (patches) computed at corresponding points in two images taken from different viewpoints. Despite the change in appearance, patches capture the same visual information.

that even though the transformation is not an affinity, the patches still match, since a homography can be locally approximated by an affinity.
The paper is organized as follows. In Section 2 we review the related work. Then in Section 3 we summarize the results of [16] which motivates the definition of the similarity measure that we use. Section 4 is devoted to the proposed inpainting method. Section 5 covers some implementation details. In Section 6 we present experiments, asserting the validity of our theoretical approach, together with a comparison with well-known exemplar based methods. Finally, Section 7 concludes the paper.

2 Related work

Most inpainting methods found in the literature can be classified into two groups: geometry- and texture-oriented, depending on how they characterize the redundancy of the image.
The geometry-oriented methods formulate the inpainting problem as a boundary value problem and the images are modeled as functions with some degree of smoothness expressed, for instance, in terms of the curvature of the level lines [29, 3, 28, 10], with propagation PDE's [8], or as the total variation of the image [11]. These methods perform well in propagating smooth level lines or gradients, but fail in the presence of texture or big inpainting domains.
Texture-oriented (also called exemplar-based) methods were initiated by the work of Efros and Leung [15] on texture synthesis. In that work the idea of self-similarity is exploited for direct and non-parametric sampling of the desired texture. The self-similarity prior is one of the most influential ideas underlying the recent progress in image processing and has been effectively used for different image processing and computer vision tasks, such as denoising and other inverse problems [17, 9, 19, 34, 35]. It has also found its application to inpainting: the value of each target pixel x in the inpainting domain can be sampled from the known part of the image or even from a vast database of images [22].

The exemplar-based approach to inpainting has been intensively studied [13, 12, 38, 25, 2, 1]. However, many such methods are based on the assumption that the information necessary to complete the image is available elsewhere and can be copied *without any modification but a translation*.

Some works consider a broader family of transformations. Drori et al. [14] used heuristic criteria to vary the scale of patches. Mansfield et al. [27] and Barnes et al. [7] extended the space of available patches by testing possible rotations and scales of a source patch. The search in the space of available patches is usually performed by a collaborative random search. However, this implies that for each query patch, the position of the matching patch as well as the parameters of the transformation (scale, rotation angle, tilt, etc) must be determined. The high dimensionality of the parameter space makes the search problem very computationally expensive and the excessive variability of candidates may lead to unstable results. In order to restrict the search space, the authors of [10] propose to combine an exemplar-based approach that includes all rotated patches, with a geometric guide computed by minimizing Euler's elastica of contrasted level lines in the inpainted region.

Several authors [33, 23] have addressed this issue using some user interaction to guide the search process. For example, the user provides information about the symmetries in the image, or specifies 3D planes which are then used for rectification and the rectified planes in turn are used to look for correspondences. Recently, Huang et al. [24] proposed a method for automatic guidance that searches for appropriately transformed source patches. It starts by detecting planes and estimating their projection parameters, which are then used to transform the patches. This allows one to handle perspective transformations, in situations when representative planes can be detected.

Most of those works use a similarity measure, either explicitly or implicitly, to compute a matching cost between patches. We propose to use an affine invariant similarity measure which automatically distorts the patches being compared [16]. Our method considers a rich patch space that includes all affine-transformed patches, furthermore, for each pair of patches the transformations are uniquely determined using the image content. This effectively limits the search space, making the method more stable. Since the patch distortions depend on the texture content of the image, our technique is related in that sense to a shape-from-texture approach [20, 21, 5].

In this paper we extend the variational framework described in [38, 25, 1] proposing a new energy and an optimization algorithm for affine invariant exemplar-based inpainting.

Let us finally note that [36] proposed a self-similarity measure for image inpainting, comparing dense SIFT descriptors on square patches of a fixed size. However, the method is not fully affine invariant, for example, neither the dense SIFT descriptors nor the square patches are scale invariant. Several authors have addressed the affine distortion and affine invariance problem in other contexts such as image comparison [31], object recognition [30], and stereo [18].

3 An affine invariant similarity measure

Non-local self-similarity is an accepted prior for natural images. To formalize it, a patch similarity or comparison measure is needed. Let us consider the general problem of comparing patches on two images $u : \Omega_u \to \mathbb{R}$ and $v : \Omega_v \to \mathbb{R}$, $\Omega_u, \Omega_v \subseteq \mathbb{R}^2$ (for simplicity, we can assume the image domains to be $\mathbb{R}^2$). A widely used comparison measure between two patches centered respectively at x and y is the weighted squared Euclidean distance

$$\mathcal{D}(t, x, y) = \int_{\mathbb{R}^2} g_t(h)(u(x+h) - v(y+h))^2 \, dh, \tag{1}$$

where g_t is a given window that we assume to be Gaussian of variance t. The Gaussian g_t represents a weighted characteristic function of both patches being compared and determines the size of the patches or, in other words, the scale.

In many occasions, similar patches exist in the image but have undergone a transformation, for example due to a different position with respect to the camera. The Euclidean distance is not appropriate for detecting these similarities. Consider for example a simple case in which v is a rotated version of image u. If the rotation is known, we should use the Euclidean distance between patches in u and rotated patches in v, namely

$$\mathcal{D}^R(t, x, y) = \int_{\mathbb{R}^2} g_t(h)(u(x+h) - v(y+Rh))^2 \, dh. \tag{2}$$

In a more realistic scenario, one does not know the appropriate transformation that matches both patches being compared and even whether it exists. Some previous works addressed this issue by searching among all possible transformations [7, 27] which involves probing of all the parameters (scale, rotation angle, etc). The high dimensionality of the parameter space makes the problem very difficult. In this paper we use an affine invariant similarity measure, introduced in [16], that automatically deduces this transformation from the local texture context.

The similarity measure defined in [16] is based on affine covariant structure tensor fields *a priori* computed in each image. It was derived as an approximation to a more general framework introduced in [4], where similarity measures between images on Riemannian manifolds are studied.

In the remainder of this section we present an alternative, self-contained overview of this similarity measure. We first briefly discuss the concept of affine covariant structure tensors. Then we describe an algorithm to compute them. Finally, we show how they are used to define the affine invariant similarity measure and establish the relation between our derivation and the theory of [4, 16].

