

Con-Patch: When a Patch Meets its Context

Yaniv Romano, and Michael Elad, *Fellow, IEEE*

Abstract—Measuring the similarity between patches in images is a fundamental building block in various tasks. Naturally, the patch-size has a major impact on the matching quality, and on the consequent application performance. Under the assumption that our patch database is sufficiently sampled, using large patches (e.g. 21-by-21) should be preferred over small ones (e.g. 7-by-7). However, this “dense-sampling” assumption is rarely true; in most cases large patches cannot find relevant nearby examples. This phenomenon is a consequence of the curse of dimensionality, stating that the database-size should grow exponentially with the patch-size to ensure proper matches. This explains the favored choice of small patch-size in most applications.

Is there a way to keep the simplicity and work with small patches while getting some of the benefits that large patches provide? In this work we offer such an approach. We propose to concatenate the regular content of a conventional (small) patch with a compact representation of its (large) surroundings – its context. Therefore, with a minor increase of the dimensions (e.g. with additional 10 values to the patch representation), we implicitly/softly describe the information of a large patch. The additional descriptors are computed based on a self-similarity behavior of the patch surrounding.

We show that this approach achieves better matches, compared to the use of conventional-size patches, without the need to increase the database-size. Also, the effectiveness of the proposed method is tested on three distinct problems: (i) External natural image denoising, (ii) Depth image super-resolution, and (iii) Motion-compensated frame-rate up-conversion.

Index Terms—Similarity, context, patch, nearest neighbors search, image restoration, motion estimation.

I. INTRODUCTION

PATCH matching is extensively used in many image processing applications, including image restoration [1]–[10], video compression based on motion estimation [11]–[13], saliency detection [14], and more. Consider, for example, the problem of image denoising which will accompany our discussion as a case study; given a noise-corrupted version $\mathbf{Y} \in \mathbb{R}^N$ of a clean image $\mathbf{Z} \in \mathbb{R}^N$, the task is to approximate the unknown \mathbf{Z} . “External Denoising” [1]–[5] is a popular and effective approach for estimating the underlying clean image, where patch matching is the core engine in it. More specifically, external denoising methods (i) break the input image into (possibly overlapping) patches, (ii) clean each noisy patch by applying a weighted average over its k nearest neighbors (k -NN), chosen from a large database of clean image patches, and (iii) construct the final image by aggregating the cleaned patches. Intuitively, the more meaningful and accurate the neighbors that are found, the better the restoration performance. Note that external denoising can be considered

as an extension of the popular Non Local Means (NLM) algorithm [6], often called “Internal Denoising” [3] as it seeks for NN in the image itself rather than in an external database, leaning on the self-similarity property of images.

Given a patch database that is sufficiently sampled (e.g. composed of 10^{10} clean and small patches), Levin and Nadler [1], [2] show that external denoising of the form described above approximates very well the optimal Bayesian minimum mean squared error (MMSE) estimator of a noisy patch. They also study the effect of the patch-size on the denoising performance, showing how the restoration error improves with the growth of the patch dimension. Interestingly, even with the large database of patches used in this work (10^{10} examples), the authors encounter difficulties analyzing the performance of the denoising bound for patches that go beyond 7×7 pixels.

One can conclude from the above that working with large patches rapidly becomes impractical since the size of the database should grow exponentially with the patch-size in order to guarantee appropriate and meaningful nearby examples. This explains why relatively small patches are widely preferred in many applications that lean on patches and their matching.

Broadly speaking, no matter how good the local treatment is when operating on an image, it is limited in modeling its global nature [15], [16]. Therefore, in this context, using large patches is a step towards better image modeling [1], [2], [17], but may seem as a dead-end avenue, due to the difficulties mentioned above.

In this paper we suggest a novel way to keep the advantage of working with small patches, while gaining from the potential that large patches bring. Our main contribution is the idea of *context-patch* (short name – “*con-patch*”). We suggest concatenating every small patch (conventional raw data patch, e.g. of size 7×7) with a compact representation of its large surroundings, called *context-feature* (e.g. 10 values that describe a 21×21 patch). In other words, the con-patch has the explicit content of a small patch and the implicit encapsulated information of its large surrounding. The proposed approach bypasses the curse of dimensionality since the context-feature is very compact, leading to a minor increase of the dimensions of a small patch. Thus, it does not require to increase the size of the database we work with.

Inspired by the self-similarity descriptor [18], we propose a specific context-feature chosen for the augmentation of the patch, based on the self-similarity property between image patches. This feature-vector expresses how similar a small central patch is to its large surrounding window. As demonstrated in Fig. 1, we suggest measuring the similarity weights

$$w_{ij} = \exp \left\{ -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2} \right\}, \quad \forall j \in N(i) \quad (1)$$

Y. Romano is with the Department of Electrical Engineering, Technion – Israel Institute of Technology, Technion City, Haifa 32000, Israel. E-mail address: yromano@tx.technion.ac.il. M. Elad is with the Department of Computer Science, Technion – Israel Institute of Technology, Technion City, Haifa 32000, Israel. E-mail addresses: elad@cs.technion.ac.il.

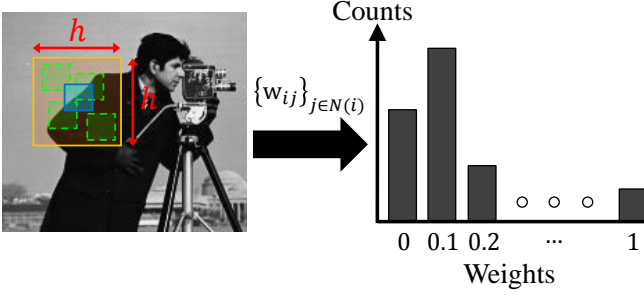


Fig. 1. The proposed context-feature. First, we measure the similarity weights w_{ij} between the central patch \mathbf{x}_i (blue solid line) and its surrounding patches \mathbf{x}_j (green dashed line) in a window of size $h \times h$ (orange solid line). Then, the weights are partitioned into a histogram of b bins, followed by a normalization step.

between a central patch \mathbf{x}_i and its surrounding patches \mathbf{x}_j in a neighborhood $N(i)$ of size $h \times h$ pixels (called “correlation surface” in [18]). In order to reduce computations, instead of having h^2 different weights, we evaluate w_{ij} in steps of m pixels in the horizontal and vertical directions, leading to $(\frac{h}{m})^2$ weights. Then, in order to reduce the dimensions, the weights are partitioned into a histogram, composed of b bins. Finally, the histogram is normalized to have a unit sum. Notice that the number of bins serves as the length of the context-feature, independent of the size of the surrounding window, describing some of the information that a large patch/window provide. In order to control the influence of the context-feature on the matching, the normalized histogram is multiplied by a scalar of choice.

It is important to emphasize that working with con-patches is not equivalent to the direct use of large patches. As such, what are the benefits of working with these extended vectors? What is the source of superiority of using the context over the conventional large or small patches? In this paper we show that the context helps by leading to better matches in the NN search; it filters irrelevant and misleading pseudo-matches that bias the search and reduce the application performance. On one hand, thanks to the compact representation of the context, we keep the advantage of working with small patches, for which there are sufficient meaningful nearby examples in the database. On the other hand, the context guides and “sharpens” the search mechanism to matches that are more appropriate by bringing a force that characterizes only large patches.

