

Needle-Match: Reliable Patch Matching under High Uncertainty

Or Lotan

Michal Irani

Dept. of Computer Science and Applied Mathematics
The Weizmann Institute of Science, ISRAEL

Abstract

Reliable patch-matching forms the basis for many algorithms (super-resolution, denoising, inpainting, etc.) However, when the image quality deteriorates (by noise, blur or geometric distortions), the reliability of patch-matching deteriorates as well. Matched patches in the degraded image, do not necessarily imply similarity of the underlying patches in the (unknown) high-quality image. This restricts the applicability of patch-based methods. In this paper we present a patch representation called “Needle”, which consists of small multi-scale versions of the patch and its immediate surrounding region. While the patch at the finest image scale is severely degraded, the degradation decreases dramatically in coarser needle scales, revealing reliable information for matching. We show that the Needle is robust to many types of image degradations, leads to matches faithful to the underlying high-quality patches, and to improvement in existing patch-based methods.

1. Introduction

Finding similar patches is a fundamental component in many computer vision algorithms. These include super resolution [12, 14], texture synthesis [11], image completion [16], image retargetting [5, 26], and many more. However, when the image quality undergoes degradation (noise, blur, geometric deformations, etc.), the reliability of patch-matching decreases dramatically. Finding two “similar” patches $\tilde{p} \approx \tilde{q}$ in the degraded image, may no longer imply that their underlying (hidden) clean patches p and q are similar. This however, is the very basic assumption in most patch-based methods. For example, *unconstrained* matching of noisy patches (globally in the noisy image, or in a database of clean images), was shown to suffer from severe overfit of the noise [25, 34]. Similarly, severe motion blur or geometric distortions harm the ability to find good matches (correspondences) across frames, hence optical flow estimation algorithms tend to fail under such degradations.

Partial solutions have been proposed for this problem, which offer only a limited remedy. For example, denoising algorithms [4, 7] typically restrict the search space to a

small image region around each patch, or use larger patches which are less prone to noise-overfit. This, however, restricts the ability to recover fine textures. Patch matching under severe blur was addressed in [3], by pre-processing the image patches to a normalized frequency range. It addresses only isotropic blur, and does not generalize to other blur types (motion blur, non-parametric blur).

This paper presents a general framework for reliable patch matching under severe image degradations. The “Needle”, a new patch descriptor, comprises of small patches obtained from multiple image scales, at the same relative image coordinates (Fig. 1a). While the patch at the finest image scale is severely degraded, hence cannot be reliably matched, the degradation decreases dramatically in coarser Needle scales, revealing reliable information for matching (Fig. 1(c,d,e)). We show that if $Needle(\tilde{p}) \approx Needle(\tilde{q})$ for two distorted patches \tilde{p} and \tilde{q} , then with high likelihood $p \approx q$ (the hidden high-quality patches).

Our paper has two main contributions: (i) A new patch descriptor which is *resilient* to many types of image degradations, (ii) Using *NeedleMatch* in patch-based algorithms (e.g., denoising, correspondence-estimation) leads to a substantial improvement under severe image degradations.

The term “Needle” was coined in [35], but for a different purpose. Each noisy patch was matched against coarser patches *inside its own needle*, looking for its cleanest version among its own needle patches. In contrast, we employ the Needle as a ‘descriptor’, which is matched against other needle-descriptors, to find good matches *elsewhere* in the image (or in other images). Contrary to [35], we do not rely on cross-scale recurrence of patches; our coarser-scale needle patches need *not* resemble the fine-scale patch (Fig. 1b).

Our Needle is different than other image descriptors (e.g., SURF, SIFT, Geometric-Blur), as its aim is to represent well the *unknown* high-quality patch signal hidden under severe degradation. A dense multiscale descriptor was also developed by [15], based on SIFT computed at multiple scales. Their aim was different – scale-invariant matching. Being a gradient-based descriptor, their descriptor is likely to fail in the presence of severe image degradation.

Incorporating information from multiple scales is commonly used in coarse-to-fine algorithms (optical flow [20], stereo [24, 33], etc.) However, these methods use multiscale

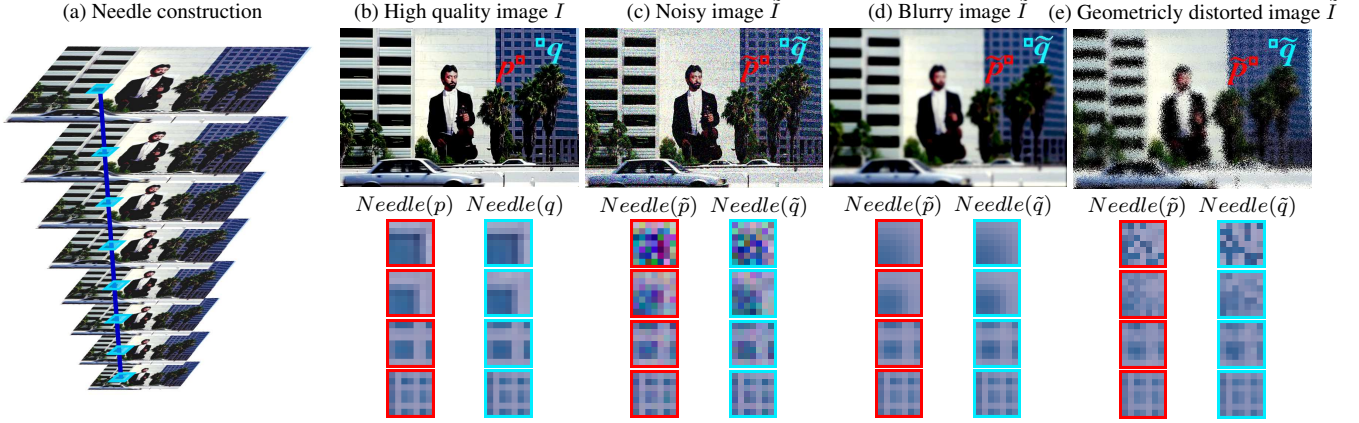


Figure 1: **The Needle Descriptor.** Needles of degraded patches (e.g., $Needle(\tilde{p})$, $Needle(\tilde{q})$) reveal high-quality signal at their coarser scales. These can be used to reliably guide the matching of degraded patches (\tilde{p} and \tilde{q}) at the original scale.

information *sequentially* (coarse to fine). Such sequential approaches fail under severe degradations (see Sec. 5.2.2), since their final optimization stage is misguided by the *last-visited but most-degraded finest scales*. In contrast, The Needle employs information from all scales *simultaneously*. Some multiscale cost aggregation methods also employ simultaneous multiscale information (e.g., [19, 32]). However, these are special purpose algorithms designed to solve a specific application. In contrast, *NeedleMatch* can be plugged into a wide variety of *existing* patch-based algorithms, simply by replacing their patch-based similarity with Needle-based similarity, thus making them robust to a wide variety of image degradations (Fig 2).

