Affine Invariant Self-similarity for Exemplar-based Inpainting

Vadim Fedorov[1], Pablo Arias[2], Gabriele Facciolo[3] and Coloma Ballester[1]
[1]*Department of Information and Communication Technologies, University Pompeu Fabra, Barcelona, Spain*
[2]*CMLA, ENS Cachan, Cachan, France*
[3]*IMAGINE/LIGM, École Nationale des Ponts et Chaussées, Marne-la-Vallée, France*

Keywords: Image Inpainting, Self-similarity, Affine Invariance.

Abstract: This paper presents a new method for exemplar-based image inpainting using transformed patches. We build upon a recent affine invariant self-similarity measure which automatically transforms patches to compare them in an appropriate manner. As a consequence, it intrinsically extends the set of available source patches to copy information from. When comparing two patches, instead of searching for the appropriate patch transformation in a highly dimensional parameter space, our approach allows us to determine a single transformation from the texture content in both patches. We incorporate the affine invariant similarity measure in a variational formulation for inpainting and present an algorithm together with experimental results illustrating this approach.

1 INTRODUCTION

Image inpainting, also known as image completion or disocclusion, 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 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 a projective transformation

Figure 1: Self-similarity under perspective distortion. The human brain can easily fill-in missing information behind the red rectangles in the examples above. We propose a method able to automatically do the same for distortions that can be locally approximated by an affinity.

(see Figure 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 Figure 1 one can easily infer what is hidden behind the red rectangle in the graffiti scene on the right, or use the non-trivially distorted context in the left 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 highly dimensional space, our approach allows us to determine a single transformation from a surrounding texture content. As opposed to some previous works

Fedorov, V., Arias, P., Facciolo, G. and Ballester, C.
Affine Invariant Self-similarity for Exemplar-based Inpainting.
DOI: 10.5220/0005728100480058
In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 3: VISAPP*, pages 50-60
ISBN: 978-989-758-175-5

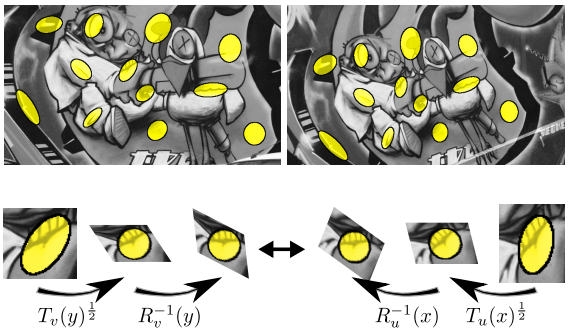

Figure 2: Affine covariant neighborhoods computed at corresponding points in two images taken from different viewpoints. The neighborhoods are computed with the algorithm described in Section 3. Note that, despite the change in appearance, the neighborhoods manage to capture the same underlying texture. An affine invariant patch comparison is achieved by normalizing the patches to circles and aligning them with suitable rotations, as depicted in the diagram at the bottom.

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 (Fedorov et al., 2015), where affine covariant 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 tensors determine elliptical patches at each location of the image domain. Due to the affine covariance property of the tensors, these patches transform appropriately when computed on an affinely transformed version of the image. Figure 2 illustrates the patches defined by the affine covariant tensors of (Fedorov et al., 2015), computed for a set of corresponding points in two images related by a homography. Note 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. Section 2 reviews the related work. In Section 3, we summarize the results of (Fedorov et al., 2015) which motivates the definition of the similarity measure we use. Section 4 is devoted to the inpainting method and algorithm we propose. In Section 5 we present some experiments asserting the validity of our theoretical approach together with a comparison with well-known exemplar based methods. Finally, Section 6 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 (Masnou and Morel, 1998; Ballester et al., 2001; Masnou, 2002; Chan and Shen, 2001b; Cao et al., 2011), with propagation PDE's (Bertalmío et al., 2000), or as the total variation of the image (Chan and Shen, 2001a). These methods perform well in propagating smooth level lines or gradients, but fail in the presence of texture or big inpainting domains.

Exemplar-based (also called texture-oriented) methods were initiated by the work of Efros and Leung (Efros and Leung, 1999) 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 (Foi and Boracchi, 2012; Buades et al., 2005; Gilboa and Osher, 2008; Peyré, 2009; Pizarro et al., 2010). 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 (Hays and Efros, 2007).