The concatenation of the context-features to the small patches enables us to easily plug the context in many algorithms, and detach it from the content whenever required. Furthermore, sticking to the original method’s distance measure, as we suggest, does not raise new limitations or additional algorithmic modifications. For instance, many methods use the l_2 -norm (Euclidean distance) as the similarity measure because of its simplicity [19]. In this case, plugging the context to the patch content is easily done.

In order to show the effectiveness of the proposed con-patch idea, we focus on several algorithms that rely on a patch matching step. We simply replace the conventional patches with our novel con-patches, and show the gained improvement in performance obtained due to the improved matching. We

demonstrate the power of con-patches on three very different applications that expose the generality of the proposed scheme: (i) External denoising, (ii) Single depth image super-resolution (SR), and (iii) Motion compensated frame rate up conversion (MC-FRUC).

Starting with external denoising [3], [4], as already mentioned, the noise in a given image is reduced by averaging clean matches that are found in the database by a NN search. The denoising performance is directly influenced by the quality of the neighbors found. With the use of the con-patch, better matches are obtained, due to the filtering effect of pseudo-matches that mislead the denoising process. Interestingly, Zontak et al. [3] demonstrate that the performance of internal denoising is better than an external one¹. This puts our work in its proper perspective, as we show that external denoising with con-patches stir things up and performs much better than the conventional internal or external denoising.

Next, we tackle the SR [8] problem, where the task is to increase the resolution of a given depth image. The authors of [8] suggest utilizing a database composed of pairs of low-resolution (LR) and high-resolution (HR) patches. They search the database for possible HR versions of the input LR patches, then construct the output image by merging the found matches. This problem enables us to demonstrate the effectiveness of the proposed context-feature for depth images. In addition, it indicates that the context is beneficial for the prediction of new spatial content.

The third problem we address with the con-patches is MC-FRUC [13], where the task is to increase the frame-rate of a given video. The temporal interpolation is achieved by estimating the local motion-trajectories, followed by a motion compensation step. Patch matching is an inherent part in these methods, where the context plays a key role. We show that the con-patches have also the ability to improve the motion estimation and thus the prediction of new temporal content.

This paper is organized as follows: In Section II we describe several related works in which the idea of context in different settings was proposed and demonstrated. In Section III we introduce our novel con-patches. Experiments are brought in Section IV, showing an improvement of the proposed idea for external denoising, SR and MC-FRUC. Conclusions are drawn in Section V.

II. RELATED WORK

There is a large ambiguity around the notion of *context* in the literature; it comes up in many applications and in different ways. For example, in object recognition there is a major importance to the contextual associations between objects [20], where the context is used in a high-level manner, e.g., by modeling the relations and interactions between objects. In contrast, in this paper we aim to benefit from the large patch statistics in a compact low-level fashion, putting emphasis on low-level tasks – image and video restoration.

In the fields of single image SR (i.e. upscaling) [21], [22] and inpainting (i.e. fill-in missing pixels/holes in an

¹In a follow-up paper [4], it has been shown that a combination of the internal and external approaches outperforms the individual performance of each of the methods.

image) [23], the context is utilized to constrain the search for candidate patches to regions of similar context. The core idea in these works is to apply a two-stage procedure: (i) seek for large regions with similar contextual behavior, and then (ii) construct the image by integrating patches from the found matched regions.

More specifically, Sun et al. [21] suggest dividing a database of images into pairs of LR and HR segments (relatively large regions with roughly uniform texture, obtained by applying a segmentation algorithm). Then, per each input pixel p , the database is searched to find several LR segments that are texturally similar to the input LR image segment (that the pixel p lies in). The similarity measure between segments is based on their responses to a derivative filter bank. Finally, a patch matching step is applied, where the potential nearby patches are constrained to the ones that belong to the matched segments.

In the case of image inpainting, Ruzic et al. [23] suggest dividing the image into variable size windows (segments) and use the responses to Gabor filter bank as a measure of contextual similarity. Then, a Markov random field model is utilized to fill-in the missing pixels by combining patches from the constrained matched windows rather than the whole image.

There are several major differences between [21]–[23] and our approach. We suggest concatenating the context to the content and apply a "one-shot" patch-matching, while the previous works choose a two-step strategy. In addition, the representation of the context is different – responses to filter bank applied on segments vs. internal self-similarities of patches in a large window. Furthermore, dividing the image into segments with similar textures may be irrelevant for some applications. For example, in MC-FRUC, when assuming relatively small movements, the motion is estimated by matching patches between two consecutive frames in a limited search window (e.g. 21×21). In this scenario, the NN search is already constrained (the search window can be considered as a matched segment), thus previous works reduce to a conventional patch-matching. In addition, finding segments with similar texture might mislead the matching step, especially when the boundaries of the segment coincide with image structures, or when the textures/edges are deformed due to local motions.

Another related work is the self-similarity descriptor [18], which is used in various high-level computer vision tasks, e.g., object recognition. Similarly to the proposed context-feature, the self-similarity descriptor computes the correlation surface, i.e., the similarity weights between a central patch to its neighbors, as defined in Eq. (1). Yet, differently from our suggestion, a log-polar mapping is applied on the correlation surface. Finally, a quantization step is applied, leading to a normalized histogram representation of the descriptor. Our added feature bares some resemblance to the self-similarity descriptor and can be considered as a simpler version of it. However, different from earlier work, we suggest using the self-similarity descriptor as a way to represent the context of the patch for improving the NN search for low-level image processing applications and it is added to the regular small patch.

We should note that, following our definition of context-feature, any compact representation of the large patch can be considered; the self-similarity feature is one out of many possibilities.

In our experiments we tested other ways to incorporate the context of the small patch, e.g., by representing the large patch using HoG descriptor [24], or by using a downscaled version of the surroundings (which closely related to the patch foveation operator [25]). Our tests show that working with the self-similarity feature is superior to these options.

III. THE PROPOSED CON-PATCHES

In this section we provide a detailed description of the proposed context-feature. Then, we show how to construct the con-patches.

A. Self-Similarity Feature

The core idea behind the context-feature is to represent the large surroundings of a patch in a compact fashion. Once we have a low dimensional feature that describes the large patch, we are able to keep the simplicity and work with small patches while having some of the flavor of working with large ones. Naturally, there are many ways to encapsulate the large patch information. In this paper we rely on the self-similarity property of images that has drawn much attention in recent years.

Fig. 1 illustrates the procedure of computing the context-feature $\mathcal{H}(\mathbf{x}_i)$, describing some of the information of the large surrounding of the patch \mathbf{x}_i . Following the self-similarity descriptor [18], we compute the correlation surface – set of similarity weights between a central patch \mathbf{x}_i and its surrounding patches $\mathbf{x}_j \in N(i)$ in a neighborhood (window) of size $h \times h$, expressed by

$$w_{ij} = \exp \left\{ -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2} \right\}, \quad \forall j \in N(i).$$

The parameter σ controls the acceptable change in illumination due to noise, compression artifacts, etc. Notice that $w_{ij} \rightarrow 1$ indicates that the patch \mathbf{x}_i and \mathbf{x}_j are highly similar, while $w_{ij} \rightarrow 0$ indicates that the two are substantially different. The size of the surrounding window can be the whole image, or a smaller portion of it. Intuitively, when the majority of w_{ij} are high, we can conclude that the small central patch \mathbf{x}_i has many co-occurrences in its surroundings, e.g., it might originate from a large flat area. Similarly, if most of the weights are small, the patch \mathbf{x}_i is unique, e.g., it might originate from a highly textured and non-repetitive area. As such, these weights provide information about the geometric structure of a large patch w.r.t. its small central patch.