3.1 Affine Covariant Structure Tensors

Given a real-valued image u, we consider an image-dependent structure tensor field T_u as a function that associates a structure tensor (a symmetric, positive semi-definite 2×2 matrix) to each point x in the image

domain. As before, for simplicity we assume the image domain to be $\mathbb{R}^2$. The structure tensor field is said to be *affine covariant* if, for any affinity A,

$$T_{u_A}(x) = A^T T_u(Ax)A, \tag{3}$$

where $u_A(x) := u(Ax)$ denotes the affinely transformed version of u. Given a structure tensor $T_u(x)$ we can associate with it an elliptical region of "radius" r centered at x

$$B_u(x,r) = \{y : \langle T_u(x)(y-x), (y-x)\rangle \leq r^2\}. \tag{4}$$

When the structure tensor is affine covariant, we have that

$$AB_{u_A}(x,r) = B_u(Ax,r).$$

This implies that the structure tensors can be used to define regions that transform appropriately via an affinity (Fig. 2).
As shown in [16], given two affine covariant structure tensors we can extract the affine transformation between the corresponding elliptical patches up to some rotation. Indeed, for any affine transformation A, there exists an orthogonal matrix R such that

$$A = T_u(Ax)^{-\frac{1}{2}} R T_{u_A}(x)^{\frac{1}{2}}. \tag{5}$$

This last equation provides an intuitive geometric relationship between the structure tensors, the associated elliptical regions and the affinity. Consider a point x and the corresponding affine covariant elliptic neighborhood $B_{u_A}(x)$. Mapping $B_{u_A}(x)$ by the affinity yields $B_u(Ax)$. The application of A can be decomposed in three steps. First, applying $T_{u_A}(x)^{\frac{1}{2}}$, we transform $B_{u_A}(x)$ into a circle or radius r. We refer to the resulting patch as a *normalized patch*. Then, a rotation is applied to the normalized patch. Finally, $T_u(Ax)^{-\frac{1}{2}}$ maps the rotated normalized patch to the elliptical neighborhood $B_u(Ax)$.
To fully determine the affinity A, one needs to find the rotation R. Any rotation would yield an affinity that maps the elliptical neighborhood associated with T_{u_A} at x to the one associated with T_u at Ax. For a wrong value of the rotation, the image content inside both neighborhoods will not match. Therefore, the right value for the rotation can be computed by aligning the image content of both patches. For this aim, we decompose the rotation as

$$R = R_u(Ax)R_{u_A}^{-1}(x), \tag{6}$$

where $R_u(Ax)$ and $R_{u_A}(x)$ are estimated from the image content in the patches. In practice, we calculate them by aligning the dominant orientation of the normalized patches to the horizontal axis. To compute the dominant orientation we use histograms of gradient orientations as in the SIFT descriptors [26].

3.2 Computation of Affine Covariant Structure Tensors

The following iterative algorithm introduced in [16] allows us to compute a dense field of affine covariant structure tensors and the associated

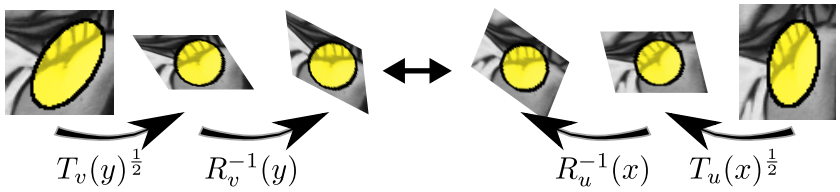


Fig. 3. An affine invariant patch comparison which is achieved by normalizing the patches to discs and aligning them with suitable rotations.

neighborhoods on an image u:

$$T_u^{(k)}(x) = \frac{\int_{B_u^{(k-1)}(x,r)} Du(y) \otimes Du(y)\, dy}{\text{Area}(B_u^{(k-1)}(x,r))}, \tag{7}$$

where u is the given image and $B_u^{(k)}$ is the elliptical patch related to $T_u^{(k)}$, defined by

$$B_u^{(k)}(x,r) = \{y : \langle T_u^{(k)}(x)(y-x), (y-x)\rangle \leq r^2\} \tag{8}$$

for $k \geq 1$, and

$$B_u^{(0)}(x,r) = \{y : |Du(x)(y-x)| \leq r\} \tag{9}$$

for $k = 0$.
Throughout this paper we follow the notation of [16] and denote by $T_u(x)$ the affine covariant structure tensor $T_u^{(k)}(x)$ for a fixed value of k ($k = 30$ in all the experiments) and a given value of r ($r > 0$ is a free parameter which is in range $[250, 350]$ in our experiments). We denote by $B_u(x)$ the affine covariant neighborhood $B_u^{(k)}(x,r)$.
Notice that the structure tensor (7) is guaranteed to be affine covariant at any iteration of the scheme, therefore, the purpose of it is not to enforce affine covariance property, but rather to diminish dependency on the very first iteration.

3.3 An Affine invariant Patch Similarity

Previously in this section we were considering two images u and u_A, related by a global affinity. For the patch comparison problem we can generalize our reasoning and consider two arbitrary images u and v. Let x and y be two given points in images u and v respectively. The structure tensors $T_u(x)$ and $T_v(y)$ define elliptical patches $B_{T_u}(x)$ and $B_{T_v}(y)$ around these points. In order to compare the patches, equations (5) and (6) suggest the following mapping between the elliptical patches:

$$P(x,y) = T_v(y)^{-\frac{1}{2}} R_v(y) R_u^{-1}(x) T_u(x)^{\frac{1}{2}}. \tag{10}$$

We can interpret $P(x, y)$ as an affinity, mapping the elliptical patch associated with $T_u(x)$ into the one associated with $T_v(y)$. If u in the vicinity of x is an affinely transformed version of v in the vicinity of y, then $P(x, y)$ recovers the true affinity. An affine invariant patch similarity measure could be built by computing the distance between the elliptical patch at y and the elliptical patch at x transformed by $P(x, y)$. In practice, it is more convenient to transform both elliptical patches to the circle of radius r (Fig. 3) and compare the aligned normalized patches:

$$\mathcal{D}^{\mathrm{a}}(t, x, y) = \int_{\Delta_t} g_t(h) \left(u(x + T_u^{-\frac{1}{2}} R_u(x)h) - v(y + T_v^{-\frac{1}{2}} R_v(y)h) \right)^2 dh, \quad (11)$$

where Δ_t is a disc centered at the origin with radius proportional to the scale $t > 0$ and big enough to contain the effective support of the weighting function g_t. The distance $\mathcal{D}^{\mathrm{a}}$ provides an affine invariant distance between the patches $p_u(x)$ and $p_v(y)$. Here the patch $p_u(x) := p_u(x, \cdot)$ is defined by $p_u(x, h) := u(x + T_u(x)^{-\frac{1}{2}} h)$, with h belonging to Δ_t. We will apply it in Section 4 to exemplar-based inpainting. Let us also remark that formula (11) has the same complexity as the patch comparison formula (1).