The exemplar-based approach to inpainting has been intensively studied (Demanet et al., 2003; Criminisi et al., 2004; Wexler et al., 2007; Kawai et al., 2009; Aujol et al., 2010; Arias et al., 2011). 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. (Drori et al., 2003) used heuristic criteria to vary the scale of patches. Mansfield et al. (Mansfield et al., 2011) and Barnes et al. (Barnes et al., 2010) 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, authors of (Cao et al., 2011) 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 (Pavić et al., 2006; Huang et al., 2013) 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 for searching for correspondences. Recently, Huang et al. (Huang et al., 2014) 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 (Fedorov et al., 2015). Our method considers a rich patch space that includes all affine-transformed patches, however, 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 (Gårding, 1992; Gårding and Lindeberg, 1996; Ballester and Gonzalez, 1998).

Let us remark that this similarity measure applies to any transformation that can locally be approximated by an affinity. Moreover, since it has the same complexity as the usual weighted Euclidean distance between patches, it is thus well-suited for practical applications.

In this paper we extend the variational framework described in (Wexler et al., 2007; Kawai et al., 2009; Arias et al., 2011) proposing a new energy and an optimization algorithm for affine invariant exemplar-based inpainting.

Let us finally note that (Wang, 2008) 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 (Mikolajczyk and Schmid, 2004), object recognition (Matas et al., 2004), and stereo (Garding and Lindeberg, 1994).

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 $v : \Omega_v \to \mathbb{R}$, $\Omega_u, \Omega_v \subseteq \mathbb{R}^2$. A widely used comparison measure between two patches centered respectively at x and y is the weighted squared Euclidean distance

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

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

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

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 (Barnes et al., 2010; Mansfield et al., 2011) 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 (Fedorov et al., 2015), that automatically deduces this transformation from the local texture context.

The similarity measure defined in (Fedorov et al., 2015) is based on affine covariant tensor fields *a priori* computed in each image. It was derived as an approximation to a more general framework introduced

in (Ballester et al., 2014), 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 tensors. Then we show how they are used to define the affine invariant similarity measure and establish the relation between our derivation and the theory of (Ballester et al., 2014; Fedorov et al., 2015). Finally, we describe an algorithm to compute the affine convariant tensors.

Affine Covariant Structure Tensors. We consider an image-dependent tensor field T_u as a function that associates a tensor (a symmetric, positive semi-definite 2×2 matrix) to each point x in the image domain. The 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. A geometric interpretation of this property is the following. Given a tensor $T_u(x)$ we can associate to 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 tensor is affine covariant, we have that $AB_{u_A}(x,r) = B_u(Ax,r)$. This implies that the tensors can be used to define regions that transform appropriately via an affinity (Figure 2).

The affine covariant tensors at two corresponding locations allow to extract the affine distortion between corresponding neighborhoods (or patches), up to a rotation, as shown in (Fedorov et al., 2015). 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 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 $T_u(x)^{1/2}$, 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 to T_{u_A} at x to the one associated to 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 $R = R_v(y)R_u^{-1}(x)$, where $R_v(y)$ and $R_u(x)$ 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 (Lowe, 2004).

An Affine Invariant Patch Similarity. We are interested in comparing the neighborhoods around two points x, y defined in images u and v. The neighborhoods are defined by the local metric given by the tensors $T_u(x)$ and $T_v(y)$. In order to compare both neighborhoods, a mapping between them is needed. Eq. (5) suggest the following definition:

$$P_R(x,y) = T_v(y)^{-\frac{1}{2}} R_v(y) R_u^{-1}(x) T_u(x)^{\frac{1}{2}}. \tag{6}$$

We can interpret $P_R(x,y)$ as an affinity, mapping the elliptical patch associated to $T_u(x)$ into the one associated to $T_v(y)$. If v is an affinely transformed version of u, 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 result of applying $P_R(x,y)$ to the patch at x. In practice, it is more suitable to transform both neighborhoods to the circle of radius r (as depicted in the second line of Figure 2) and compare the aligned normalized patches:

$$\mathcal{D}^{\mathrm{a}}(t,x,y) = \int_{\Delta_t} g_t(h) \cdot \left(u(x + T_u^{-\frac{1}{2}} R_u(x)h) - v(y + T_v^{-\frac{1}{2}} R_v(y)h) \right)^2 dh, \tag{7}$$

where Δ_t is a disc centered at the origin with radius proportional to the scale t and big enough such that the weighting function g_t has effective support in Δ_t. $\mathcal{D}^{\mathrm{a}}$ is an affine invariant patch distance which intrinsically extends the set of available patches. We will apply it in Section 4 to exemplar-based inpainting. Let us also remark that formula (7) has the same complexity of the patch comparison formula (1).