Next, in order to have a compact representation, we suggest a rearrangement of the correlation surface $\{w_{ij}\}_{j \in N(i)}$ of the patch \mathbf{x}_i into a histogram of b -bins (e.g. $b = 10$). Finally, histogram normalization step is applied, leading to the context-feature $\mathcal{H}(\mathbf{x}_i)$, which is an empirical distribution of the co-occurrences of the central patch in its large surroundings. Put differently, $\mathcal{H}(\mathbf{x}_i)$ measures the correlations between the central patch to its surroundings: When the correlation is

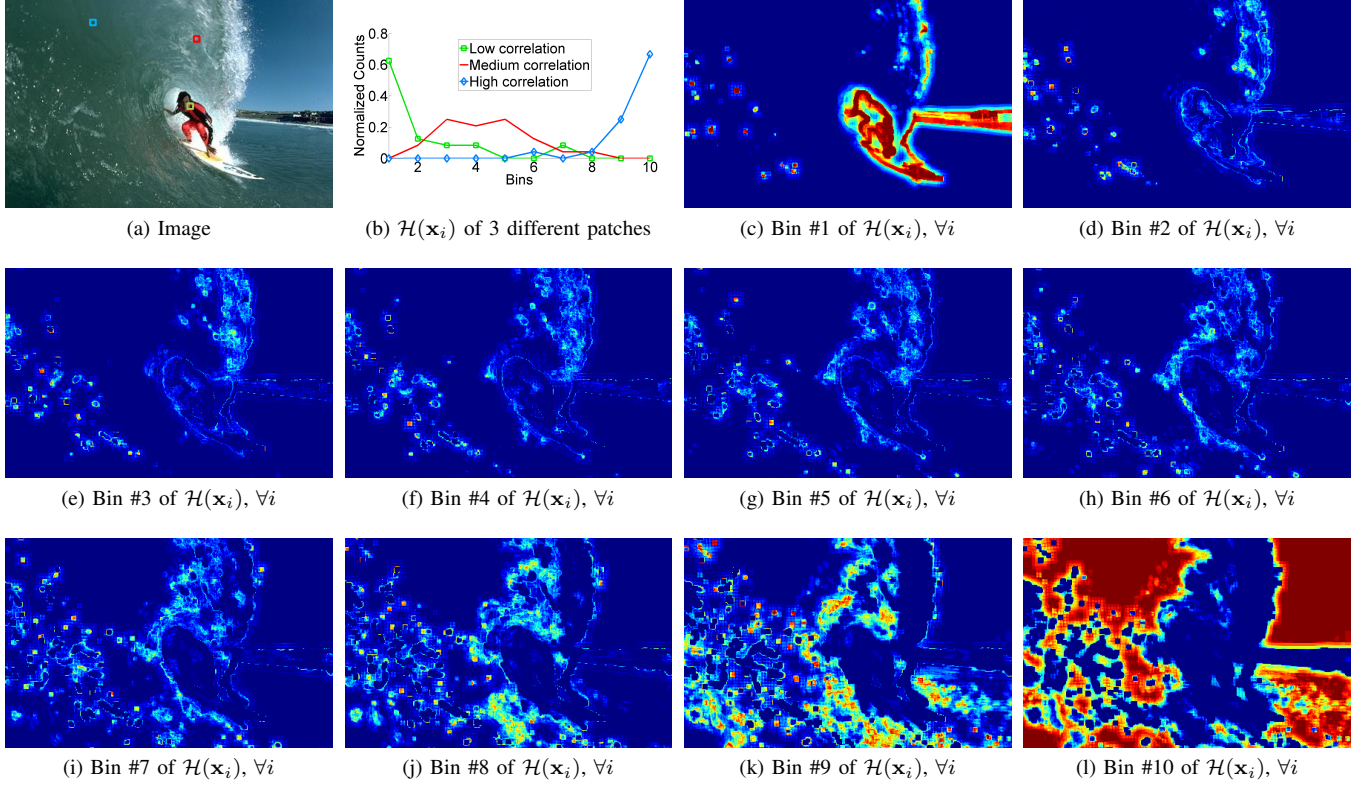


Fig. 2. Visualization of the context-feature, computed on *Surfer* image. (a) The input image, (b) The histogram $\mathcal{H}(\mathbf{x}_i)$ of the green, red and blue patches taken from *Surfer* image, (c)-(l) The values of the normalized histogram $\mathcal{H}(\mathbf{x}_i)$ per each patch and bin, warm and cold colors indicate low and high values, respectively.

relatively high (many co-occurrences) the histogram is biased towards $\delta(b_k - 10)$, and when the correlation is low the histogram concentrates around $\delta(b_k - 1)$, where b_k is the k -th bin of the histogram.

A visualization of the context-feature is given in Fig. 2. We compute the proposed feature $\mathcal{H}(\mathbf{x}_i)$ for each patch \mathbf{x}_i of *Surfer* image. The size of the small patch and its surrounding window are 7×7 and 21×21 , respectively. The similarity weights $\{w_{ij}\}_{j \in N(i)}$ are partitioned into histogram of 10 bins, where the different weights are evenly spaced (the space is 0.1)². An illustration of the normalized histogram of 3 different patches, taken from *Surfer* image, is given in Fig. 2b. The green patch in Fig. 2a has a small amount of co-occurrences in its large surroundings, thus most of its similarity weights are small, leading to the green curve (with square marker) that concentrates around the first bin (weights between 0 to 0.1). On the other hand, the blue patch in Fig. 2a is highly correlated to its surrounding patches, leading to the blue curve (with diamond marker), which concentrates around the last two bins (weights between 0.8 to 1). The red patch has “medium” amount of co-occurrences, thus the red curve in Fig. 2b is spread over wide range of bins.

The 10 heat-maps in Fig. 2c – Fig. 2l plot the values of the features, computed per each patch, where warmer color corresponds to larger value. Notice that the i -th pixel in

each heat-map corresponds to a specific bin of the histogram, representing one element of $\mathcal{H}(\mathbf{x}_i)$. The i -th pixel along these 10 heat-maps constructs the whole feature $\mathcal{H}(\mathbf{x}_i)$, i.e., the 1D normalized histogram. Fig. 2c plots the values of the first bin in $\mathcal{H}(\mathbf{x}_i)$ for all the patches, while Fig. 2l plots the values of the last bin. As can be seen, warm pixels in Fig. 2c correspond to patches that are different than their surroundings (e.g. the foam of the wave and the surfer’s face), while warm pixels in the latter heat-map correspond to patches that are similar to their neighbors (e.g. the sky which is a flat area). In some sense, this feature is a measure of the complexity of the patch, “simple” patches tend to be correlated with their neighbors, while complex patches are unique and has a few (if any) nearest neighbors.

In the case of external image denoising, Fig. 3 demonstrates the effect of the length of the histogram on the denoising performance for various noise levels. The restoration is evaluated by averaging the Peak Signal to Noise Ratio (PSNR) over 10 images, randomly chosen from BSDS300 dataset [26], where the higher the PSNR the better the restoration. The similarity weights w_{ij} are partitioned into histogram of b equally-spaced bins ($b \in \{0, 2, 4, 8, 16\}$), followed by a normalization step. The importance of the context-feature (a scalar that multiplies the histogram, see more details in the next subsection) is optimally tuned per each histogram-length and noise-level. As can be seen, increasing the number of bins from $b = 0$ (i.e., use the conventional patch) to $b = 8$ consistently improves the restoration, demonstrating the

²A detailed discussion about the effect of the length of the histogram is given hereafter.