The rest of the paper is organized as follows. Sec. 2 presents the Needle construction. Sec. 3 analyzes the source of its resilience to severe degradations. Sec. 4 analyzes why the Needle is able to accurately lock onto fine details despite using large contextual information at coarser scales. Sec. 5 shows the superiority of the Needle matching over regular patch matching in various patch-based applications.

2. “Needle” - A Multiscale Patch Descriptor

Let I be a high-quality image, and \tilde{I} be its degraded version. $\tilde{p} \in \tilde{I}$ is a degraded version of the patch $p \in I$. When seeking a “similar” patch $\tilde{q} \in \tilde{I}$, we wish to find a patch which is similar in its *underlying (unknown) high-quality signal*, i.e., $q \approx p$. The difficulty in finding reliable matches stems from the fact that the “*distortion component*” in a degraded patch is often larger than its “*signal component*” p (i.e., $\|\tilde{p} - p\| > \|p\|$). This leads to an overfit of the distortion component instead of the signal, resulting in bad patch matches. In other words, a good fit of \tilde{q} to \tilde{p} in image \tilde{I} *does not necessarily imply a good fit of q to p in image I* .

We compute a Needle-descriptor for each image patch, such that if $Needle(\tilde{p}) \approx Needle(\tilde{q})$, then with high likelihood $p \approx q$. To build the descriptor, the image \tilde{I} is down-scaled to generate multiscale images $\{\tilde{I}_s\}$ (we used a bicu-

bic downscaling kernel). $Needle(\tilde{p}) = \{\tilde{p}_s\}$ is obtained by stacking all the patches $\tilde{p}_s \in \tilde{I}_s$ which reside at the *same relative image coordinates* as \tilde{p} ; i.e., if (x, y) are the center coordinates of \tilde{p} in \tilde{I} , then (sx, sy) are the center coordinates of the patch \tilde{p}_s in \tilde{I}_s , at *sub-pixel accuracy* (Fig. 1a).

In order to match image patches, a Nearest-Neighbor (NN) search is conducted in the *needle space*. For two corrupted patches \tilde{p} and \tilde{q} to match, their corresponding needles $\{\tilde{p}_{s_\ell}\}_{\ell=0}^L$ and $\{\tilde{q}_{s_\ell}\}_{\ell=0}^L$ must match (have low ℓ_2 distance). We refer to this process as *NeedleMatch*. Note that \tilde{p} need not necessarily be of same size as its needle patches $\{\tilde{p}_{s_\ell}\}$. For example, a 5×5 patch \tilde{p} at the original scale may be represented using a multi-scale needle of 3×3 patches. In our experiments we typically used needles of 3×3 multi-scale patches, obtained from 8 scales $s_\ell = 0.75^\ell$ ($\ell = 0, 1, \dots, 7$).

Three Needle properties lead to reliable patch matching:

(a) High-quality signal is revealed at coarse Needle scales:

While the patch at the finest scale of a degraded image is severely distorted, the degradation decreases dramatically in coarser scales. Additive Gaussian noise, general image blur and general geometric distortions – all decrease linearly with scale, *revealing high-quality patches at coarser needle scales* (Fig. 1.c,d,e). These claims are derived analytically in Sec. 3. Note that the *different* types of degradations in Fig. 1 reveal *similar* signals in their coarser scales. This is a strong property of the Needle, as *NeedleMatch* need not assume a specific degradation model for its search. In fact, *NeedleMatch* can be used to match across images with *different types of degradation* (e.g., matching patches in a noisy image to patches in a blurry image).

(b) Context without penalty: *NeedleMatch* benefits from local context information, without suffering from the inherent limitation of matching large rigid regions. Wider and wider image regions around the patch are represented by increasingly coarser scales of its needle. This provides reliable context information for matching. However, one would

expect that this should harm the ability of the Needle to find accurate well-localized matches in textured regions (e.g., tree leaves, bushes, hair, etc.) Surprisingly, this is not the case. *NeedleMatch* provides very good matches in such regions (e.g., see the tree leaves in Fig. 6h). Sec. 4 explains and analyzes this surprising phenomenon. We show that *NeedleMatch* gives rise to patch matches which only *roughly* share the same local context. It permits non-rigid deformations of the surrounding region, while enforcing high match accuracy of the center patch. Such deforming contexts cannot be exploited by matching *large rigid patches*, as these are known to lack good matches [18].

(c) Simultaneous use of all scales: The needle does not commit to any specific scale, but rather uses all scales simultaneously. This has several benefits: (i) Different image regions may benefit from using different scales, depending on the amount of details they contain. (ii) We do not know a-priori the degree of blur/noise/distortion, hence do not know ahead of time which scale is most informative. (iii) Coarse needle levels provide high-quality signal information and *rough* context information, thus guaranteeing *high match reliability*. Fine needle levels, on the other hand, allow to match the fine details of the signal (e.g., texture, edges), thus guaranteeing *high localization accuracy*.

These theoretical benefits of the needle are derived analytically in Sec. 3 and 4, and are supported by the quantitative and qualitative experiments in Sec. 5. We further show the superiority of *NeedleMatch* over: (i) Patch matching at the original scale alone (of either small or large patches), and (ii) superiority over *sequential* coarse-to-fine patch matching. In particular, *under severe degradation*, *NeedleMatch* provides more reliable correspondences than advanced coarse-to-fine optical-flow methods [28] (Sec. 5).

3. Reliable Signal is Revealed at Coarse Scales

We next show that signal uncertainty, which is high in fine scales of the degraded image, reduces dramatically in coarser scales, revealing high quality signal. This is what provides *NeedleMatch* with its robustness.

(a) Noise Uncertainty Reduces with Scale: Let $\tilde{p} = p + n$ be a noisy patch. Let $\tilde{q} = q + \hat{n}$ be a ‘Nearest-Neighbor’ (NN) of \tilde{p} . When the noise n in the patch is larger than its signal p (e.g., in uniform patches p), the *patchwise* Signal-to-Noise Ratio [25]: “PatchSNR” = $\sqrt{\frac{\text{var}(p)}{\text{var}(n)}}$ is low. This leads to *overfit of the noise* instead of the signal (See [25]).