The similarity measure corresponding to (7) was derived in (Fedorov et al., 2015) as a computationally tractable approximation of the linear case of the multiscale similarity measures introduced in (Ballester et al., 2014). 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). 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_R 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 (Ballester et al., 2014) for details).

WKB approximation method, named after Wentzel, Kramers and Brillouin, was used in (Fedorov et al., 2015) to find this approximate solution to a linear partial differential equation with spatially varying coefficients as a convolution with a short-time space-varying kernel.

Computation of Affine Covariant Tensors. The following iterative algorithm introduced in (Fedorov et al., 2015) allows us to compute a dense field of affine covariant tensors and the associated neighborhoods on an image u

$$T_u^{(k)}(x) = \frac{\int_{B_u^{(k-1)}(x,r)} Du(y) \otimes Du(y)\, dy}{\text{Area}(B_u^{(k-1)}(x,r))}, \qquad (8)$$

where $B_u^{(k)}$ is the elliptical region associated to T_u^k given by (4) for $k \geq 2$, and $B_u^{(0)}(x,r) = \{y : |Du(x)(y-x)| \leq r\}$ for $k = 1$.

In this paper we follow the notation of (Fedorov et al., 2015) and denote by $T_u(x)$ the affine covariant structure tensor $T_u^{(k)}(x)$ for a fixed value of k ($k = 30$) 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 $B_u^{(k)}(x,r)$.

To simplify notation in the following sections, we are going to assume that $R_u = Id$ and $R_v = Id$. In order to stress that $\mathcal{D}^{\mathrm{a}}$ in (7) refers to a patch distance, we will use interchangeably $\mathcal{D}^{\mathrm{a}}(t,x,y)$ or $\mathcal{D}_t^{\mathrm{a}}(p_u(x), p_v(y))$, where $p_u(x)$ denotes the elliptic patch centered at x. The patch $p_u(x) := p_u(x,\cdot)$ is defined by $p_u(x,h) := u(x + T_u(x)^{-\frac{1}{2}}h)$, where h belongs to a disc centered at $0 \in \mathbb{R}^2$. Note that the scale t in (7) reflects the support of the Gaussian g_t and, hence, the size of the patch used for the comparison.

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 (Wexler et al., 2007; Kawai et al., 2009; Arias et al., 2011) by using the affine invariant similarity measure $\mathcal{D}_t^{\mathrm{a}}$ given in (7). We formulate the problem of inpainting from affinely transformed patches via the minimization of the following energy functional

$$E(u,\varphi) = \int_{\widetilde{O}} \mathcal{D}_t^{\mathrm{a}}\left(p_u(x), p_{\hat{u}}(\varphi(x))\right) dx, \qquad (9)$$

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 (i.e., the centers of unknown patches) and $\widetilde{O}^c$ is its complement, i.e. the centers of fully known patches. The minimization of (9) aims at finding a visually plausible completion u of $\hat{u}$ in the unknown region O. While 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.

This energy compares patches defined on elliptic domains centered at x and $\varphi(x)$. In the known part of the image, these domains are defined by the affine structure tensors $T_{\hat{u}}$. Since the image is unknown inside the inpainting domain we have to estimate the tensors together with the image. The relationship between u and T_u introduces a complexity in the minimization of (9). Therefore, we propose to relax it and consider the minimization of the energy

$$\widetilde{E}(u,\varphi,G) = \int_{\widetilde{O}} \int_{\Delta_t} g_t(h) \left(u(x + G(x)^{-\frac{1}{2}}h) - \hat{u}(\varphi(x) + T_{\hat{u}}(\varphi(x))^{-\frac{1}{2}}h)\right)^2 dh dx \qquad (10)$$

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 affine structure tensors T_u. Instead, we consider them 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 structure tensors, it turns out that the G can be estimated from $T_u(x)$, as will be explained in the next section.

4.1 Approximate Minimization Algorithm

We compute a local minimum of the energy with an alternating optimization scheme on the variables u, G and φ.