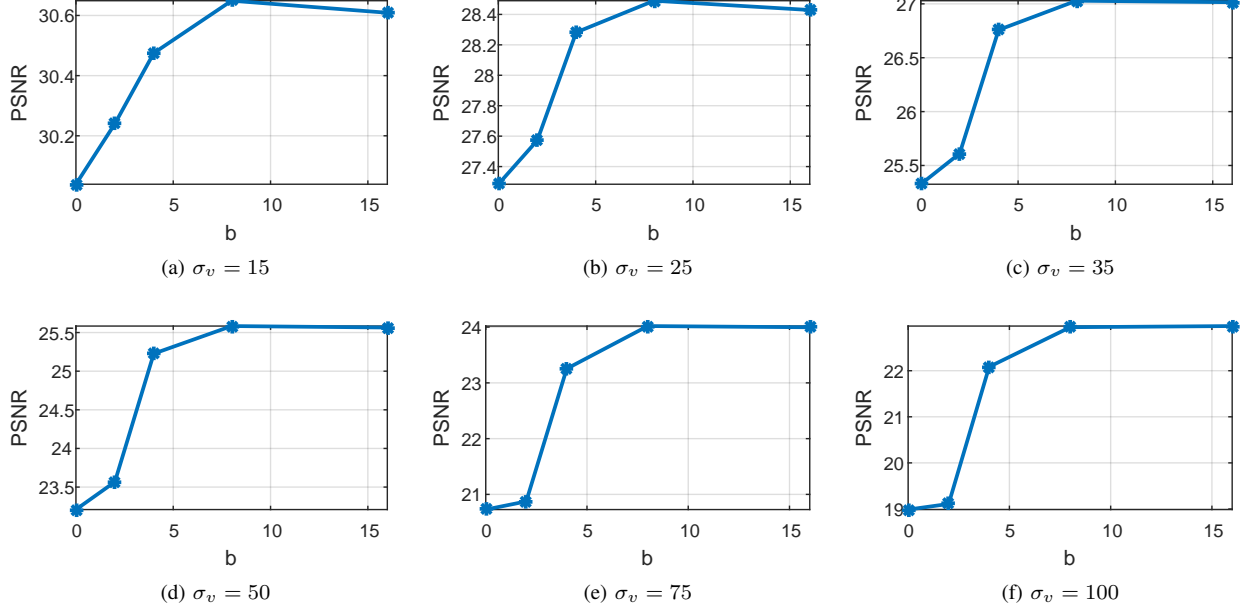


Fig. 3. The effect of the length of the context-feature \mathcal{H} on external denoising performance for different noise-levels, varying from (a) $\sigma_v = 15$ up-to (f) $\sigma_v = 100$. The results obtained by averaging the PSNR over 10 randomly images, taken from the test-set of BSDS300 database. Per each noise level, the similarity weights are partitioned into b equally-spaced bins ($b \in \{0, 2, 4, 8, 16\}$), leading to context-feature of length b . Note that conventional external filtering is obtained when $b = 0$, serving as a baseline for comparison.

effectiveness of the context. Using a finer grid than 8 bins does not gain additional improvement, indicating the compactness of the proposed feature.

B. Constructing the Con-Patches

In the following discussion we assume that the distance between two patches is defined as

$$d^2(\mathbf{q}, \mathbf{x}_j) = \|\mathbf{q} - \mathbf{x}_j\|_2^2, \quad (2)$$

where $\mathbf{q} \in \mathbf{R}^{h \times h}$ is an input query patch and $\mathbf{x}_j \in \mathbf{R}^{h \times h}$ is a sample from the database. Both \mathbf{q} and \mathbf{x}_j are large patches of size $h \times h$, held as column vectors after lexicographic ordering. The rational behind the con-patches is to leverage the additional force that large patches bring but still work with small patches. As such, we aim to replace a portion from the large patch with its compact representation (context-feature). More specifically, using norm properties we can rewrite Eq. (2) as

$$d^2(\mathbf{q}, \mathbf{x}_j) = \|\mathbf{P}_c \mathbf{q} - \mathbf{P}_c \mathbf{x}_j\|_2^2 + \|\mathbf{P}_s \mathbf{q} - \mathbf{P}_s \mathbf{x}_j\|_2^2, \quad (3)$$

where $\mathbf{P}_c \in \mathbf{R}^{c^2 \times h^2}$ and $\mathbf{P}_s \in \mathbf{R}^{s^2 \times h^2}$ are operators that extract the central and surrounding parts of a patch, respectively. The central part is the $c \times c$ small patch, while the surrounding is of size $s \times s$ pixels ($c \ll s$). Notice that in our terminology, $\mathbf{P}_c \mathbf{q}$ is the regular patch, while $\mathbf{P}_s \mathbf{q}$ is its context, *but in high dimensions*.

We can dramatically reduce the size of the large patch by replacing the surrounding of the central patch with its context-

feature. As such, the modified distance between two patches is expressed by

$$\begin{aligned} d_{\mathcal{H}}^2(\mathbf{q}, \mathbf{x}_j) &\approx \|\mathbf{P}_c \mathbf{q} - \mathbf{P}_c \mathbf{x}_j\|_2^2 + \alpha \|\mathcal{H}(\mathbf{q}) - \mathcal{H}(\mathbf{x}_j)\|_2^2 \quad (4) \\ &= \left\| \begin{bmatrix} \mathbf{P}_c \mathbf{q} \\ \sqrt{\alpha} \mathcal{H}(\mathbf{q}) \end{bmatrix} - \begin{bmatrix} \mathbf{P}_c \mathbf{x}_j \\ \sqrt{\alpha} \mathcal{H}(\mathbf{x}_j) \end{bmatrix} \right\|_2^2, \end{aligned}$$

where α is a scalar (gain) that controls the importance/weight of the context. Notice that the con-patches are obtained by concatenating the context-feature to the regular patch. In addition, the size of the con-patch is much smaller than the large patch, while being slightly bigger than the central one.

C. Testing the Power of Con-Patches

The above description may suggest that the true distance between the large patches and the approximated one obtained via the con-patches tend to be similar. In this section we address this interpretation and show that while this is not exactly the case, a related and highly desired property does hold true.

The following experiment illustrates the superiority of applying k -NN search using the proposed context-based distance $d_{\mathcal{H}}^2(\mathbf{q}_i, \mathbf{x}_j)$ over both the conventional distance between large $d^2(\mathbf{q}_i, \mathbf{x}_j)$ or small $d^2(\mathbf{P}_c \mathbf{q}_i, \mathbf{P}_c \mathbf{x}_j)$ patches, all in the context of noisy data. In what follows we propose an experiment in which we show that the con-patches lead to more appropriate matches, when compared to the above two alternatives.

The size of a small patch is 7×7 , which is the center part of a large patch of size 17×17 . The con-patch is composed of the small central patch (of size 7×7), concatenated to an 8 bin context-feature that represents its 17×17 surroundings.