The similarity measure corresponding to (11) was derived in [16] as a computationally tractable approximation of the linear case of the multiscale similarity measures introduced in [4]. There, the authors show that all scale spaces of similarity measures $\mathcal{D}(t, x, y)$ satisfying a set of appropriate axioms are solutions of a family of degenerate elliptic partial differential equations (PDE) in the variables (x, y). Images are considered in those papers as Riemannian manifolds endowed with a metric defined by a tensor field. If this tensor field is affine covariant, the resulting similarity measure is affine invariant. In this Riemannian framework $P(\cdot, \cdot)$ defines an isometry between the tangent spaces in two manifolds. The authors refer to it as the *a priori* connection, since it is related to the notion of connection appearing in parallel transport (see [4] for details). WKB approximation method, named after Wentzel, Kramers and Brillouin, was used in [16] to find the approximate solution to a linear partial differential equation with spatially varying coefficients as a convolution with a short-time space-varying kernel.

The affine invariant patch distance (11) is used in the following section in a variational formulation for exemplar-based image inpainting and Section 6 will present inpainting results for both gray and color images. Let us note that for the color case we consider a generalization of (11) to multi-channel images. Let $\overrightarrow{u} : \Omega_u \to \mathbb{R}^M$ and $\overrightarrow{v} : \Omega_v \to \mathbb{R}^M$, $\Omega_u, \Omega_v \subseteq \mathbb{R}^2$ be multi-channel images (e.g., $M = 3$ for color images), then the corresponding affine invariant similarity measure is defined as

$$\mathcal{D}^{\mathrm{a,M}}(t, x, y) = \int_{\Delta_t} g_t(h) \left\| \overrightarrow{u}(x + T_u^{-\frac{1}{2}} R_u(x)h) - \overrightarrow{v}(y + T_v^{-\frac{1}{2}} R_v(y)h) \right\|_2^2 dh, \quad (12)$$

where $\|\cdot\|_2$ denotes the Euclidean norm of vectors in $\mathbb{R}^M$ and the affine covariant structure tensors and neighborhoods are computed using the corresponding gray-value images associated to $\overrightarrow{u}$ and $\overrightarrow{v}$ by the iterative algorithm of Section 3.2.

4 Inpainting formulation

Exemplar-based inpainting methods aim at filling-in the image so that each patch in the inpainting domain is similar to some known patch. This requires comparing known patches with partially or completely unknown patches. For this we extend the variational framework described in [38, 25, 1] by using the affine invariant similarity measure $\mathcal{D}^{\mathrm{a}}$ given in (11). We formulate the problem of inpainting from affinely transformed patches via the minimization of the following energy functional

$$E(u,\varphi) = \int_{\widetilde{O}} \mathcal{D}^{\mathrm{a}}\left(t, x, \varphi(x)\right)\, dx, \tag{13}$$

where $O \subset \Omega \subset \mathbb{R}^2$ is the inpainting domain, $\hat{u} : \Omega \setminus O \to \mathbb{R}$ is the known part of the image, $\widetilde{O}$ includes all the centers of patches intersecting O and $\widetilde{O}^c$ is its complement, that is, $\widetilde{O}^c$ contains centers of fully known patches (see Fig. 4). The minimization of (13) aims at finding a visually plausible completion u of $\hat{u}$ in the unknown region O. The additional variable $\varphi : \widetilde{O} \to \widetilde{O}^c$ determines for each unknown target patch the location of a source patch from which the information will be copied.

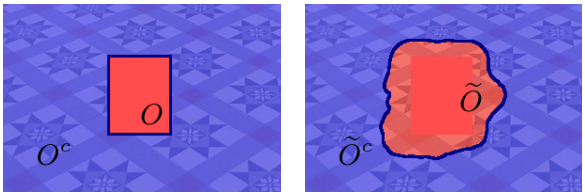


Fig. 4. Schematic representation of the sets O, O^c, $\widetilde{O}$ and $\widetilde{O}^c$.

This energy compares patches defined on elliptical regions centered at x and $\varphi(x)$. In the known part of the image, these regions are defined by the affine covariant structure tensors $T_{\hat{u}}$. Since the image is unknown inside the inpainting domain we have to estimate the structure tensors together with the image. The relationship between u and T_u introduces a complex dependency in the energy (13), which complicates its minimization. Therefore, we propose to relax it and consider instead the

minimization of the energy

$$\widetilde{E}(u,\varphi,G) = \int_{\widetilde{O}} \int_{\Delta_t} g_t(h) \Big(u(x + G(x)^{-\frac{1}{2}}h) - \hat{u}(\varphi(x) + T_{\hat{u}}(\varphi(x))^{-\frac{1}{2}} R_{\hat{u}}(\varphi(x))h)\Big)^2 \, dh dx \tag{14}$$

where $G(x)$ is an invertible 2×2 matrix, $\forall x \in \widetilde{O}$. For now, we will not restrict the tensor field G to be given by the structure tensors T_u. Instead, we consider it as an additional variable, in principle independent of u. In this way, we do not have to deal with the complex dependency between T_u and u. In practice, due to the properties of the affine covariant structure tensors, it turns out that the $G(x)$ can be estimated from $T_u(x)$ and the additional rotation $R_u(x)$, as will be explained later in this section.
We compute a local minimum of the energy with an alternating optimization scheme on the variables u, G and φ which is summarized in Algorithm 1.