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:

$$u(z) = \frac{1}{C(z)} \int_{\widetilde{O}} g_t(G(x)^{\frac{1}{2}}(z-x)) \\ \hat{u}\left(\varphi(x) + T_{\hat{u}}(\varphi(x))^{-\frac{1}{2}} G(x)^{\frac{1}{2}}(z-x)\right) |G(x)^{\frac{1}{2}}| dx, \tag{11}$$

where $C(z)$ is normalization factor such that the sum is an average. The field G determines elliptic 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 tensor $T_{\hat{u}}(\varphi(x))$. The corresponding patch is then warped via the affinity $P(x,\varphi(x)) = T_{\hat{u}}(\varphi(x))^{-\frac{1}{2}} G(x)^{\frac{1}{2}}$, and aggregated in the inpainting domain. Note that if $G(x) = T_u(x)$, then $P(x,\varphi(x))$ coincides with Eq. (6) (recall that for simplicity in the presentation we are assuming that the rotations in Eq. (6) are the identity).

Affine Correspondence Update Step. Given a fixed u, the minimization with respect to (φ, G) can be performed by independently minimizing the patch error function $\mathcal{D}_t^{\mathrm{a}}$ 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 structure tensors to estimate 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 $T_{\hat{u}}(\varphi(x))^{-\frac{1}{2}} G^{\frac{1}{2}}(x) = A$ will lead to a correct mapping and zero patch distance. On the other hand, using (6) we can find this affinity as $A = T_{\hat{u}}(\varphi(x))^{-\frac{1}{2}} R T_u(x)^{\frac{1}{2}}$ 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 error function $\mathcal{D}_t^{\mathrm{a}}$ 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 (Barnes et al., 2009), detailed in Section 4.2, and the additional rotation $R(x)$ can be determined as described in Section 3. Of course, if the neighborhood of x does not match any affinely transformed patch, then the estimated G might not minimize the patch error function $\mathcal{D}_t^{\mathrm{a}}$.

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

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

Require: Initial condition u^0 at O, tolerance $\tau > 0$.

1: **repeat**
2: Compute affine structure tensors $T_{u^{k-1}}(x)$ and rotations $R^{k-1}(x)$ for all $x \in \widetilde{O}$.
3: Estimate optimal correspondences φ^k using the modified PatchMatch (see Sect 4.2).
4: Update image: $u^k = \arg\min_u \widetilde{E}(u, \varphi^k, G^k)$, subject to $u^k = \hat{u}$ in O^c.
5: **until** $\|u^k - u^{k-1}\| < \tau$.

4.2 Implementation Details

Image Update Step. The actual implementation of (11) is

$$u(z) = \frac{1}{C(z)} \sum_{x \in \widetilde{O}} g_t(T_u^{\frac{1}{2}}(x)(z-x))\, m_c(x) w(x,\varphi(x)) \\ \hat{u}\left(\varphi(x) + \widetilde{P}_R(x,\varphi(x))(z-x)\right) |T_u^{\frac{1}{2}}(x)|, \tag{12}$$

where $\widetilde{P}_R(x,\varphi(x)) = T_{\hat{u}}^{-\frac{1}{2}}(\varphi(x)) R T_u^{\frac{1}{2}}(x)$ is the estimated a priory connection between x and $\varphi(x)$. The tensor field T_u is computed using the inpainted image u from the previous iteration. m_c is a confidence mask that takes values from 1 to 0, decreasing with the distance to the set of known pixels O^c. This mask is usual in exemplar-based inpainting, e.g. (Arias et al., 2011), since 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. Finally, let

us comment on the additional weight $w(x,\varphi(x))$. Usually, all patches containing a pixel z contribute to its color value. To control the amount of contributors, we introduce an auxiliary Gaussian weight $w(x,\varphi(x))$ that depends on our patch distance $\mathcal{D}^{\mathrm{a}}(t,x,\varphi(x))$. It allows us to cut off contributors with low similarity (high distance) values, which in turn results in sharper reconstructions.

The energy (9) is non-convex and has several local minimia. As a consequence, there is a dependency on the initialization. To alleviate this dependency, we aid the propagation of information from the boundary towards the interior of the inpainting domain in the following way. Recall that the extended domain $\widetilde{O}$ contains the centers of all ellipses overlapping the inpainting domain. We dilate $\widetilde{O}$ a few pixels to capture a narrow band around the inpainting domain of completely known elliptic patches. Since these patches do not intersect the inpainting domain, they do not yet contribute to the inpainting. For these pixels, during the image update step we extend their neighborhoods by setting a larger radius value $2r$ (for this we do not recompute the tensors, therefore, we only change the size but not the shape of the associated neighborhood). This results in bigger ellipses, which now may overlap the domain. This boosts the information propagation at the boundaries of the inpainting domain during the very first iterations of inpainting.