While the noise severely corrupts the original image scale, the noise level drops significantly at coarser image scales. For additive Gaussian noise: $\tilde{I} = I + N(0, \sigma^2)$, the noise in an α -times coarser image will be distributed like [35]: $n^{(\alpha)} \sim N(0, (\frac{\sigma}{\alpha})^2)$, i.e., α -times smaller. Thus, $\text{PatchSNR}^{(\alpha)}(\tilde{p}) \geq \sqrt{\frac{\text{var}(p)}{\text{var}(n^{(\alpha)})}} = \alpha \cdot \text{PatchSNR}(\tilde{p})$, revealing more reliable signal for matching (Fig. 1c). The

average PatchSNR of the needle is thus much higher. This significantly lowers the chance to overfit noise, leading to reliable patch matching (see Figs. 1c, 2c).

(b) Blur Uncertainty Reduces with Scale: Let $\tilde{I} = I * b$ be a blurry image obtained by convolving a sharp image I with a severe blur b . The difficulty in finding reliable matches stems from the fact that the blur b eliminates much of the high frequency information of I , leaving patches in \tilde{I} with very few details for reliable matching.

When the blurry image \tilde{I} is scaled down by a factor of $\alpha > 1$, its blur also decreases by a factor of α . For example, an edge smeared over 10 pixels in \tilde{I} would appear smeared over only 5 pixels in its half-scaled version $(\tilde{I} \downarrow_2)(x, y) = \tilde{I}(2x, 2y)$. More formally [23], for any scaling factor α : $\tilde{I} \downarrow_\alpha(x, y) = I \downarrow_\alpha(x, y) * (\alpha^2 \cdot b \downarrow_\alpha(x, y))$. Note that $\alpha^2 \cdot b \downarrow_\alpha(x, y) = \alpha^2 \cdot b(\alpha x, \alpha y)$ is α -times narrower in x and in y , and α^2 -times taller than the original blur $b(x, y)$, thus maintaining an integral of 1. Since $\lim_{\alpha \rightarrow \infty} (\alpha^2 \cdot b \downarrow_\alpha(x, y)) = \delta(x, y)$, significant down-scaling of a blurry image will result in a sharp image with no blur (an observation also used in Blind-Deblurring [23, 29]). Thus, *blur uncertainty reduces with needle scales* (Fig. 1d), resolving match ambiguities.

(c) Geometric Distortion Reduces with Scale: While noise and blur introduce uncertainty in *pixel intensities*, we next analyze uncertainty in *pixel locations*, namely *geometric uncertainty*. Such uncertainty can be caused, e.g., by atmospheric turbulence, imaging through smoke, underwater photography, etc. These distortions presents a challenging task for patch matching.

However, a pixel dislocated by Δ in the original scale, is dislocated by only $\frac{1}{\alpha} \Delta$ in an α -times coarser image. Thus, *the geometric uncertainty reduces with needle scales*, leading to reliable signal-fit (Fig. 1e). For our empirical evaluations (Sec. 5.1), we used a simplistic model to geometrically distort images, by randomly displacing each pixel using a 2D-Gaussian distribution. A pixel $p \in I$, located at (x, y) , is randomly displaced to $(x + \Delta_x, y + \Delta_y)$ where $\Delta_x, \Delta_y \sim N(0, \sigma)$. This process results in severely degraded images (e.g., Figs. 1e, 6i, for $\sigma = 4.5$).

4. Context without Penalty

We saw that cleaner/sharper signals emerge in coarser needle-scales, resulting in Needles with much higher SNR than their degraded source patches. *But why should the Needle, which captures information from a large image region (substantially larger than the patch), be able to find good and well-localized patch matches of fine structures?* After all, even if two patches p and q are good matches, their surrounding contexts P and Q may be quite different (see Fig. 3a). This section shows that *NeedleMatch* can match patches well, despite some context variations. In fact, *contexts usually help improve the matches*.

Context information has been shown to significantly im-

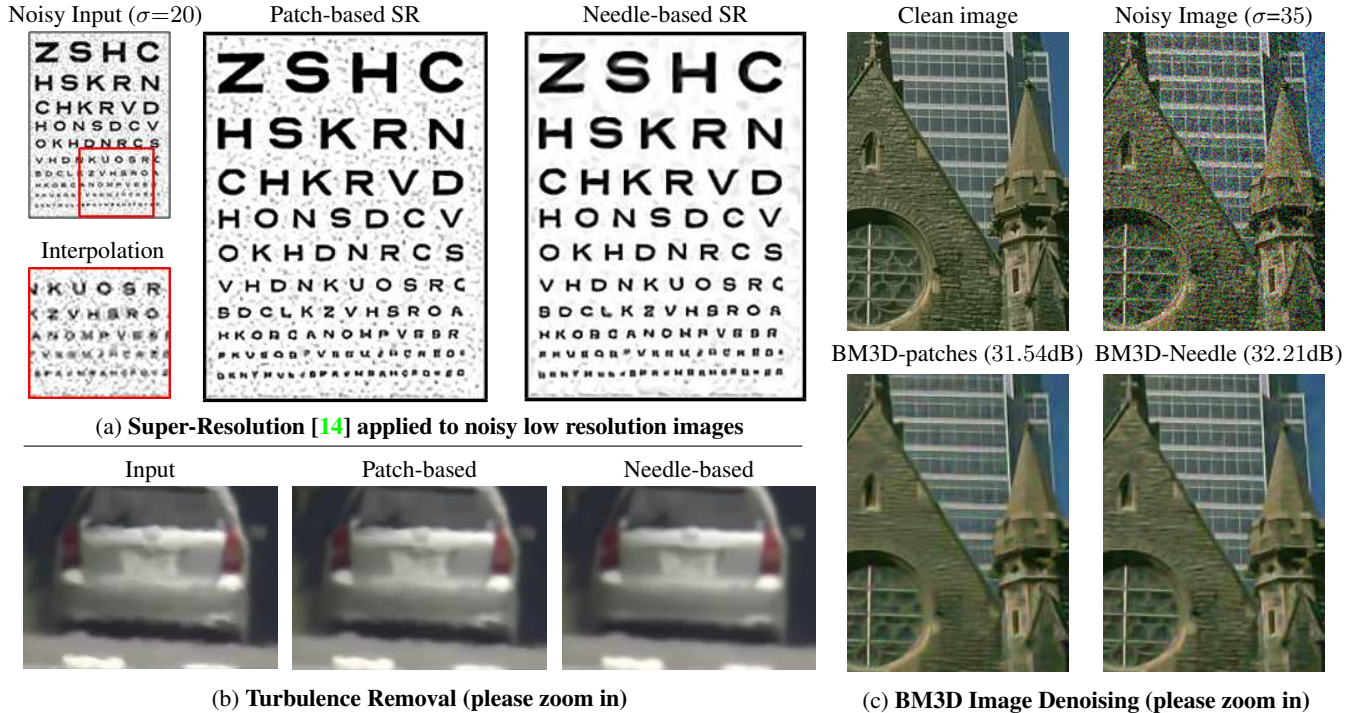


Figure 2: **Replacing ‘patch-based similarity’ with ‘Needle-based similarity’ improves a variety of patch-based applications.** (a) SR; (b) Turbulence removal (See Sec. 5.2 for details); (c) BM3D denoising. Please zoom in to see details.