4.1 Image Update Step

In the image update step, φ and G are fixed, and the energy is minimized with respect to u. With the change of variables $z = x + G(x)^{-\frac{1}{2}}h$, the Euler-Lagrange equation leads to the following expression:

$$u(z) = \frac{1}{\varrho(z)} \int_{\widetilde{O}} g_t \left(G(x)^{\frac{1}{2}}(z-x)\right) \hat{u}\left(\varphi(x) + T_{\hat{u}}(\varphi(x))^{-\frac{1}{2}} R_{\hat{u}}(\varphi(x)) G(x)^{\frac{1}{2}}(z-x)\right) \left|G(x)^{\frac{1}{2}}\right| dx, \tag{15}$$

where $\varrho(z)$ is normalization factor such that the sum is an average. The field G determines elliptical patches centered at each $x \in \widetilde{O}$. For each one of these patches a matching patch centered at $\varphi(x)$ is known, as well as its shape which is given by the structure tensor $T_{\hat{u}}(\varphi(x))$. The corresponding patch is then warped via the affinity

$$\widetilde{P}(\varphi(x), x) = G(x)^{-\frac{1}{2}} R_{\hat{u}}^{-1}(\varphi(x)) T_{\hat{u}}(\varphi(x))^{\frac{1}{2}}$$

and aggregated in the inpainting domain. Note that if $G(x)$ is given by $T_u(x)$, then $\widetilde{P}(\varphi(x), x)$ coincides with Equation (10).

4.2 Affine Correspondence Update Step

Given a fixed u, the minimization of the energy with respect to (φ, G) can be performed as independent minimization of the patch distance function $\mathcal{D}^{\mathrm{a}}$ for each $x \in \widetilde{O}$. This problem is very complex to solve since it is a nearest neighbor search where we also optimize for the affine transformation of the patch at x, given by G.

We will exploit the properties of the affine covariant structure tensors to obtain an approximate solution. For that, let us consider a completion candidate u and assume that a local vicinity of x on u is an affinely transformed version of a local vicinity of $\varphi(x)$ on $\hat{u}$. That is, $u(x+h) = \hat{u}(\varphi(x)+Ah)$, which is the case when x and $\varphi(x)$ do actually correspond. Setting $G(x)$ such that $G^{-\frac{1}{2}}(x)R_{\hat{u}}^{-1}(\varphi(x))T_{\hat{u}}(\varphi(x))^{\frac{1}{2}} = A$ will lead to a correct mapping and thus to the zero patch distance. On the other hand, using (10) we can find this affinity as $A = T_u(x)^{-\frac{1}{2}}RT_{\hat{u}}(\varphi(x))^{\frac{1}{2}}$ where R is some orthogonal 2×2 matrix and T_u is calculated on u. Then $G(x)$ such that $G^{\frac{1}{2}}(x) = R(x)T_u^{\frac{1}{2}}(x)$, together with $\varphi(x)$, will be global minimizers of the patch distance function $\mathcal{D}^{\mathrm{a}}$ at x. Therefore, we need to search only for $\varphi(x)$ and $R(x)$. An approximate $\varphi(x)$ can be found efficiently using our modified version of the PatchMatch algorithm [6], detailed in Section 5. The additional rotation $R(x)$ is determined as described in Section 3.1, in the same way as for the known part of an image. Notice that for notation consistency we should write $R(x) := R_u(x)^{-1}$.
Of course, if the neighborhood of x does not match any affinely transformed patch, then the estimated G might not minimize the patch distance $\mathcal{D}^{\mathrm{a}}$.

Algorithm 1: Approximate minimization of $\widetilde{E}(u,\varphi,G)$.

Input: Initial condition u^0 at O, tolerance $\tau > 0$
Output: Image completion u

repeat
- Compute affine covariant structure tensors $T_{u^{k-1}}(x)$ and rotations $R_{u^{k-1}}(x)$ for all $x \in \widetilde{O}$;
- Estimate optimal correspondences φ^k using the modified PatchMatch (see Section 5.2);
- Update image $u^k = \arg\min_u \widetilde{E}(u,\varphi^k,G^k)$, subject to $u^k = \hat{u}$ in O^c;

until $\|u^k - u^{k-1}\| < \tau$;

Another interpretation of the approximate minimization can be given by adding to the minimization of $\widetilde{E}(u,\varphi,G)$ the constraint that $G^{\frac{1}{2}}(x) = R_u(x)^{-1}T_u^{\frac{1}{2}}(x)$ for all $x \in \widetilde{O}$ and for some rotation matrix $R_u(x)$, namely,

$$\min \widetilde{E}(u,\varphi,G) \quad \text{subject to } G^{\frac{1}{2}} = R_u^{-1}T_u^{\frac{1}{2}}.$$

The correspondence update step corresponds to the constrained minimization of the energy with respect to φ, G for a fixed image u. In the image update step the energy is minimized with respect to u, but without enforcing the constraint. Therefore, our approximate minimization can be seen as an alternating minimization applied to a constrained problem. The constraint is enforced only when minimizing with one of the variables (the pair φ, G). There are no theoretical guarantees for the

convergence of such a scheme, although we have not yet encountered a practical case where the algorithm failed to converge.

5 Numerical Implementation

5.1 Image Update Step

The actual implementation of (15), that we use in our method, is

$$u(z) = \frac{1}{C(z)} \sum_{x \in \tilde{O}} g_t(T_u^{\frac{1}{2}}(x)(z-x))\, m_c(x) w(x,z) \\ \hat{u}\big(\varphi(x) + P(x,\varphi(x))(z-x)\big)\, |T_u^{\frac{1}{2}}(x)|, \quad (16)$$

where $P(x,\varphi(x)) = T_{\hat{u}}^{-\frac{1}{2}}(\varphi(x)) R_{\hat{u}}(\varphi(x)) R_u^{-1}(x) T_u^{\frac{1}{2}}(x)$ is the estimated affinity mapping the target patch at x onto the source patch at $\varphi(x)$. The structure tensor field T_u is computed using the inpainted image u from the previous iteration.

Of course, in the discrete setting some kind of interpolation needs to be done after transforming one elliptical patch into another by $P(x,\varphi(x))$. For that we use the Nadaraya-Watson estimator [32, 37] with Gaussian kernel.