Affine Correspondence Update Step. During the update of the correspondence map we compute an approximation of the nearest neighbor field using PatchMatch (Barnes et al., 2009; Barnes et al., 2010). The PatchMatch 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 elliptic patches which might be arbitrarily rotated, we adapt the PatchMatch propagation scheme to take it 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 Figure 3). 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. In contrast, we calculate the direction $\hat{d} = \widetilde{P}_R(y,\hat{y})d$, where $\widetilde{P}_R(\cdot,\cdot)$ is the a priori connection, and we try a few positions along that direction. This generalization gives more meaningful propagation along edges.

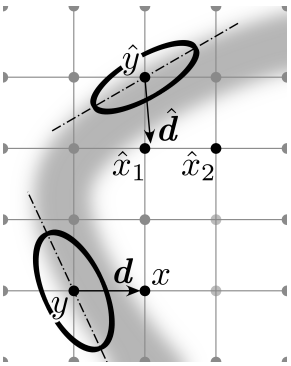

Figure 3: Propagation directions in the modified scheme.

At early iterations of our algorithm, the inpainted image may be blurry. It comes from inconsistency between color values proposed for filling-in pixels in the inpainting domain. As discussed in (Fedorov et al., 2015), the tensors are sensitive to blur, tending to larger neighbourhoods in blurry regions. To compensate for this, we allow the parameter r (in equation 8) to vary during the correspondence map estimation. That is, while $T_{\hat{u}}(\varphi(x))$ 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 values of r smaller than r_0 (around 5) and select the one giving the smallest patch distance $\mathcal{D}_t^{\mathrm{a}}$ between $p_u(x)$ and $p_{\hat{u}}(\varphi(x))$. Let us note, that to be able to compare patches, computed with different values of r, we scale the normalized circles to circles of radius one.

5 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 (Wexler et al., 2007; Kawai et al., 2009) which we find to be a representative exemplar-based image inpainting method operating with only translations of patches. Whenever possible, we also compare against the method of (Mansfield et al., 2011) with a single scale and considering rotations, and the method of (Huang et al., 2014). In both cases we use the authors' implementations.

As a sanity check we first test the proposed method on a synthetic example, displayed in Figure 4. 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 (Mansfield et al., 2011) and (Huang et al., 2014) for this experiment, since the

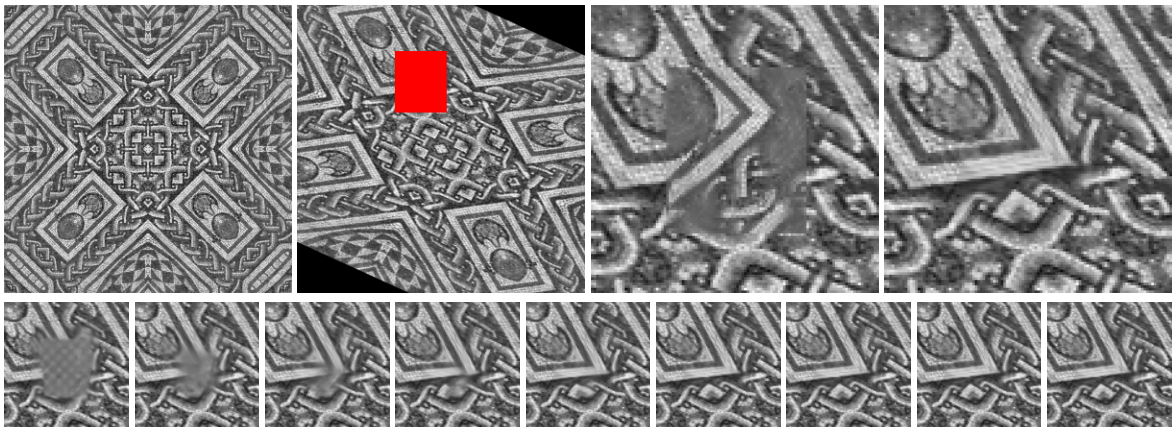

Figure 4: First row: source image, target image with the inpainting domain shown in red, and close-ups around the inpainting area of the NL-Means result and the result of our method. Second row: evolution of the inpainting domain over iterations of our method (every third iteration).