prove the performance of detection and recognition [8, 9, 10, 31], super-resolution [30], and is the guiding principle behind the fast PatchMatch method of [1]. *NeedleMatch* benefits from this property. When certainty is low (e.g., under degradation), the context information guides the search for NNs to more likely image regions. For example, when searching for a leave patch (e.g., the small red patch in Fig. 3a), we are more likely to find reliable matches for it among other leave patches, than among building patches. However, two similar leave patches may have quite different contexts, despite sharing similar appearance properties (see two larger orange regions in Fig. 3a). This prohibits the ability to match the two small leave patches via matching their larger context regions. We next show that larger context regions, which share similar appearance properties, but differ at finer image scales, are *very similar in coarser Needle scales*. Thus, coarse needle scales guide the search to reliable matches in *similar*, but not identical contexts. This phenomenon is explained and analyzed next.

Let p and q be two corresponding patches of size $m \times m$, whose surrounding $M \times M$ regions P and Q (their “contexts”) deform with respect to each other (Fig. 3a). Since P and Q share similar appearance properties (similar colors/intensities), we can describe one of them as a non-rigid deformation of the other. The arrows (u, v) in Fig. 3b illustrate a deformation between the contexts P and Q (namely, how far a pixel from P has to move in order to find a pixel in Q with similar color). Let’s assume that $Needle(p) = \{p_{s_\ell}\}_{\ell=1}^L$ and $Needle(q) = \{q_{s_\ell}\}_{\ell=1}^L$

are composed of patches of size $m \times m$ (where $m \ll M$, e.g., $m=3$ and $M=30$). Let’s further assume that the coarsest needle-patches p_{s_L} and q_{s_L} captures information from their entire surrounding contexts P and Q , respectively. Namely, $P \downarrow_\alpha = p_{s_L}$ and $Q \downarrow_\alpha = q_{s_L}$, where $\alpha=1/s_L$.

We analyze the case where p and q are clean image patches that match well (i.e., if placed on top of each other, the two patches will be well aligned, with very similar colors). However, their surrounding context regions P and Q are *not* well aligned. We assume that the contexts do not change drastically, but rather deform gradually with respect to each other (an assumption which holds for the vast majority of the patches that do not lie near occlusion boundaries). In other words, we allow for increasing deformation (larger “relative misalignments” (u, v) between P and Q) as a function of the distance from their centers (Fig. 3b). We make the simplifying assumption that the *magnitudes* of these “relative misalignments” are *proportional* to their distance from the region center $(0, 0)$:

$$\|(u(x, y), v(x, y))\| = t \cdot \|(x, y)\| \quad (1)$$

for some scalar t . Note that this assumption holds not only for gradual non-rigid deformations, but also for contexts deforming by global *linear* parametric transformations (e.g., zoom, rotation, affine deformation). We next show that although the average misalignment is large for the contexts P and Q (Claim 1), it is small for the corresponding Needles of p and q (Claim 2), leading to good matches.

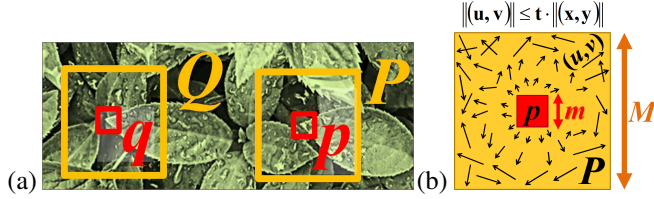


Figure 3: **Similar patches with different contexts.** (a) The patches p and q are similar but their surrounding contexts P and Q are not. (b) “Relative misalignments” between the two contexts increase with distance from their centers.

Claim 1 [Context Misalignment:] Under the assumption of Eq. (1), the average misalignment per-pixel between two context regions of size $M \times M$ is:

$$\text{AvgMisalignment}(P, Q) = \frac{Mt}{\sqrt{6}} \quad (2)$$

Proof: For simplicity, we perform the derivation in the continuum. The average misalignment per pixel between P and Q is:

$$\begin{aligned} & \text{AvgSquareMisalignment}(P, Q) = \\ &= \frac{1}{M^2} \int_{-M/2}^{M/2} \int_{-M/2}^{M/2} (u(x, y)^2 + v(x, y)^2) dx dy \\ &= \frac{1}{M^2} \int_{-M/2}^{M/2} \int_{-M/2}^{M/2} t^2(x^2 + y^2) dx dy = \frac{M^2 t^2}{6} \end{aligned}$$

Therefore, $\text{AvgMisalignment}(P, Q) = \frac{Mt}{\sqrt{6}}$ ■

Large contexts P and Q have large misalignment. E.g., for $M=25$ (25×25 regions) and $t=0.2$ (e.g., zoom by 0.8), the average pixel misalignment between P and Q is ≈ 2 pixels. This results in a large grayscale difference, especially in textured regions, hence a bad match between P and Q .

In contrast, when the *needles* of two matching patches p and q are compared, $\text{Needle}(p) = \{p_{s_\ell}\}_{\ell=1}^L$ and $\text{Needle}(q) = \{q_{s_\ell}\}_{\ell=1}^L$, their average misalignment will be significantly smaller. The misalignment between corresponding needle-patches at coarser scales will still grow linearly with distance from the patch centers, with the same scalar t . When two image regions are scaled-down by a factor of α , their relative misalignments become α -times smaller. In our case: $\|(u \downarrow_\alpha(x, y), v \downarrow_\alpha(x, y))\| = \frac{1}{\alpha} \|(u(\alpha x, \alpha y), v(\alpha x, \alpha y))\| = \frac{1}{\alpha} t \|(\alpha x, \alpha y)\| = t \|(x, y)\|$. Hence, the scalar t remains the same in all scales. Nevertheless, the resulting average misalignment per needle-pixel will be much smaller, since all its patches are small ($m \times m$). This is shown next:

Claim 2 [Needle Misalignment:] Under the assumption of Eq. (1), the average misalignment per needle-pixel between the two Needles of patch-size $m \times m$ is:

$$\text{AvgMisalignment}(\text{Needle}(p), \text{Needle}(q)) = \frac{mt}{\sqrt{6}} \quad (3)$$

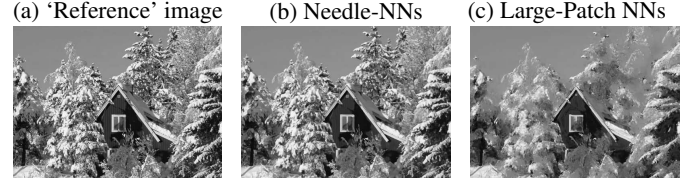


Figure 4: **Matching Needles vs. Matching Large Patches.** A sharp ‘reference’ image (a) was zoomed by 0.75 to obtain a ‘target’ image. Reference patches searched for their 1st NN in the target image. Reconstructed images (b and c) were generated using the grayscale of the center pixel of each NN. Needle-NNs find good matches even in textured regions, whereas 25×25 patches do not (Please zoom).