The extra term m_c in (16) is a so-called *confidence mask* that takes values from 1 to 0, exponentially decreasing with the distance to the set of known pixels O^c. This mask is usual in exemplar-based inpainting, for instance, it is used in [12, 1]. It helps to guide the flow of information from the boundary towards the interior of the inpainting domain, eliminating some local minima and reducing the effect of the initial condition. More precisely, we compute the confidence mask as

$$m_c(x) = (1 - c_0)\exp\left(-\frac{d(x,O^c)}{c_t}\right) + c_0,$$

where $d(x,O^c)$ is the distance from a point x to the boundary of the O^c set, such that $d(x,O^c) = 0$ when $x \in O^c$. Parameter $c_0 > 0$ defines the smallest (asymptotic) value that m_c can take and $c_t > 0$ controls the rate of decay. This confidence mask never changes during the inpainting process and can be precomputed for a given inpainting domain.

There is also another additional weighting term $w(x,z)$ in (16). In principle, all patches containing a pixel z contribute to its color value. To control the amount of contributors, we introduce the auxiliary Gaussian weight, that depends on the patch distance between a contributing patch and its corresponding known patch

$$w(x,z) = \exp\left(-\frac{\big(\mathcal{D}^{\mathrm{a}}(t,x,\varphi(x)) - \min(\mathbf{D}(z))\big)^2}{2\sigma^2_{\text{cut-off}}(z)}\right), \quad (17)$$

where $\mathbf{D}(z) = \{\mathcal{D}^{\mathrm{a}}(t,y,\varphi(y)) : z \in B_{T_u}(y)\}$ is a set of patch distances to known patches, computed among all patches contributing to z, and $\sigma_{\text{cut-off}}(z)$ defines a soft threshold for the patch distance values. This

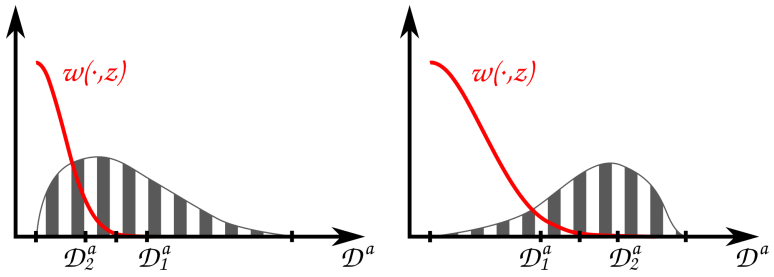


Fig. 5. Schematic depiction of two different cases of patch distance distribution. Gray filled-in curves represent histograms of distance values computed between patches overlapping $z \in O$ and their most similar known counterparts from $\widetilde{O}^c$. Weighting functions are shown in red. In both cases $\gamma_{\text{cut-off}} \approx 0.45$.

weight allows us to cut off contributors with low similarity (high distance) values, which in turn results in sharper reconstructions.
To compute $\sigma_{\text{cut-off}}(z)$ we begin by computing the first estimate for the cut-off distance

$$\mathcal{D}_1^{\mathrm{a}}(z) = \gamma_{\text{cut-off}}\big(\max(\mathbf{D}(z)) - \min(\mathbf{D}(z))\big),$$

where $\gamma_{\text{cut-off}} \in (0,1)$ is a parameter of the method. Since distance values are usually distributed unevenly, the initial distance threshold $\mathcal{D}_1^{\mathrm{a}}$ might discard too few or too many contributors. Therefore, we very roughly estimate the density of values that fall below $\mathcal{D}_1^{\mathrm{a}}$ and refine the initial cut-off distance by

$$\mathcal{D}_2^{\mathrm{a}}(z) = \gamma_{\text{cut-off}} \frac{\mathcal{D}_1^{\mathrm{a}}\ |\mathbf{D}(z)|}{N_1},$$

where $|\mathbf{D}(z)|$ stands for the total number of elements in the set $\mathbf{D}(z)$ and $N_1 = |\{\mathcal{D}^{\mathrm{a}} \in \mathbf{D}(z) : \mathcal{D}^{\mathrm{a}} - \min(\mathbf{D}(z)) < \mathcal{D}_1^{\mathrm{a}}(z)\}|$ is the number of distance values retained by $\mathcal{D}_1^{\mathrm{a}}$. Then the final cut-off threshold is given by

$$\sigma_{\text{cut-off}}(z) = \frac{1}{6}\left(\mathcal{D}_1^{\mathrm{a}}(z) + \mathcal{D}_2^{\mathrm{a}}(z)\right).$$

The factor $\frac{1}{6}$ in the formula above implies that the Gaussian (17) approaches zero at the average cut-off distance between the first and second estimates. Fig. 5 illustrates the cut-off distances for two different cases of distance values distribution.
The energy (13) is non-convex and might have several local minima. As a consequence, there is a dependency on the initialization. To alleviate this dependency, we try to promote the propagation of information from the boundary towards the interior of the inpainting domain during the very first iterations of inpainting. Recall that the extended domain $\widetilde{O}$ contains the centers of all elliptical patches overlapping the inpainting domain. We enlarge $\widetilde{O}$ by a few pixels to capture a narrow stripe $\widetilde{O}^+ = (\widetilde{O} \oplus B) \setminus \widetilde{O}$ around the inpainting domain, that contains centers of completely known elliptical patches. Obviously, these elliptical patches

in $\widetilde{O}^+$ do not intersect the inpainting domain. To make them contribute to the inpainting we should enlarge them first. For that we recompute them doubling the value of r. Notice that we use $r' = 2r$ only for the points within the stripe $\widetilde{O}^+$ and only in the image update step. We do not recompute the corresponding structure tensors, thus we only increase the sizes of these elliptical patches and do not modify their shapes. This additional contribution from elliptical patches, that do not depend by any means on the inpainting domain, boosts the information propagation at the boundaries of the inpainting domain. The width of the stripe $\widetilde{O}^+$ is set to 6 pixels in all our experiments.

5.2 Affine Correspondence Update Step

During the update of the correspondence map we compute an approximation of the nearest neighbor field using PatchMatch [6, 7]. The Patch-

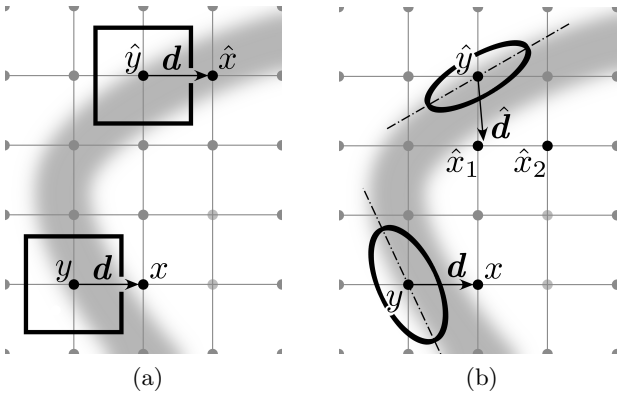


Fig. 6. Propagation directions in PatchMatch: (a) original scheme, (b) modified scheme.