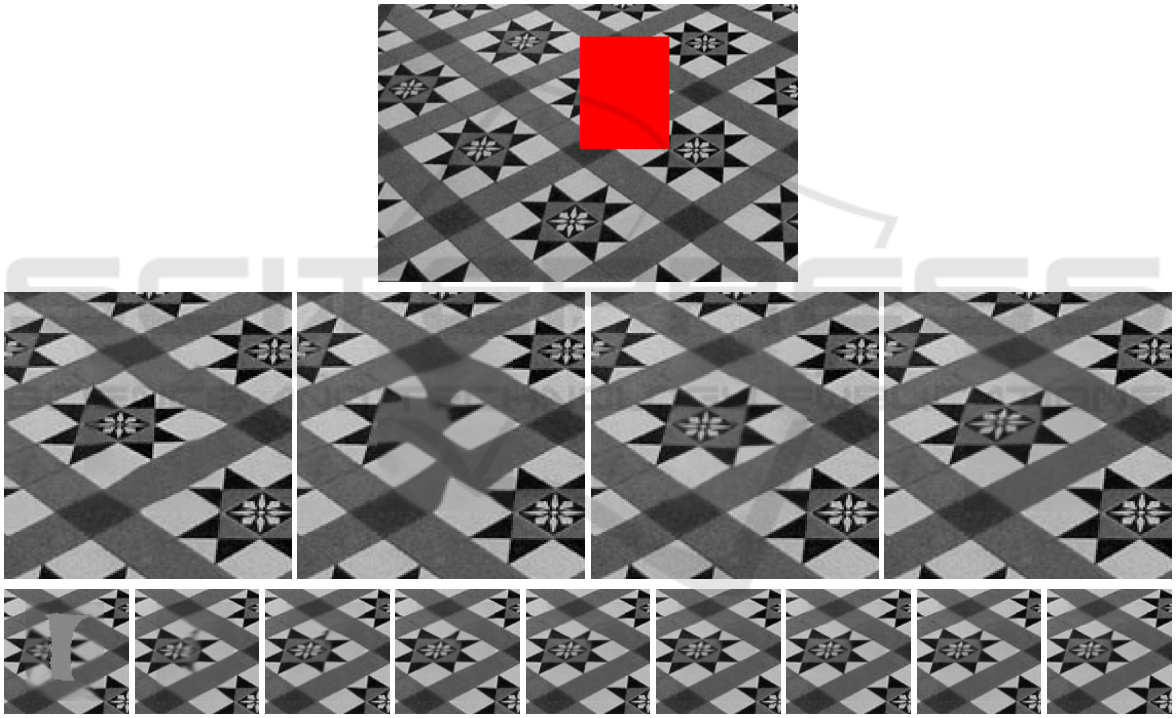

Figure 5: First row: image with the inpainting domain shown in red. Second row: close-ups around the inpainting area of the NL-Means result, the result of (Mansfield et al., 2011) (considering rotations), the result of (Huang et al., 2014), and the result of our method. Third row: evolution of the inpainting domain over iterations of our method (every third iteration).

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 Figure 5) 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 translations, and additionally with the method of (Mansfield et al., 2011) in the mode when the rotations are also considered, and the method of (Huang et al., 2014). Note that the latter method successfully determines a single plane in the image and, as expected, achieves a good reconstruction.