Proof:

$$\begin{aligned} & \text{AvgSquareMisalignment}(\text{Needle}(p), \text{Needle}(q)) = \\ &= \frac{1}{L} \sum_{\ell=1}^L \text{AvgSquareMisalignment}(p_{s_\ell}, q_{s_\ell}) \\ &= \frac{1}{L} \sum_{\ell=1}^L \left(\frac{1}{m^2} \int_{-m/2}^{m/2} \int_{-m/2}^{m/2} (t^2(x^2 + y^2) dx dy) \right) \\ &= \frac{1}{L} \sum_{\ell=1}^L \left(\frac{m^2 t^2}{6} \right) = \frac{m^2 t^2}{6} \\ &\Rightarrow \text{AvgMisalignment}(\text{Needle}(p), \text{Needle}(q)) = \frac{mt}{\sqrt{6}} \quad \blacksquare \end{aligned}$$

The expression for the average misalignment per needle-pixel is very similar to that per context-pixel ($\frac{mt}{\sqrt{6}}$ vs. $\frac{Mt}{\sqrt{6}}$), with the important distinction that $m \ll M$. It is interesting to note that the average misalignment per needle-pixel does not depend on the number L of levels (patches) along the needle, nor on the size M of the spatial context P covered by the needle (as long as Eq. (1) holds for the entire context P). It only depends on the size m of the needle patches (which is typically small), and on the rate t of non-rigid distortion. For example, if the needle patches are of size 3×3 ($m=3$), and $t=0.2$ (same as above), the average misalignment error per pixel between $\text{Needle}(p)$ and $\text{Needle}(q)$ will be ≈ 0.25 pixel – an order of magnitude smaller than the average misalignment between their contexts P and Q !

The small misalignment leads to high similarity between the two needles. E.g., for textured regions with gradient magnitude of ~ 10 grayscales per pixel, a rough calculation (assuming linearized brightness constancy) yields $\text{RMSE}(\text{Needle}(p), \text{Needle}(q)) \approx \|\text{grad}\| \cdot \|(u, v)\| = 10 \cdot 0.25 = 2.5$ grayscales per pixel. The small error indicates a good match between the patches p and q . In contrast, the RMSE between their contexts is large: $\text{RMSE}(P, Q) \approx \|\text{grad}\| \cdot \|(u, v)\| = 10 \cdot 2 = 20$.

Note that two random needles of non-matching patches will not have small misalignments. This is because the basic assumptions (that the central patches are well aligned, and that the misalignments between their surrounding contexts grow slowly with distance from the centers), do not hold.

In summary, *NeedleMatch* benefits from local context information, without suffering from the inherent limitation of large region matching (with the exception of patches near occlusion boundaries). This is what allows patches on tree-leave in Figs. 4 and 6h to find other matching tree-leave patches. Similar tree leaves share roughly the same local context, up to mild non-rigid deformations.

Note that the assumption of Eq. (1), holds not only for gradual non-rigid deformations, but also for linear parametric transformations (e.g., rotation, zoom, affine deformations). The ability of *NeedleMatch* to find good matches under such transformations is illustrated in Fig. 4, and is further contrasted with matching large rigid patches. A ‘reference’ image (Fig. 4a) was zoomed by 0.75 ($t=0.25=1-0.75$, for all image patches¹), to generate a ‘target’ image. Each patch in the reference image searched for its *best* match (1st NN) in the target image, once using *NeedleMatch* and once using large 25×25 rigid patches (the spatial support covered by the needles). New images were reconstructed (Fig. 4b,c), by replacing each pixel in the ‘reference’ image with its corresponding pixel in the ‘target’ image, according to the computed NNs. Needle-NNs find good matches, even in highly textured regions (e.g., tree leaves), whereas large patches do not, resulting in a blurry unwarped image (please zoom in). This experiment was repeated for a rotation by 15° ($t \approx 0.26$ for all patches), yielding similar results.

5. Experiments

A wide variety of algorithms rely on patch similarity to solve computer vision tasks. These include image denoising [4, 7], image deblurring [23], super resolution [14], optical flow estimation [20], inpainting [6] and more. The success of these algorithms relies on the assumption that when we search for “similar patches” (in the same image, or in other images), we find other patches with the same signal. This assumption, however, is not true when the image (signal) undergoes degradations. In such cases, patch-based similarity often leads to patches with *similar degradations*, instead of *similar signal*, resulting in high reconstruction errors in the above algorithms. *NeedleMatch*, on the other hand, provides more accurate “signal-fit” in the presence of degradations. Thus, replacing the patch-based similarity with needle-based similarity leads to improved results².

We first compare the ability of regular *patch matching* vs. *NeedleMatch* to find good NNs (patches with *similar underlying signal*). Sec. 5.1 presents such quantitative evaluation on hundreds of images, undergoing different degradation types. We then show that replacing *patch-based similar-*

¹ Note that zoom around the origin $(0, 0)$ by a scale factor α , can be rewritten as zoom by α around *any* other point in the image (e.g., any patch center), plus a global image shift. The same holds also for image rotation.

² Average runtimes per 321×481 image (on a single CPU): When using KDtree-based search – 5.2s for matching patches, and 16s for the needles. When using an efficient PatchMatch-like implementation [1] for finding NNs, runtimes decrease to 0.8s for patches, and 1.5s for Needles.