Match algorithm speeds up the computation of optimal correspondences by exploiting the correlation between patches so that they can be found collectively. Since we are working with elliptical patches which might be arbitrarily rotated, we adapt the PatchMatch propagation scheme to take this into account. Let x be the current pixel and $d_1 = (\pm 1, 0)$, $d_2 = (0, \pm 1)$ be the directions of propagation. Then, the adjacent pixels $y_i = x - d_i$ $(i = 1, 2)$ are tested during the propagation. Assume $i = 1$ (see Fig. 6). Pixel $\hat{y} = \varphi(y)$ is the current nearest neighbor candidate for y. The standard PatchMatch would try to propagate position $\hat{y} + d$ to pixel x (Fig. 6a). In contrast, we calculate the direction $\hat{d} = P(y, \hat{y})d$, where P is the a priori connection, and we try a few positions along that direction (Fig. 6b). This generalization gives a more meaningful propagation along edges.

Recall that to compare two elliptical patches we first transform them into discs of the same radius (see Section 3.3). As a result of this normalization we obtain two sets of scattered point, each of which is described by real-valued coordinates and a color value. To be able to compare these sets, we use the Nadaraya-Watson estimator [32, 37] with Gaussian kernel to interpolate them to a regular grid.
At early iterations of our algorithm, the inpainted image may be blurry. This is typical for iterative patch-based inpainting methods and is caused by aggregating patches that do not coincide in their overlap area at the initial iterations. As discussed in [16], the structure tensors are sensitive to blur, tending to larger elliptical patches in blurry regions. Essentially, smoothing of an image suppresses small details and produces the same effect as scaling the image down. The elliptical patches in turn capture larger areas. To compensate for this, we allow the parameter r to vary during the correspondence map estimation. That is, while $T_{\tilde{u}}(\varphi(x))$ is always computed with the fixed r, say r_0 (a given parameter of the method), in the computation of $T_u(x)$ we consider a few (around five) values of r smaller than r_0 and select the one giving the smallest patch distance $\mathcal{D}^{\mathrm{a}}$ between points x and $\varphi(x)$. Let us note, that to be able to compare patches, computed with different values of r, we scale the normalized patches to discs of radius one.

6 Experimental results

In this section we present results obtained by the proposed method. For all the experiments in this section, we compare our results with the ones obtained by the multiscale NL-Means method [38, 25] which we find to be a representative exemplar-based image inpainting method operating only with translations of patches. Whenever possible, we also compare against the method of [27] with a single scale and considering rotations, and the method of [24]. In both cases we use implementations provided by the authors.
As a sanity check we first test the proposed method on a synthetic example, displayed in Fig. 7. We take a textured image and create an affinely transformed version of it. We select a part of the transformed image as the inpainting domain. Instead of using the rest of the transformed image to copy information from, we make *the original (not transformed) image to be the source*. Let us remark that the ground truth affinity is not provided to the algorithm, hence, we test the ability of the proposed method to identify and copy affinely transformed patches. We do not show any results for [27] and [24] for this experiment, since the available implementations do not support the use of a separate image as a source.
A more realistic case would be associated with a more general transformation. Since for planar objects a projective transformation can be locally approximated by an affinity, in the second example (shown in Fig. 8) we test the robustness of our method in the reconstruction of an image distorted by perspective. As usual in inpainting applications, in this experiment we use *the known part of the image as source*. We compare our method with the NL-Means method, that works only with



Fig. 7. First row: source image, target image with the inpainting domain shown in red, and close-ups around the inpainting area of the NL-Means result and the result of our method. Second row: evolution of the inpainting domain over iterations of our method (every second iteration).

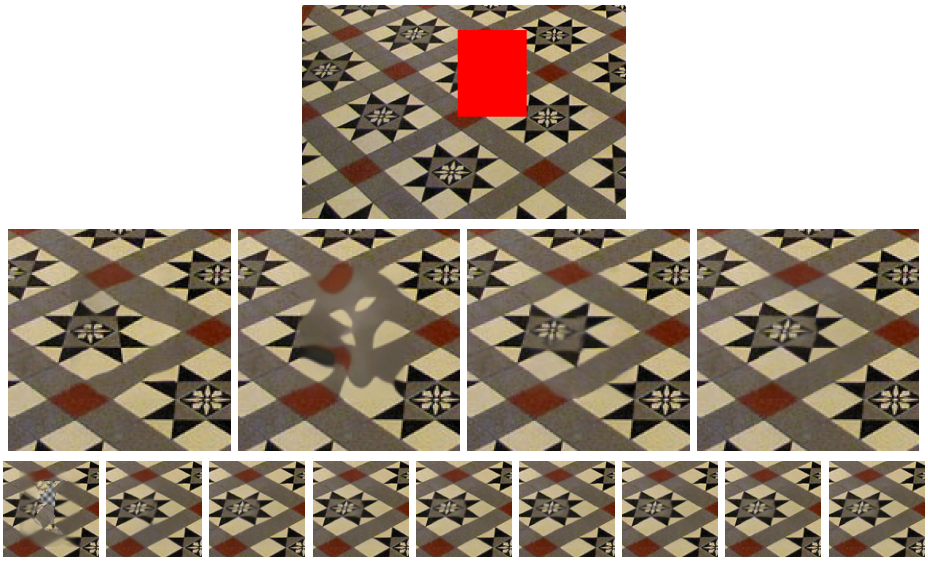


Fig. 8. First row: image with the inpainting domain shown in red. Second row: close-ups around the inpainting area of the NL-Means result, the result of [27] (considering rotations), the result of [24], and the result of our method. Third row: evolution of the inpainting domain over iterations of our method (every third iteration).

translations, and additionally with the method of [27] in the mode when the rotations are also considered, and the method of [24]. Note that the