We gather two sets of clean large image patches of size 17×17 pixels: query-set $\{\mathbf{q}_i\}_{i=1}^{N_T}$ and example-set $\{\mathbf{x}_j\}_{j=1}^{N_D}$,

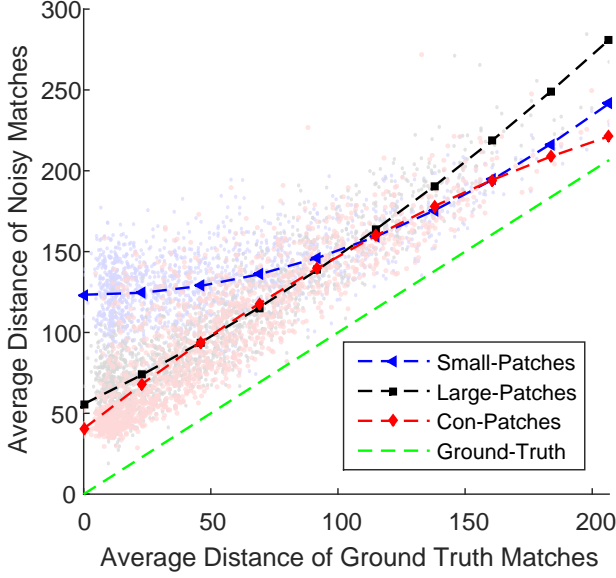


Fig. 4. Comparison between applying k -NN using small (blue triangles), large (black squares) and con-patches (red diamonds). Refer to Section III-C for a detailed explanation.

where $N_T = 8 \times 10^5$ and $N_D = 35 \times 10^5$. In addition, a noisy version of the query-set, denoted by $\{y_i\}_{i=1}^{N_T}$, is created by contaminating the patches with zero-mean Gaussian noise with $\sigma_v = 35$. All these patches are randomly chosen from BSDS300 database [26], where we ensure that the noisy query-set patches $\{y_i\}_{i=1}^{N_T}$ have at least $k = 20$ NNs in the example-set $\{x_j\}_{j=1}^{N_D}$ (in the sense of having an Euclidean distance that falls below the noise level).

A set of ground-truth matches per each query-patch is created by applying k -NN search in the example-set based on the *small central part* (of size 7×7) of the *clean query-patch* $P_c q_i$. Then, we evaluate the average l_2 distance between $P_c q_i$ and its $k = 20$ ground-truth matches, denoting this distance by $E_{GT}(i)$. Since this computation is done using the clean patches, these stand as the ultimate distances between each small query patch and the candidate neighbors from the example-set.

Next, k -NN search is applied for each *noisy query-patch*, using the 3 variants of the distances:

- 1) Using small patches of size 7×7 , $d^2(P_c y_i, P_c x_j)$,
- 2) Using large patches of size 17×17 , $d^2(y_i, x_j)$, and
- 3) Using the con-patches, with the whole center part of size 7×7 and an augmented feature of length 8 to represent the surrounding of size 17×17 , $d_{\mathcal{H}}^2(y_i, x_j)$,

leading to $E_{small}(i)$, $E_{large}(i)$ and $E_{con}(i)$, respectively. Notice that the matching process is done based on the *noisy* patches, but the average error/distance E is always computed based on the small central part of the clean patches. As such, the closer the average distance to the average ground-truth distance $E_{GT}(i)$, the better the matching process.

Fig. 4 illustrates these average distances between 1500 randomly chosen query-patches and their matches. The horizontal coordinate of each point i in the scatter-plot represents the value of $E_{GT}(i)$, while the vertical coordinates represent

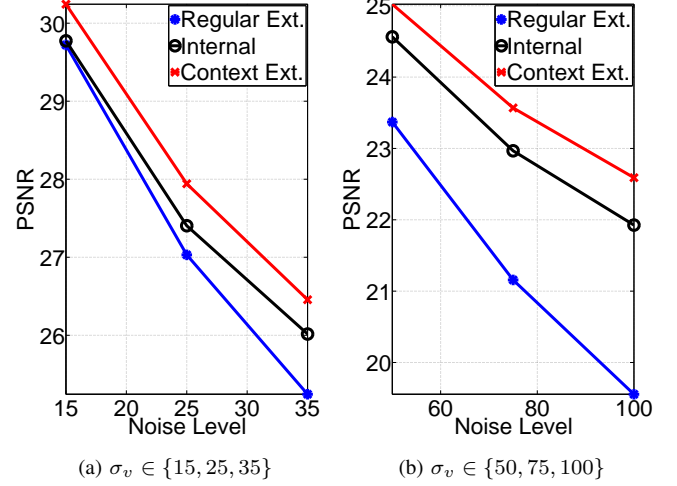


Fig. 5. Denoising results (in terms of PSNR) of conventional external denoising, internal NLM, and the proposed context-based external denoising, averaged over 100 test-set images of BSDS300 database. (a) Comparison for noise levels 15, 25, 35, and (b) Comparison for noise levels 50, 75, 100.

the values of $E_{small}(i)$, $E_{large}(i)$, and $E_{con}(i)$, respectively. In addition, the dashed blue-triangle, black-square, and red-diamond curves correspond to polynomial fits (of degree 2) of E_{small} , E_{large} , and E_{con} , respectively, where the fits are computed based on the whole query-set (8×10^5 patches). The dashed green line represents the case in which the horizontal and vertical coordinates are equal, thereby fulfilling it implies that all the right neighbors are found.

Following Fig. 4, when the test-set patches have many meaningful nearby samples in the example-set (in the region that $E_{GT} < 100$), the curves of E_{large} and E_{con} are comparable and much better than E_{small} . In this region the “dense-sampling” assumption holds, and applying k -NN search using large or con-patches results in equally good set of neighbors, while working with small patches tends to err. In this regime the proposed con-patches successfully seize the important information that large patches provide, in a very compact way while being robust to the noise.

On the other hand, for relatively large ground-truth distances ($E_{GT} > 100$), the E_{con} curve is found to be the best. In this regime there is a limited amount of nearby samples, and thus working with large patches is inferior to both small or con-patches.

This experiment shows that the context proposed is beneficial across the board – it leads to better matches than the ones obtained by small or large patches. This experiment also demonstrates that our context-feature exploits the unique force of large patches, along with the attractive “dense-sampling” property of small ones.

To conclude, the proposed con-patch benefits from the simplicity and convenience of working with small patches, as it adds a short feature vector to each patch. On the other hand, con-patches bring an additional force that relates to large patches, and does so without the need to have a denser sampling of the patch manifold. We have seen that beyond these properties, con-patches introduce a robustness to noise