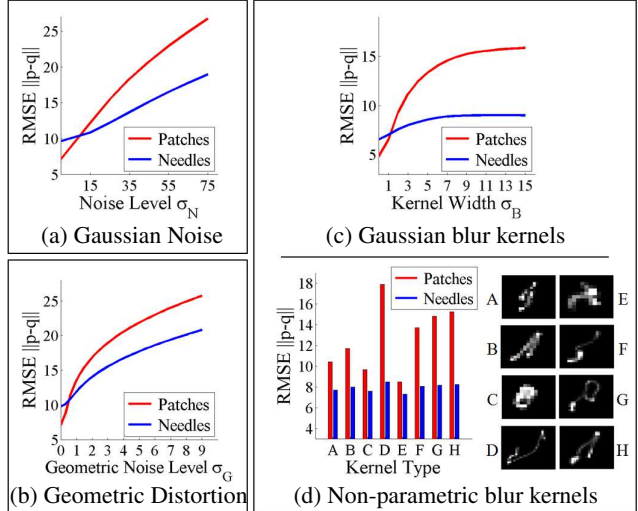


Figure 5: **Reliability of “signal fit” in degraded images.** Matches are searched for corrupted patches ($\tilde{q} = NN(\tilde{p})$), but errors $\|p - q\|$ are measured between the true underlying signals. (Red for Patch-NNs; Blue for Needle-NNs). Already for small levels of degradation, *NeedleMatch* shows a dramatic improvement over regular patch-matching.

ity with *Needle-based similarity* in a variety of patch-based applications, leads to improved results (Sec. 5.2).

5.1. Empirical Evaluation: “Signal Fit” Accuracy

We compare the *match-quality* of patch matching vs. *NeedleMatch* under 3 types of degradation (noise, blur, geometric distortion). Given a degraded 5×5 patch \tilde{p} , we search for a similar patch $\tilde{q} = NN(\tilde{p}, \tilde{I})$ among all other patches in the degraded image \tilde{I} (omitting self matches). The “signal-fit” between the degraded patches \tilde{p} and \tilde{q} is assessed by:

$$SignalMatchError(\tilde{p}, \tilde{q}) = \|p - q\|,$$

where p and q denote the ground-truth *high quality* patches underlying \tilde{p} and \tilde{q} in the ground-truth clean image I .

The graphs in Fig. 5 report *SignalMatchError*, averaged over all 5×5 patches from hundreds of degraded images (degradations applied to the DB of [21] and [29]). Statistics were collected at varying degradation levels: (i) Noisy images were generated with varying Gaussian noise levels ($\sigma_N = 0, \dots, 75$). (ii) Blurry images were generated via convolution with Gaussian blurs of varying width ($\sigma_B = 0, \dots, 15$), as well as with the 8 non-parametric blur kernels of [29]. (iii) Geometrically deformed images were generated by *randomly displacing* pixels with 2D-Gaussian perturbations ($\sigma_G = 0, \dots, 9$). The NN search used KDtree [13], applied once in the space of 5×5 degraded patches ($\tilde{q} = NN(\tilde{p}, \tilde{I})$), and once in the space of $3 \times 3 \times 8$ needles emerging from degraded patches ($\tilde{q} = NN_{needle}(\tilde{p}, \tilde{I})$). The graph indicates that for patches with no degradation ($\sigma_N = \sigma_B = \sigma_G = 0$), regular patch matching is better than *NeedleMatch* (due to better matches near occluding boundaries). However, *already*

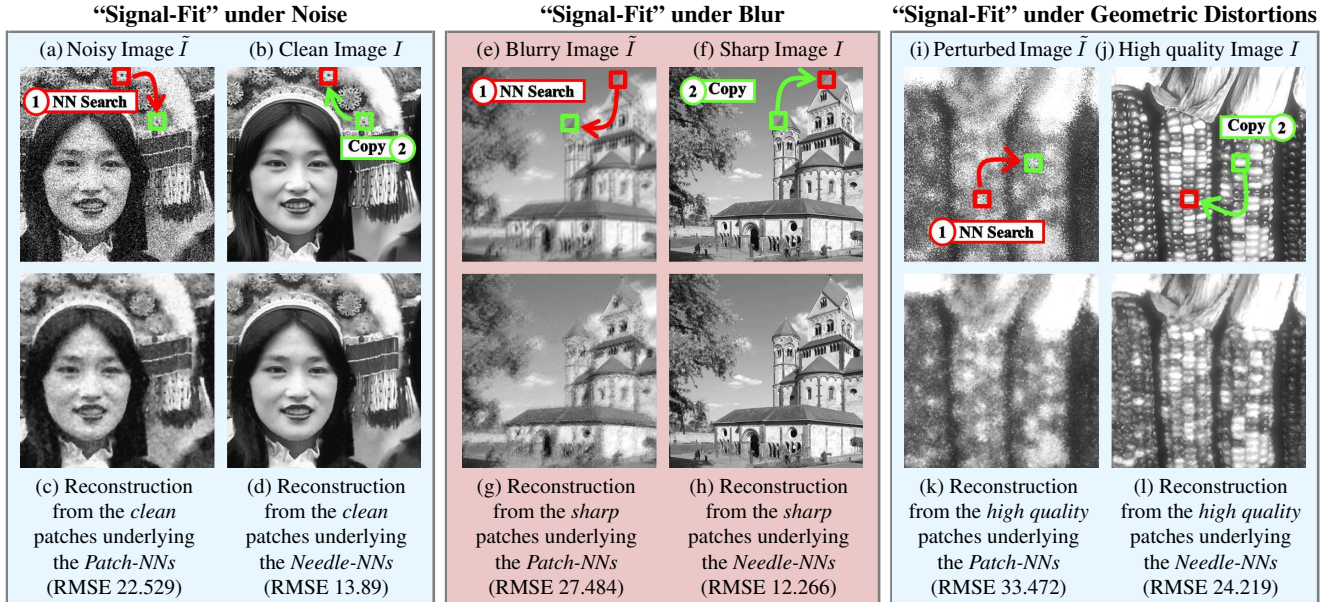


Figure 6: Visual example of “signal-fit” errors in degraded images (Patch-NNs vs. Needle-NNs). NNs are searched in the corrupted image (upper-left), but the reconstructed image is built from the underlying high-quality patches in the ground-truth image (upper-right). Patch-NNs lead to bad “signal-fit”, hence to poor reconstruction (bottom-left), even though built from high quality patches. Needle-NNs provide much better “signal-fit”, hence much better reconstruction (bottom-right) (Please zoom-in). Please note that this is NOT an application; only a visualization of the error shown in the graphs of Fig. 5.

at small levels of degradations, there is a dramatic improvement in “signal-fit” for NeedleMatch over regular patch matching (in all 3 degradation types).

Fig. 6 is a visualization of the signal-fit errors of Fig. 5. It shows image reconstructions based on the NNs found. The NNs were searched in the corrupted images (e.g., Fig. 6a for noise), but the reconstructed images were built from the underlying clean patches (Fig. 6b). When the “signal-fit” is good, the reconstructed images should resemble the ground-truth images (as in Fig. 6d); but when the “signal-fit” is poor, the reconstructed images are of low quality (Fig. 6c), even though they were constructed from clean patches. NeedleMatch provides much higher signal-fit reliability than regular patch-matching, even in highly textured regions (e.g., the hat in Fig. 6d, the tree leaves in Fig. 6h). We next show a few example applications of NeedleMatch.