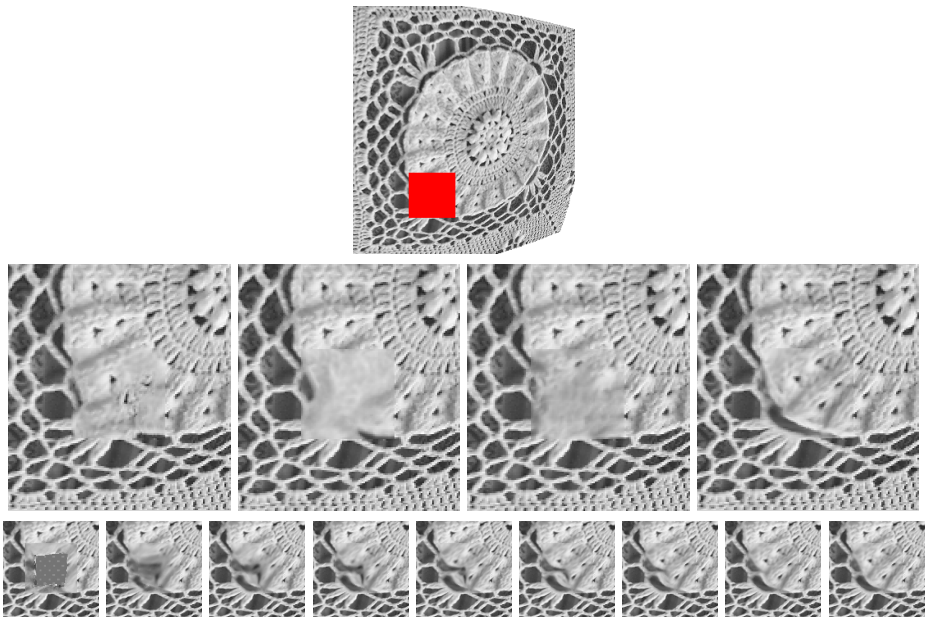


Fig. 9. First row: image with the inpainting domain shown in red. Second row: close-ups around the inpainting area of the NL-Means result, the result of [27] (considering rotations), the result of [24], and the result of our method. Third row: evolution of the inpainting domain over iterations of our method (every third iteration).



Fig. 10. First row: source image, target image with the inpainting domain shown in red, and close-ups around the inpainting area of the NL-Means result and the result of our method. Second row: evolution of the inpainting domain over iterations of our method (every second iteration).

latter method successfully determines a single plane in the image and, as expected, achieves a good reconstruction.

The third example (Fig. 9) demonstrates the reconstruction of a texture with some lens distortion applied to it. *The known part of the image is used as a source* and, like in all other experiments, just a rotation of source patches is not sufficient to obtain a good result. As in the previous case, here we compare our method with the NL-Means method (translations), the method of [27] (translations and rotations), and the method of [24] (projective transformation).
A final experiment, which is also potentially interesting for real applications, consists in inpainting one view of a scene using information from another view of the same scene. Fig. 10 shows the results of this experiment where we have applied the proposed method to two views related by an unknown homography. As before, we compare our result with the result of the NL-Means method.
Let us note that the method of [27] also supports rotations plus scalings. However, we could not obtain meaningful results on these examples for this mode. It seems that the additional variability added by the scalings makes it easier for the algorithm to be trapped in a bad local minimum. For example, a constant region can be produced by scaling a small uniform patch.
Finally, we briefly discuss the limitations of the proposed method. Since the transformation between two patches is estimated from the surrounding texture, the method fails when there is not enough textural information (Fig. 11, first row). Severe transformations between pairs of patches may be recovered incorrectly. This can be illustrated by replacing the source image in the last experiment with a much more slanted view (Fig. 11, second row). The proposed method does not exploit the common multiscale scheme which limits the maximum possible size of the inpainting domain (Fig. 11, third row).

7 Conclusions

In this work we propose a new variational formulation for exemplar-based image inpainting that, for the first time, considers local full affine transformations with a tractable approximate optimization scheme. This is possible thanks to the use of an affine invariant patch similarity measure constructed from affine covariant structure tensors, both introduced in [16]. These structure tensors provide an efficient way to determine a unique affinity putting in correspondence any pair of patches. If the patches being compared are related by an affinity, then this affinity is recovered. We show that our method can be applied to complete a perspectively distorted color texture as well as to automatically inpaint one view of a scene using other view of the same scene as a source.
The problem of exemplar-based inpainting is a complex non-convex problem with many local minima. As pointed out in [10], adding transformations of patches makes it even more complex. Intuitively, the added variability makes it harder to distinguish "good" minima from other minima (a single pixel can be scaled to match a constant patch). The structure tensors are beneficial in this respect, because they constrain the number of ways in which a source patch can be transformed to match a target

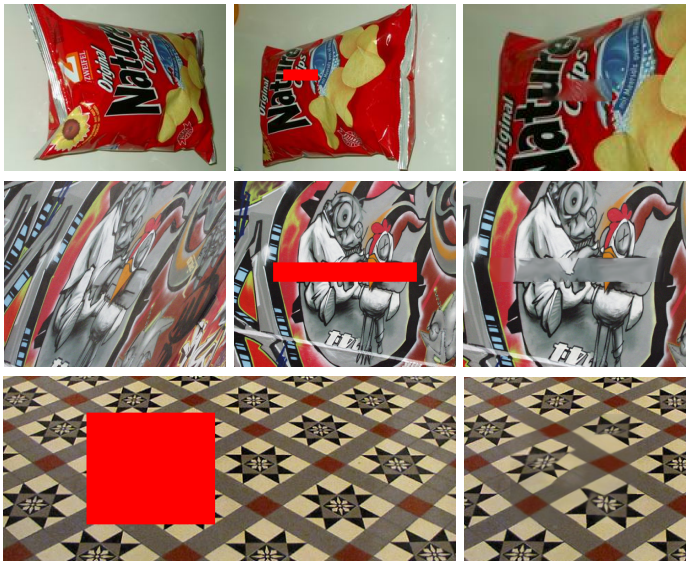


Fig. 11. Failure cases. From top to bottom: insufficient textural information, severe distortion, too big inpainting domain. More details are given in the text.

patch, thus eliminating some of the variability. This also allows us to design faster and more accurate minimization algorithms without the need to search the parameter space of the transformation family.

The proposed method works at a single scale. To better handle larger inpainting domains it would be desirable to develop a multiscale scheme, as is customary in the literature [38, 25, 1]. However, extending the multiscale approach to the problem of inpainting using affinely transformed patches is not trivial, since the filtering with an isotropic Gaussian breaks the affine invariance. Adapting multiscale inpainting approaches to this context is an interesting direction for future research.