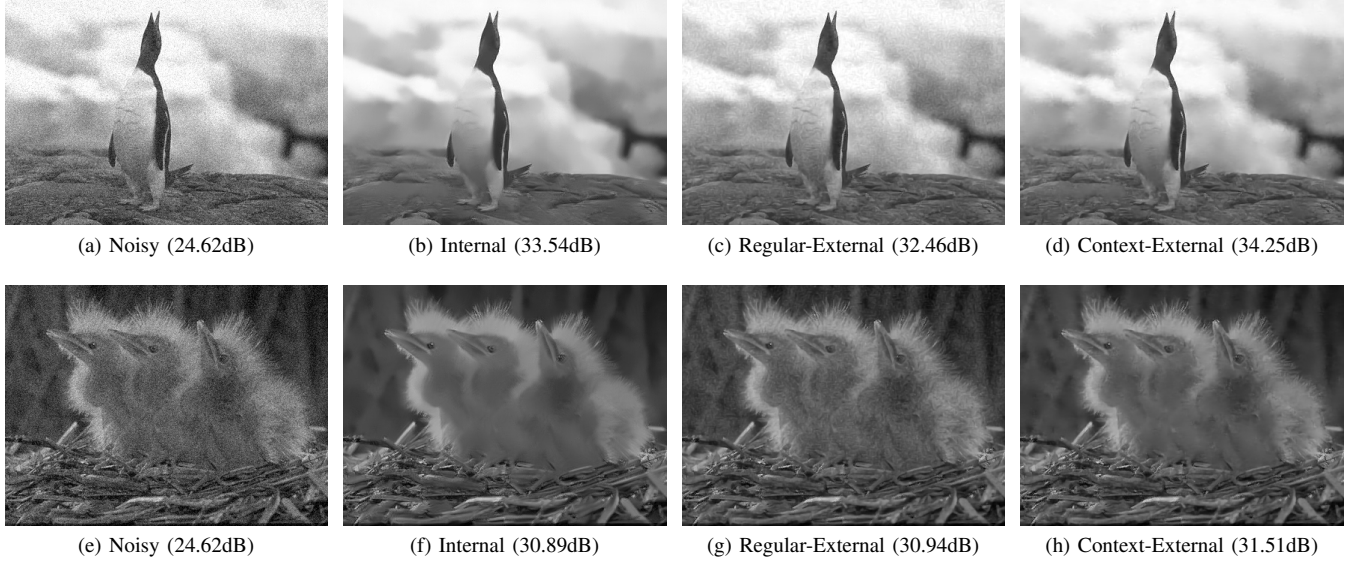


Fig. 6. Visual comparison for noisy ($\sigma = 15$) Penguin and Baby-Birds images: (a,e) Noisy images, (b,f) Internal NLM, (c,g) External denoising using conventional patches, and (d,h) External denoising using con-patches.

which enables them to perform even better than large patches. In the following section we demonstrate the effectiveness of the con-patches on various applications.

IV. EXPERIMENTAL RESULTS

In this section we test the proposed idea on three problems: (i) External image denoising, (ii) Depth single image SR, and (iii) MC-FRUC. The core restoration mechanism in these problems is a NN search, while each problem has different degradation model: additive noise, missing of spatial or temporal content, respectively. We show that a significant boost in performance is achieved by simply plugging the context-feature to the conventional patches.

A. External Denoising

The image denoising problem [1]–[6], [15], [27]–[31] assumes that a noisy measurement $\mathbf{Y} \in \mathbf{R}^N$ of a clean image $\mathbf{Z} \in \mathbf{R}^N$ is obtained by a degradation of the form

$$\mathbf{Y} = \mathbf{Z} + \mathbf{V}, \quad (5)$$

where $\mathbf{V} \in \mathbf{R}^N$ is an i.i.d. zero-mean Gaussian noise with standard-deviation σ_v . Our goal is to restore the unknown underlying image \mathbf{Z} from \mathbf{Y} using a large database of clean images. More formally, external denoising methods break the noisy image \mathbf{Y} into overlapping patches \mathbf{y}_i and denoise each patch by applying a weighted average over the k -NN of the patch \mathbf{y}_i , expressed by

$$\hat{\mathbf{x}}_i = \frac{1}{\sum_{j=1}^k \exp \left\{ -\frac{\|\mathbf{y}_i - \mathbf{x}_i^j\|_2^2}{2\sigma_v^2} \right\}} \sum_{j=1}^k \exp \left\{ -\frac{\|\mathbf{y}_i - \mathbf{x}_i^j\|_2^2}{2\sigma_v^2} \right\} \mathbf{x}_i^j, \quad (6)$$

where $\hat{\mathbf{x}}_i$ is the denoised version of the noisy patch \mathbf{y}_i , and $\{\mathbf{x}_i^j\}_{j=1}^k$ are the nearby patches that are found in the database. Then, the denoised image $\hat{\mathbf{x}}$ is constructed from the locally

cleaned patches by returning each patch $\hat{\mathbf{x}}_i$ to its i -th location in the global image, followed by averaging over the overlaps.

Generally speaking, given the noisy patches \mathbf{y}_i , the challenge is to find several clean database samples $\{\mathbf{x}_i^j\}_{j=1}^k$ that are close to the unknown patch \mathbf{z}_i . The closer the matched samples to \mathbf{z}_i , the better the restoration. In this experiment we show that improved restorations are achieved when using con-patches in the NN search rather than the conventional patches.

In order to plug the context we should compute the context-feature, both on the clean database images and the input noisy ones. The parameter σ in Eq. (1), which controls the acceptable change in illumination, is set to be a fixed value ($\sigma = 5$) for the clean database images, and equal to the noise-level σ_v for the noisy images. Per each noise level σ_v , we tuned the parameter α , the gain which controls the importance of the context feature, resulting in

$$\sqrt{\alpha} = \begin{cases} 0.9 & \sigma_v < 50 \\ 1.1 & 50 \leq \sigma_v < 100 \\ 1.3 & \sigma_v \geq 100 \end{cases}$$

The rest of the parameters are fixed for all noise-levels; the small patch is of size $c \times c = 7 \times 7$, the context-feature is of length $b = 8$, computed on a window of size $h \times h = 21 \times 21$ with step-size of $m = 4$ pixels. The denoising is done by searching for the $k = 500$ NN in an external database of 4.5×10^6 randomly sampled patches, taken from the training-set images of BSDS300 dataset [26]. Notice that the results of external denoising using conventional patches are obtained by setting $\alpha = 0$ in the NN search.

The proposed algorithm is evaluated based on the 100 test-set images of BSDS300 database. Fig. 5 demonstrates the denoising results for various noise levels in terms of PSNR, where we compare the context-based external denoiser to its conventional version [1]–[4], and also to internal NLM [32]. As demonstrated in Fig. 5, external denoising using con-

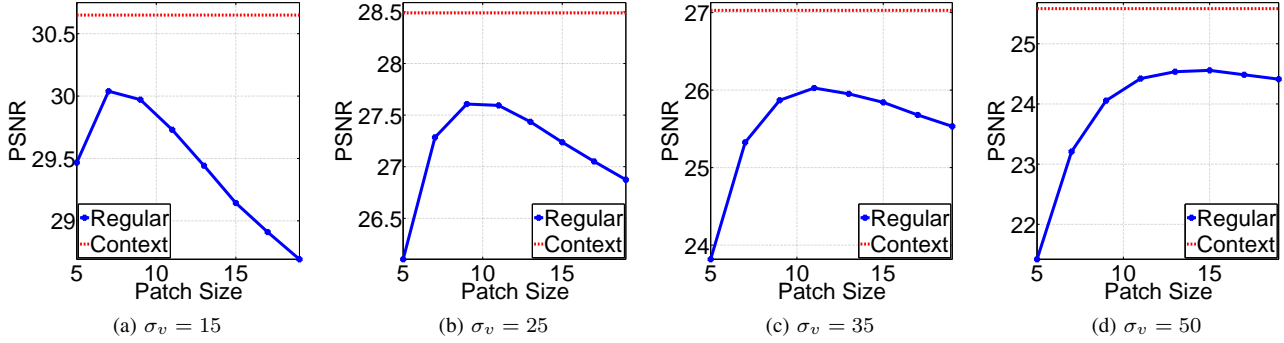


Fig. 7. The effect of the patch-size on conventional external denoising performance for different noise-levels (blue solid curves), along with a comparison to the proposed context-based approach (red dashed line). The size of the con-patch is fixed to 7×7 patch, concatenated to a feature of length $b = 8$. We measure the average PSNR over 10 randomly chosen images from the test-set of BSDS300 database, obtained by applying conventional external denoising with patches of varying size, starting from 5×5 up to 19×19 . The effect of the patch-size is tested for the following noise levels: (a) $\sigma_v = 15$, (b) $\sigma_v = 25$, (c) $\sigma_v = 35$, and (d) $\sigma_v = 50$.

patches is significantly better than the conventional patches, indicating that the context has a major impact on the patch matching mechanism. Quantitatively, on average, for noise levels of 15, 25, 35, 50, 75, and 100 plugging the context leads to improvement of 0.52dB, 0.91dB, 1.21dB, 1.65dB, 2.41dB, and 3.03dB, respectively.