5.2. Example Applications

We examined a variety of patch-based reconstruction methods, once using their regular patch-based similarity, and once using our Needle-based similarity. The latter significantly improves the results in the presence of degradation. Figs. 2 and 7 show visual results for a variety of such applications, whereas Secs. 5.2.1 and 5.2.2 provide a more extensive empirical evaluation for two of these applications.

Fig. 2a shows results of applying the Single-Image Super-Resolution (SR) method of [14] to a low resolution image with mild Gaussian noise ($\sigma=20$). The goal in SR

is to reconstruct new high frequency details which are invisible in the low-res image (beyond the Nyquist limit). However, SR with regular patch-matching tries to increase not only the details of the signal, but also the noise [27]. NN-search for noisy uniform patches tries to fit their noise (as best possible) with signal-details elsewhere in the image [25, 34]. Replacing the patch-matching in [14] with NeedleMatch, results in a super-resolved image with reconstructed high-frequency signal details, while discarding the noise (please zoom in on Fig. 2a). This is due to the ability of NeedleMatch to fit the signal but not the noise.

Even in image denoising, where the images are inherently assumed to be corrupted, employing NeedleMatch improves the results (Fig. 2c). We replaced the patch-based similarity of BM3D [7] with Needle-based similarity. The improvement is mostly visible in the fine textures (please zoom-in on Fig. 2c). More details and an extensive quantitative comparison are provided in Sec. 5.2.1 and Table. 1.

Fig. 2b shows “removal” of geometric distortion caused by heat turbulence. Every patch in the turbulent frame was replaced by the median of its k-NNs, searched in other turbulent frames in the same video. Patch-NNs preserve the geometric distortion, while Needle-NNs match the signal, allowing recovery of straight lines (e.g., see line above license plate, or at the bottom of the car – please zoom in).

Even in correspondence estimation, NeedleMatch outperforms advanced coarse-to-fine optical flow methods in the presence severe degradations (Fig. 7 and Sec. 5.2.2).

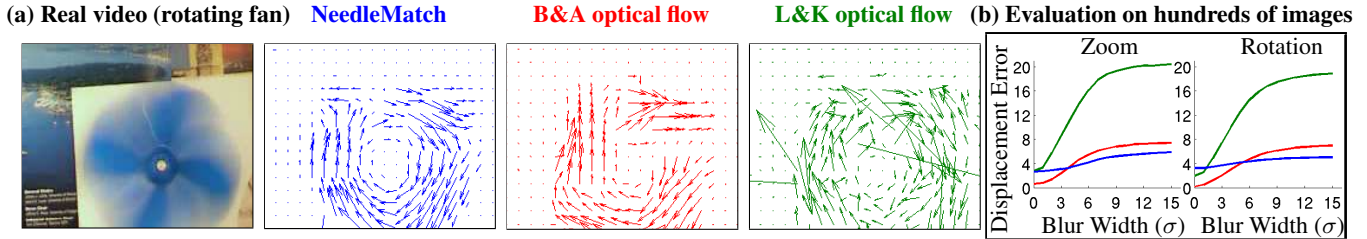


Figure 7: **Needle-correspondences vs. Coarse-to-fine Optical-Flow.** (a) A fast rotating fan induces severe motion blur. *NeedleMatch* provides better frame-to-frame correspondences on the blurry fan. (b) Errors averaged on hundreds of image pairs with varying blurs (see text). Optical flow is better for small blurs, while *NeedleMatch* is superior under severe blurs.

5.2.1 BM3D Denoising – Patches vs. Needles

As observed in [34] and in Sec. 5.1, unrestricted patch matching suffers from severe overfit of noise. Patch-based denoising methods (e.g., [4, 7]) avoid this problem by restricting their search region to a *local* area around each patch (e.g., 30×30). A small surrounding region contains many other patches with a similar (hidden) signal, while limiting the risk to overfit the noise [34]. However, this restricted search region also restricts their ability to find good matches for textured patches, resulting in blurry details.

We ran BM3D on hundreds of natural images from the dataset of [21], contaminated by Gaussian noise at varying levels ($\sigma=15, \dots, 75$). We applied the BM3D code of [17], once with patch-based similarity, and once with Needle-based similarity (We used $3 \times 3 \times 8$ -needles and 8×8 -patches for $\sigma = 15, 35$, and $5 \times 5 \times 8$ -needles and 12×12 -patches for $\sigma = 55, 75$). As expected, applying BM3D to the entire image using its regular patch-based similarity, leads to worse results (see Table. 1 - BM3D local vs. global). This deterioration is due to the noise-fitting problem of patch matching in unrestricted domains. Global *NeedleMatch*, on the other hand, avoids noise fitting, leading to an overall improvement, especially at high noise-levels (0.42-0.81dB - see Table. 1). Using a larger search space without being prone to noise-fitting, offers better reconstruction in textured regions (see Fig. 2c). We further ran the experiment also on the 20 images of BM3D benchmark for $\sigma=15, 35, 55$, obtaining an average improvement of 0.25, 0.36, 0.66dB on color images, and 0.18, 0.3, 0.51dB on the BW images.

5.2.2 Correspondence Estimation under Severe Blur

Coarse-to-fine optical flow (such as Lucas&Kanade) can be regarded as a *sophisticated* coarse-to-fine patch matching between two images. We therefore compared the robustness of *NeedleMatch* vs. coarse-to-fine optical flow, under severe blur. We compared to 2 optical flow methods – Lucas&Kanade (L&K) [20] and Black&Anandan (B&A) [2] (using the advanced implementations of [22, 28], respectively). We generated a database of hundreds of pairs of degraded images, by applying large transformations (relative zoom of 0.75 and relative rotation of 15°) to the images of [29], and then severely blurring them (one image from each pair was blurred vertically, while its transformed ver-

σ	BM3D local (regular)	BM3D global	BM3D needle	Improvement (over BM3D-regular)
15	31.04	26.91	31.19	+0.15
35	27.34	26.17	27.57	+0.24
55	24.96	24.83	25.38	+0.42
75	24.13	23.67	24.94	+0.81

Table 1: **BM3D Denoising results.** Average PSNR results for hundreds of contaminated images from [21]. *NeedleMatch* leads to an overall PSNR improvement, especially at high noise-levels. (see Sec. 5.2.1 for more details)

sion was blurred horizontally). The width of the horizontal and vertical blur kernels varied from zero-width (no blur) to 14 pixels wide (severe blur). The graphs in Fig. 7b show the resulting errors (RMSE) between the computed displacements, and the ground-truth displacements. Optical flow estimation is superior to *NeedleMatch* under small blur. However, under severe blur, *NeedleMatch* provides more accurate correspondences, even though it uses no smoothness constraint (each patch independently seeks its 1st Needle-NN). Deep $3 \times 3 \times 12$ Needles were used for correspondence estimation, since the goal is to find a *single best* accurate correspondence (not multiple NNs). Fig. 7a shows results on a real-video. *NeedleMatch* produces much better correspondences on the blurry fan. Incorporating smoothness into *NeedleMatch* is likely to improve its results.