Acknowledgments. The first, second and fourth authors acknowledge partial support by the MINECO/FEDER project with reference TIN2015-70410-C2-1-R , the MICINN project with reference MTM2012-30772, and by GRC reference 2014 SGR 1301, Generalitat de Catalunya. The second and third authors were partly founded by the Centre National dEtudes Spatiales (CNES, MISS Project), BPIFrance and Région Ile de France, in the framework of the FUI 18 Plein Phare project, the European Research Council (advanced grant Twelve Labours n246961), the Office of Naval research (ONR grant N00014-14-1-0023), and ANR-DGA project ANR-12-ASTR-0035.

References

1. P. Arias, G. Facciolo, V. Caselles, and G. Sapiro. A variational framework for exemplar-based image inpainting. *International Journal of Computer Vision*, 93:319–347, 2011.
2. J.-F. Aujol, S. Ladjal, and S. Masnou. Exemplar-based inpainting from a variational point of view. *SIAM Journal on Mathematical Analysis*, 42(3):1246–1285, 2010.
3. C. Ballester, M. Bertalmío, V. Caselles, G. Sapiro, and J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Transactions on Image Processing*, 10(8):1200–1211, 2001.
4. C. Ballester, F. Calderero, V. Caselles, and G. Facciolo. Multiscale analysis of similarities between images on riemannian manifolds. *SIAM Journal Multiscale Modeling and Simulation*, 12(2):616–649, 2014.
5. C. Ballester and M. Gonzalez. Affine invariant texture segmentation and shape from texture by variational methods. *Journal of Mathematical Imaging and Vision*, 9(2):141–171, 1998.
6. C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. PatchMatch: a randomized correspondence algorithm for structural image editing. In *ACM SIGGRAPH 2009 Papers*, pages 1–11, New York, NY, USA, 2009. ACM.
7. C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein. The generalized PatchMatch correspondence algorithm. In *European Conference on Computer Vision*, 2010.
8. M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of SIGGRAPH*, pages 417–424, 2000.
9. A. Buades, B. Coll, and J.-M. Morel. A non local algorithm for image denoising. In *Proceedings of the IEEE Conference on CVPR*, volume 2, pages 60–65, 2005.
10. F. Cao, Y. Gousseau, S. Masnou, and P. Prez. Geometrically guided exemplar-based inpainting. *SIAM Journal on Imaging Sciences*, 4(4):1143–1179, 2011.
11. T. Chan and J. H. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. App. Math.*, 62(3):1019–43, 2001.
12. 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.
13. L. Demanet, B. Song, and T. Chan. Image inpainting by correspondence maps: a deterministic approach. *Applied and Computational Mathematics*, 1100:217–50, 2003.
14. I. Drori, D. Cohen-Or, and H. Yeshurun. Fragment-based image completion. In *ACM SIGGRAPH 2003 Papers*, volume 22, pages 303–12, 2003.
15. A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Proceedings of the IEEE ICCV*, pages 1033–38, September 1999.
16. V. Fedorov, P. Arias, R. Sadek, G. Facciolo, and C. Ballester. Linear multiscale analysis of similarities between images on riemannian manifolds: Practical formula and affine covariant metrics. *SIAM Journal on Imaging Sciences*, 8(3):2021–2069, September 2015.

17. A. Foi and G. Boracchi. Foveated self-similarity in nonlocal image filtering. *Human Vision and Electronic Imaging XVII*, 8291(1):829110, February 2012.
18. J. Garding and T. Lindeberg. Direct estimation of local surface shape in a fixating binocular vision system. In *Eklundh, Lecture Notes in Computer Science*, pages 365–376, 1994.
19. G. Gilboa and S. J. Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling and Simulation*, 7(3):1005–1028, 2008.
20. J. Gårding. Shape from texture for smooth curved surfaces in perspective projection. *Journal of Mathematical Imaging and Vision*, 2(4):327–350, 1992.
21. J. Gårding and T. Lindeberg. Direct computation of shape cues using scale-adapted spatial derivative operators. *International Journal of Computer Vision*, 17(2):163–191, 1996.
22. J. Hays and A.A. Efros. Scene completion using millions of photographs. In *SIGGRAPH*, New York, NY, USA, 2007. ACM.
23. J-. B. Huang, J. Kopf, N. Ahuja, and S. B. Kang. Transformation guided image completion. In *International Conference on Computational Photography*, pages 1–9, April 2013.
24. J. B. Huang, S. B. Kang, N. Ahuja, and J. Kopf. Image completion using planar structure guidance. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2014)*, 33(4):129:1–129:10, 2014.
25. N. Kawai, T. Sato, and N. Yokoya. Image inpainting considering brightness change and spatial locality of textures and its evaluation. In *Advances in Image and Video Technology*, pages 271–282. 2009.
26. D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
27. A. Mansfield, M. Prasad, C. Rother, T. Sharp, P. Kohli, and L. van Gool. Transforming image completion. In *Proceedings of BMVC*, pages 121.1–121.11, 2011.
28. S. Masnou. Disocclusion: a variational approach using level lines. *IEEE Transactions on Image Processing*, 11(2):68–76, 2002.
29. S. Masnou and J.-M. Morel. Level lines based disocclusion. In *Proceedings of IEEE ICIP*, volume 3, pages 259–263, 1998.
30. J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.
31. K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.
32. E. A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.
33. D. Pavić, V. Schonefeld, and L. Kobbelt. Interactive image completion with perspective correction. *The Visual Computer*, 22(9-11):671–681, 2006.
34. G. Peyré. Manifold models for signals and images. *Computer Vision and Image Understanding*, 113(2):249–260, 2009.
35. L. Pizarro, P. Mrázek, S. Didas, S. Grewenig, and J. Weickert. Generalised nonlocal image smoothing. *International Journal of Computer Vision*, 90:62–87, 2010.

36. Z. Wang. Image affine inpainting. In *Image Analysis and Recognition*, volume 5112 of *Lecture Notes in Computer Science*, pages 1061–1070. 2008.
37. Geoffrey S. Watson. Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)*, 26(4):359–372, 1964.
38. Y. Wexler, E. Shechtman, and M. Irani. Space-time completion of video. *IEEE Transactions on PAMI*, 29(3):463–476, 2007.