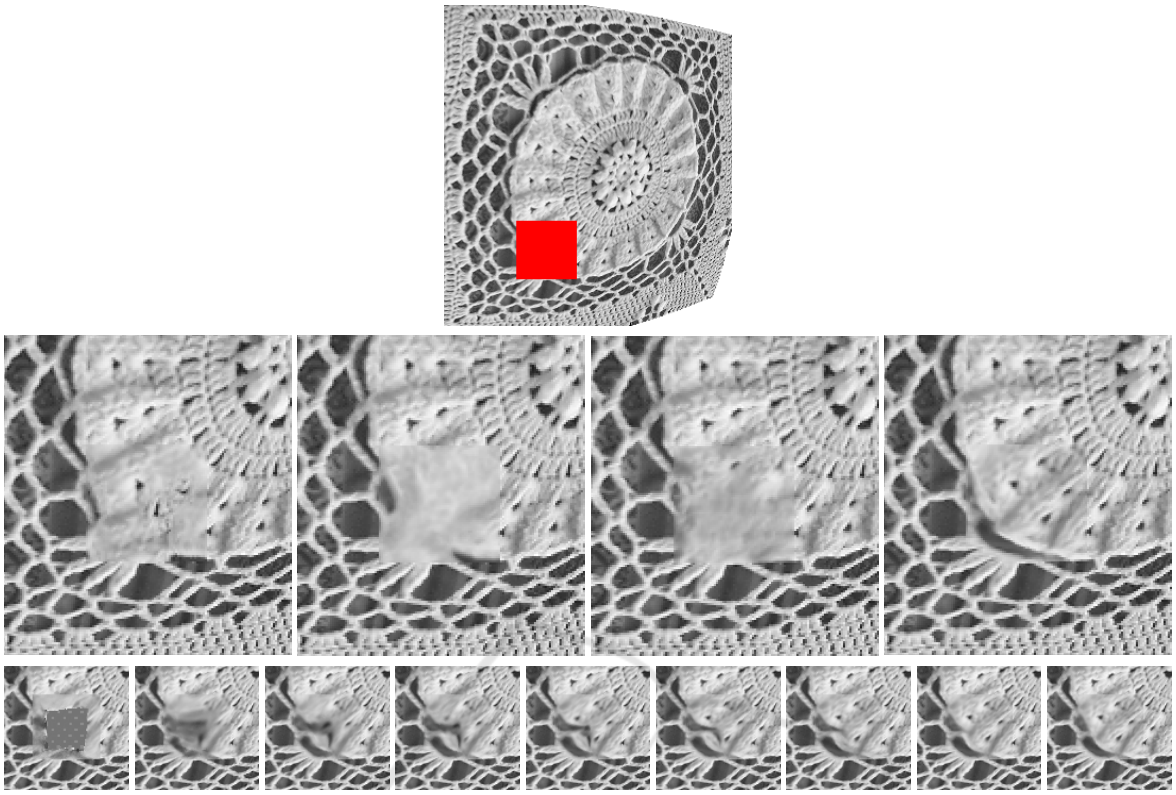

Figure 6: First row: image with the inpainting domain shown in red. Second row: close-ups around the inpainting area of the NL-Means result, the result of (Mansfield et al., 2011) (considering rotations), the result of (Huang et al., 2014), and the result of our method. Third row: evolution of the inpainting domain over iterations of our method (every third iteration).

Figure 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 third iteration).

The third example (Figure 6) 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 (Mansfield et al., 2011) (translations and rotations), and the method of (Huang et al., 2014) (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. Figure 7 shows the results of this experiment where we have applied the proposed method to two views related by an unknown homog-

raphy. As before, we compare our result with the result of the NL-Means method.

Let us finally note that the method of (Mansfield et al., 2011) 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.

6 CONCLUSIONS

In this work we propose a new variational formulation for exemplar-based 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 the affine covariant tensors and the associated affine invariant metric, both introduced in (Fedorov et al., 2015). These 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.

The problem of exemplar-based inpainting is a complex non-convex problem with many local minima. As pointed out in (Cao et al., 2011), 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). We believe that the 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 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 (Wexler et al., 2007; Kawai et al., 2009; Arias et al., 2011). 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.

ACKNOWLEDGEMENTS

The first, the second and the fourth authors acknowledge partial support by MICINN project, 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 d'Etudes Spatiales (CNES, MISS Project), BPIFrance and Rgion 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

Arias, P., Facciolo, G., Caselles, V., and Sapiro, G. (2011). A variational framework for exemplar-based image inpainting. *International Journal of Computer Vision*, 93:319–347.

Aujol, J.-F., Ladjal, S., and Masnou, S. (2010). Exemplar-based inpainting from a variational point of view. *SIAM Journal on Mathematical Analysis*, 42(3):1246–1285.

Ballester, C., Bertalmío, M., Caselles, V., Sapiro, G., and Verdera, J. (2001). Filling-in by joint interpolation of vector fields and gray levels. *IEEE Transactions on Image Processing*, 10(8):1200–1211.

Ballester, C., Calderero, F., Caselles, V., and Facciolo, G. (2014). Multiscale analysis of similarities between images on riemannian manifolds. *SIAM Journal Multiscale Modeling and Simulation*, 12(2):616–649.

Ballester, C. and Gonzalez, M. (1998). Affine invariant texture segmentation and shape from texture by variational methods. *Journal of Mathematical Imaging and Vision*, 9(2):141–171.

Barnes, C., Shechtman, E., Finkelstein, A., and Goldman, D. B. (2009). PatchMatch: a randomized correspondence algorithm for structural image editing. In *ACM SIGGRAPH 2009 Papers*, pages 1–11. ACM.

Barnes, C., Shechtman, E., Goldman, D. B., and Finkelstein, A. (2010). The generalized PatchMatch correspondence algorithm. In *European Conference on Computer Vision*.

Bertalmío, M., Sapiro, G., Caselles, V., and Ballester, C. (2000). Image inpainting. In *Proceedings of SIGGRAPH,*, pages 417–424.