Although using multiple scales, optical-flow methods do so *sequentially*. Their accuracy deteriorates at finer scales, as fine-scale patches are very blurry, misleading the final search. In contrast, *NeedleMatch* uses information from all scales simultaneously, not committing to any specific scale.

Note that matching patches at a fixed predetermined coarse scale will not achieve Needle performance. Different image regions benefit from different scales, depending on their *patchwise Signal-to-Degradation Ratio* (Patch-SDR). E.g., texture patches have higher Patch-SDR than uniform patches, thus benefit from finer scales in denoising and SR (Fig. 2). The degradation may also vary across the image: patches on the blurry fan (Fig. 7a) benefit from correspondences at coarse scales, whereas patches on the sharp static background benefit from finer scales. *NeedleMatch* inherently handles this ambiguity: coarse needle levels guarantee *high match reliability*, while fine needle levels guarantee *high localization accuracy*, when such details are available.

References

- [1] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 28(3), Aug. 2009. 4, 6
- [2] M. Black and P. Anandan. A framework for the robust estimation of optical flow. In *International Conference on Computer Vision (ICCV)*, pages 231–236, May 1993. 8
- [3] D. Bradley and T. Beeler. Local signal equalization for correspondence matching. *IEEE International Conference on Computer Vision (ICCV)*, 0:1881–1887, 2013. 1
- [4] A. Buadess, B. Coll, and J. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling and Simulation*, 4(2):490–530, 2005. 1, 6, 8
- [5] T. S. Cho, M. Butman, S. Avidan, and W. T. Freeman. The patch transform and its applications to image editing. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008. 1
- [6] A. Criminisi, P. Prez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13:1200–1212, 2004. 6
- [7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3d transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095, 2007. 1, 6, 7, 8
- [8] S. Divvala, A. Efros, M. Hebert, and S. Lazebnik. Unsupervised patch-based context from millions of images. In *CMU Tech Report*, 1010. 4
- [9] S. K. Divvala, D. Hoiem, J. Hays, A. A. Efros, and M. Hebert. An empirical study of context in object detection. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2009. 4
- [10] C. Doersch, A. Gupta, and A. A. Efros. Context as supervisory signal: Discovering objects with predictable context. In *European Conference on Computer Vision (ECCV)*, pages 362–377. Springer, 2014. 4
- [11] A. Efros and T. Leung. Texture synthesis by non-parametric sampling. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1033–1038, 1999. 1
- [12] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Comput. Graph. Appl.*, 22(2):56–65, March 2002. 1
- [13] J. H. Friedman, J. L. Bentley, and R. A. Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematics Software*, 3(3):209–226, September 1977. 6
- [14] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *IEEE International Conference on Computer Vision (ICCV)*, Tokyo, Japan, 2009. 1, 4, 6, 7
- [15] T. Hassner, V. Mayzels, and L. Zelnic-Manor. On sifts and their scales. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June. 2012. 1
- [16] K. He and J. Sun. Statistics of patch offsets for image completion. In *Proceedings of the 12th European Conference on Computer Vision (ECCV)*, pages 16–29, 2012. 1
- [17] M. Lebrun. An Analysis and Implementation of the BM3D Image Denoising Method. *Image Processing On Line*, 2:175–213, 2012. 8
- [18] A. Levin, B. Nadler, F. Durand, and W. T. Freeman. Patch complexity, finite pixel correlations and optimal denoising. In *Proceedings of the 12th European Conference on Computer Vision (ECCV)*, pages 73–86, 2012. 3
- [19] L. Li and H. Yan. Cost aggregation strategy with bilateral filter based on multi-scale nonlinear structure tensor. *Journal of Networks*, 6(7):958–965, 2011. 2
- [20] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'81*, pages 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc. 1, 6, 8
- [21] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *International Conference on Computer Vision (ICCV)*, volume 2, pages 416–423, July 2001. 6, 8
- [22] J. Marzat, Y. Dumortier, and A. Ducrot. Real-time dense and accurate parallel optical flow using cuda. In *Proceedings of the 17th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, Plzen, Czech Republic, february 2009. 8
- [23] T. Michaeli and M. Irani. Blind deblurring using internal patch recurrence. In *European Conference on Computer Vision (ECCV)*, pages 783–798. Springer, 2014. 3, 6
- [24] D. B. Min and K. Sohn. Cost aggregation and occlusion handling with WLS in stereo matching. *IEEE Transactions on Image Processing*, 17(8):1431–1442, 2008. 1
- [25] I. Mosseri, M. Zontak, and M. Irani. Combining the power of internal and external denoising. In *International Conference on Computational Photography (ICCP)*, pages 1–9, Harvard University, Cambridge, USA, April 2013. 1, 3, 7
- [26] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani. Summarizing visual data using bidirectional similarity. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2008. 1
- [27] A. Singh, F. Porikli, and N. Ahuja. Super-resolving noisy images. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, Columbus, Ohio, USA, June 2014. 7
- [28] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2432–2439, 2010. 3, 8
- [29] L. Sun, S. Cho, J. Wang, and J. Hays. Edge-based blur kernel estimation using patch priors. *International Conference on Computational Photography (ICCP)*, 0:1–8, 2013. 3, 6, 8
- [30] L. Sun and J. Hays. Super-resolution from internet-scale scene matching. In *International Conference on Computational Photography (ICCP)*, 2012. 4
- [31] A. Torralba. Contextual priming for object detection. *International Journal of Computer Vision (IJCV)*, 53(2):169–191, 2003. 4
- [32] K. Zhang, Y. Fang, D. Min, L. Sun, S. Yang, S. Yan, and Q. Tian. Cross-scale cost aggregation for stereo matching. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1590–1597, Columbus, OH, USA, 2014. 2

- [33] K. Zhang, J. Lu, and G. Lafruit. Cross-based local stereo matching using orthogonal integral images. *IEEE Trans. Circuits Syst. for Video Technol.*, 19(7):1073–1079, July 2009. [1](#)
- [34] M. Zontak and M. Irani. The internal statistics of a single natural image. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, June 2011. [1](#), [7](#), [8](#)
- [35] M. Zontak, I. Mosseri, and M. Irani. Separating signal from noise using patch recurrence across scales. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 0, pages 1195–1202, Los Alamitos, CA, USA, 2013. IEEE Computer Society. [1](#), [3](#)