A visual comparison between external denoising with and without integrating the context is given in Fig. 6. As can be seen, con-patches lead to a pleasant restoration of the underlying image, both in flat (e.g. the skies) and structured areas (e.g. the rocks or twigs). On the other hand, conventional patches result in an effective restoration of structured/active areas, but fail to recover the flat ones. This observation stands in agreement with [3], [4], which demonstrate that conventional external denoising tends to over-fit the noise details. Put differently, it has been shown [4] that random-noise patches have high correlation with natural image patches, explaining why the conventional external denoiser introduces undesired structures, especially in flat areas.

Turning to internal NLM, the results in Fig. 5 are obtained by fine-tuning the NLM [32] parameters to produce the best PSNR for patches of size 7×7 and search window of size 21×21 . Note that we do not test the proposed con-patches on internal NLM since it already seeks for NNs in a limited window, which can be considered as the internal context of the patch. As demonstrated, internal NLM outperforms its conventional external variant, complying with the conclusions of [3]. However, this observation is no longer valid when integrating the context of the patch. As can be seen, external denoising with con-patches outperforms the internal version with average gaps of 0.47dB, 0.54dB, 0.44dB, 0.46dB, 0.60dB, and 0.66dB, for noise-levels 15, 25, 35, 50, 75, and 100, respectively. The achieved gain in performance is also reflected visually, as demonstrated in Fig. 6 (notice the improved restoration in the rocks area of Penguin image and the twigs area of Baby-Birds image).

Fig. 7 demonstrates that the proposed con-patch approach benefits from the advantages of working with both large and small patches. It plots the restoration performance of external denoising using the proposed con-patches (flat red dashed-

line), compared to conventional patches of varying size (blue curves) – starting from 5×5 ($c = 5$) up to 19×19 ($c = 19$). In this experiment the *size of the con-patch is fixed*, obtained by concatenating 7×7 patch to context-feature of length 8. Broadly speaking, as seen in Fig. 7, the PSNR increases with the conventional patch-size due to improved matching, up to a point where the PSNR starts decreasing due to the lack in valid NN. Note that this observation fits the theoretical conclusions of [1], [2]. However, the context-based approach leads to restoration that outperforms the conventional approach (for all patch-size), demonstrating the power of the context, which benefits from improved patch matching while keeping the “dense-sampling” property.

To summarize, we illustrated that a significant improvement in denoising performance is achieved simply by plugging the context-feature to the conventional patch in the NN search. We also show how context-based external denoising outperforms the conventional internal version. Finally, by comparing the performance of the context-based denoising to the one obtained by conventional large patches, we show that working with con-patches is not equivalent and even might be much better than the direct use of large patches.

B. Depth Image Super Resolution

Single depth image SR is the process of increasing the resolution of a LR depth image. In this experiment, we build upon the external SR algorithm of Aodha et al. [8], and demonstrate the effectiveness of the proposed con-patches.

Given a generic database of LR-HR pairs of depth patches, the authors of [8] suggest searching for the k -NN of each of the input LR image patches, leading to k HR candidates per patch. Then, using an MRF model, these candidates are combined together and form the desired HR image. Naturally, the quality of the matches affect the restoration performance. Therefore, we find it interesting to integrate the context-feature to the NN search and compare the results to the conventional approach. In order to have a fair comparison, we follow the experiments of the external SR work [8] and test the two variants of the algorithm on Middlebury stereo dataset [33]. The LR images

TABLE I
COMPARISON BETWEEN THE RESULTS [PSNR] OF THE CONVENTIONAL/REGULAR EXTERNAL SR [8] AND THE PROPOSED CONTEXT-BASED VARIANT.
THE BEST RESULTS PER EACH IMAGE AND SCALING FACTOR ARE HIGHLIGHTED.

Method (Scaling)	Cones	Teddy	Tsukuba	Venus	Scan 21	Scan 30	Scan 42	Average
Regular (x2)	32.51	30.64	20.62	33.78	41.80	42.59	37.97	34.27
Context (x2)	33.11	31.57	21.55	34.50	42.05	42.88	38.49	34.88
Regular (x4)	29.15	28.56	19.16	30.22	39.62	39.74	32.55	31.29
Context (x4)	29.98	29.30	19.33	32.68	40.18	40.85	34.33	32.38

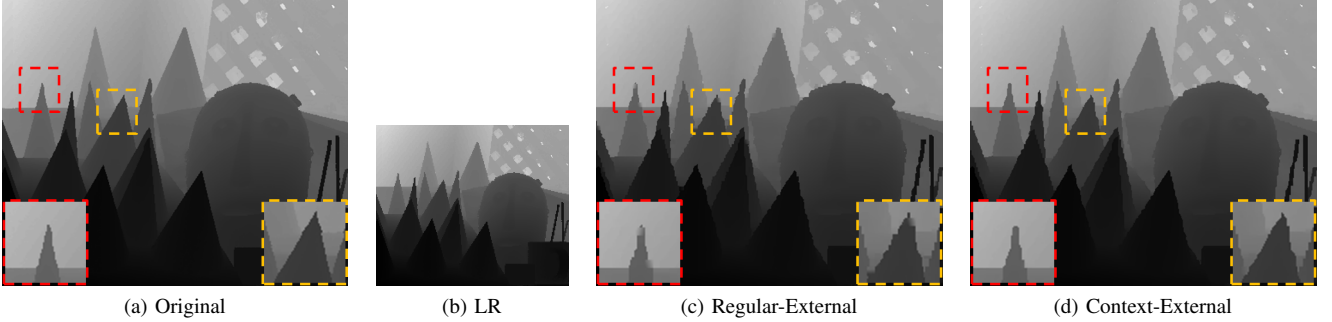


Fig. 8. Visual comparison for upscaling *Cones* depth image by a factor of 2. (a) Original image, (b) LR input image, (c) External SR using conventional patches [8] – 32.51dB, and (d) External SR using con-patches – 33.11dB.

are generated by decimating the dataset images by factors of 2 and 4, both in the horizontal and vertical dimensions. The reconstructed images are obtained by applying the original software of [8] on the generated LR images.

The context-feature is computed by measuring the similarity of 11×11 patch to its 21×21 neighborhood, where $\sigma = 5$ and $m = 4$ (both for the dataset and input images). Similarly to image denoising, this feature is plugged *only* in the NN search (with $\sqrt{\alpha} = 0.4$) and detached from the patch for the rest of the algorithm. Both in the conventional and its context-based variant, we use the very same parameters that were chosen by the authors.