Buades, A., Coll, B., and Morel, J.-M. (2005). A non local algorithm for image denoising. In *Proceedings of the IEEE Conference on CVPR*, volume 2, pages 60–65.

Cao, F., Gousseau, Y., Masnou, S., and Prez, P. (2011). Geometrically guided exemplar-based inpainting. *SIAM Journal on Imaging Sciences*, 4(4):1143–1179.

Chan, T. and Shen, J. H. (2001a). Mathematical models for local nontexture inpaintings. *SIAM J. App. Math.*, 62(3):1019–43.

Chan, T. and Shen, J. H. (2001b). Nontexture inpainting by curvature-driven diffusions. *Journal of Visual Communication and Image Representation*, 12(4):436–449.

Criminisi, A., Pérez, P., and Toyama, K. (2004). Region filling and object removal by exemplar-based inpainting. *IEEE Trans. on IP*, 13(9):1200–1212.

Demanet, L., Song, B., and Chan, T. (2003). Image inpainting by correspondence maps: a deterministic approach. *Applied and Computational Mathematics*, 1100:217–50.

Drori, I., Cohen-Or, D., and Yeshurun, H. (2003). Fragment-based image completion. In *ACM SIGGRAPH 2003 Papers*, volume 22, pages 303–12.

Efros, A. A. and Leung, T. K. (1999). Texture synthesis by non-parametric sampling. In *Proceedings of the IEEE ICCV*, pages 1033–38.

Fedorov, V., Arias, P., Sadek, R., Facciolo, G., and Ballester, C. (2015). 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.

Foi, A. and Boracchi, G. (2012). Foveated self-similarity in nonlocal image filtering. *Human Vision and Electronic Imaging XVII*, 8291(1):829110.

Garding, J. and Lindeberg, T. (1994). Direct estimation of local surface shape in a fixating binocular vision system. In *Eklundh, Lecture Notes in Computer Science*, pages 365–376.

Gilboa, G. and Osher, S. J. (2008). Nonlocal operators with applications to image processing. *Multiscale Modeling and Simulation*, 7(3):1005–1028.

Gårding, J. (1992). Shape from texture for smooth curved surfaces in perspective projection. *Journal of Mathematical Imaging and Vision*, 2(4):327–350.

Gårding, J. and Lindeberg, T. (1996). Direct computation of shape cues using scale-adapted spatial derivative operators. *International Journal of Computer Vision*, 17(2):163–191.

Hays, J. and Efros, A. (2007). Scene completion using millions of photographs. In *SIGGRAPH*, New York, NY, USA. ACM.

Huang, J.-. B., Kopf, J., Ahuja, N., and Kang, S. B. (2013). Transformation guided image completion. In *International Conference on Computational Photography*, pages 1–9.

Huang, J. B., Kang, S. B., Ahuja, N., and Kopf, J. (2014). Image completion using planar structure guidance. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2014)*, 33(4):129:1–129:10.

Kawai, N., Sato, T., and Yokoya, N. (2009). Image inpainting considering brightness change and spatial locality of textures and its evaluation. In *Advances in Image and Video Technology*, pages 271–282.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Mansfield, A., Prasad, M., Rother, C., Sharp, T., Kohli, P., and van Gool, L. (2011). Transforming image completion. In *Proceedings of BMVC*, pages 121.1–121.11.

Masnou, S. (2002). Disocclusion: a variational approach using level lines. *IEEE Transactions on Image Processing*, 11(2):68–76.

Masnou, S. and Morel, J.-M. (1998). Level lines based disocclusion. In *Proceedings of IEEE ICIP*, volume 3, pages 259–263.

Matas, J., Chum, O., Urban, M., and Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767.

Mikolajczyk, K. and Schmid, C. (2004). Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86.

Pavić, D., Schonefeld, V., and Kobbelt, L. (2006). Interactive image completion with perspective correction. *The Visual Computer*, 22(9-11):671–681.

Peyré, G. (2009). Manifold models for signals and images. *Computer Vision and Image Understanding*, 113(2):249–260.

Pizarro, L., Mrázek, P., Didas, S., Grewenig, S., and Weickert, J. (2010). Generalised nonlocal image smoothing. *International Journal of Computer Vision*, 90:62–87.

Wang, Z. (2008). Image affine inpainting. In *Image Analysis and Recognition*, volume 5112 of *Lecture Notes in Computer Science*, pages 1061–1070.

Wexler, Y., Shechtman, E., and Irani, M. (2007). Space-time completion of video. *IEEE Transactions on PAMI*, 29(3):463–476.