In Table I we provide a comparison (in terms of PSNR) between the conventional patch-based external SR [8] and its context-based variant. As can be seen, the con-patch approach results in a superior restoration performance when compared with the conventional one. Quantitatively, we gain an average improvement of 0.61dB and 1.09dB for upscaling by factors of 2 and 4, respectively. The gain in PSNR is also reflected visually, as demonstrated in Fig. 8 for *Cones* image, upscaled by a factor of 2. As illustrated, the proposed context-based restoration results in continuous edges, having less artifacts than the conventional version (notice the 2 zoomed-in regions).

To conclude, we test the proposed concept on single depth image SR, showing that the context-feature fits to depth images. This experiment demonstrates how con-patches can effectively boost the restoration of missing spatial content.

C. Motion-Compensated Frame-Rate Up Conversion

Frame Rate Up Conversion (FRUC) is the process of increasing the temporal resolution (frame-rate) of a given video, used for format conversion, improvement of low bit-rate video coding, and more [13], [34]–[40]. Frame averaging or repetition are the simplest FRUC techniques, ignoring the

motion between frames, and thereby result in motion-jerkiness or ghost artifacts. As such, interpolators that consider the motion between frames are preferred, often times leading to pleasant conversion. The widely used block/patch-based methods consist of block-matching step that seek for similar blocks in a limited search window within the consecutive (and known) frames, leading to an estimation of the local motion trajectory. Then, new interpolated frames are obtained by compensating the motions of the matched blocks.

Similarly to previous applications, the effectiveness of the context is tested by integrating the context-feature only to the block-matching step. Therefore, a gain in performance is achieved due to improved matching, leading to better local motion estimation than the conventional block-based approach. In this experiment, we build upon the popular bi-directional technique [34], [36], [37], which estimates the motion of 16×16 block in half-pixel accuracy, within a search radius of 10 pixels. The context-feature is computed by measuring the similarity between 7×7 patch to its 21×21 neighborhood ($\sigma = 10$, $m = 4$), concatenated to the conventional 16×16 block, where $\sqrt{\alpha} = 1.3$.

Table II demonstrates the effect of the context on the bi-directional MC-FRUC for various video sequences, which are commonly used in related publications. Per each sequence, we measure the average PSNR over 80 interpolated frames, temporally up-sampled by a factor of 2. Following the results in Table II, the average improvement (over all the sequences) that achieved by plugging the context-feature to the conventional-blocks is 0.11dB, expressing the ability of the context to improve the block-matching step.

A visual illustration can be found in Fig. 9, presenting the estimation of a frame from *Foreman* sequence. More specifically, given the existing frames at times $t - 1$ (Fig. 9a) and $t + 1$ (Fig. 9c), our goal is to estimate the unknown frame at

TABLE II
COMPARISON BETWEEN THE RESULTS [PSNR] OF THE CONVENTIONAL BLOCK-BASED MC-FRUC [8] AND ITS CONTEXT-BASED VARIANT, FOR TEMPORAL UP-SAMPLING BY A FACTOR OF 2. THE BEST RESULTS PER EACH VIDEO SEQUENCE ARE HIGHLIGHTED.

Method	Foreman	Flower	Tennis	Harbour	Park-Run	Ice	Average
Regular	33.34	29.82	28.16	30.40	28.07	33.70	30.58
Context	33.47	29.95	28.31	30.46	28.13	33.82	30.69

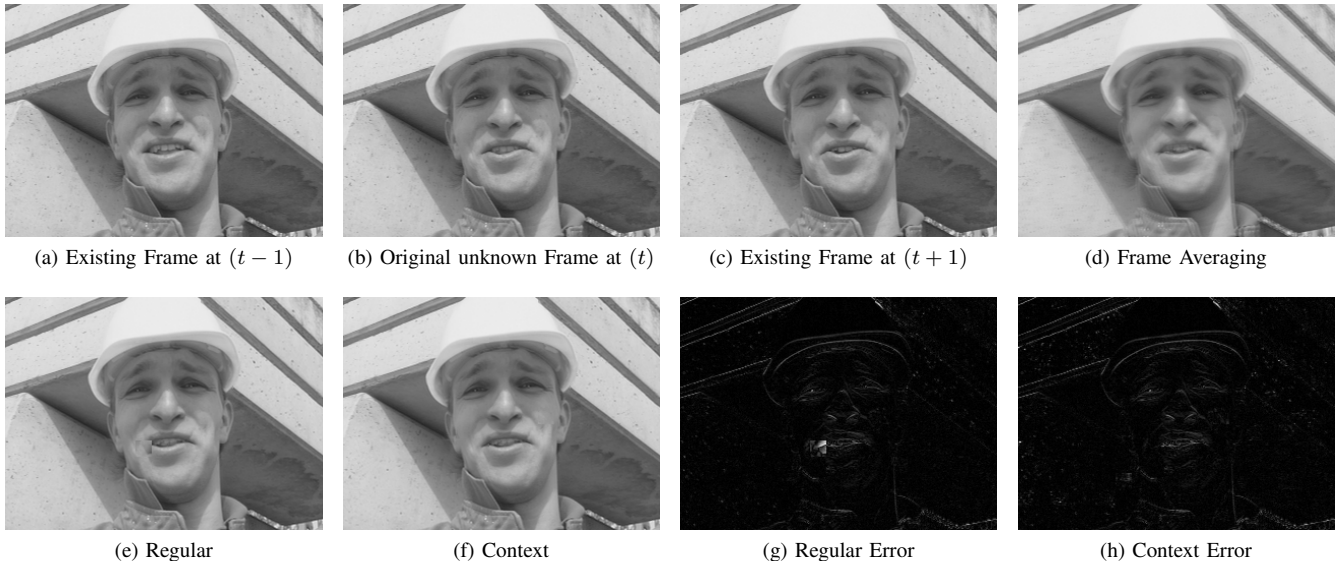


Fig. 9. Visual comparison for temporal up-sampling Foreman sequence by a factor of 2. (a) Existing frame at time $t-1$, (b) Unknown frame at time t (to be interpolated), (c) Existing frame at time $t+1$, (d) Estimation of the t -frame by frame averaging – 29.42dB, (e) Conventional MC-FRUC – 35.04dB, (f) Context-based MC-FRUC – 35.22dB, (g) Absolute value of the error of the conventional approach, and (h) the context-based approach.

time t (Fig. 9b). As can be seen, simple averaging of frames $t-1$ and $t+1$ leads to ghost artifacts (Fig. 9d), while the conventional motion-compensated approach results in much better performance (Fig. 9e). Still, due to errors in the block-matching step, it suffers from visible artifacts/outliers in the reconstruction (e.g. misaligned block close to the mouth). The context-based approach (Fig. 9f) benefits from better block-matching, leading to better estimation of the motion trajectory, and thereby resulting in improved restoration. The achieved improvement is also demonstrated in Fig. 9g and Fig. 9h that show the absolute difference between the interpolated and the original frame, for the conventional and context-based approach (linearly scaled to 0-255 range).

To summarize this experiment, we evaluate the effectiveness of working with context-blocks/patches as a way to improve temporal interpolation. Our experiments show that an improved performance can be achieved by integrating the context-feature to the block-matching step in MC-FRUC, supporting the effectiveness of the context.

V. CONCLUSIONS

The motivation of this paper emerged from the impressive work of Levin et al. [1], [2] that suggest a way to evaluate the denoising bound of image patches. Theoretically, their work claims that larger patches result in better denoising performance compared to using small ones. Nevertheless, in order to ensure meaningful matches, the size of the database (number of samples) should grow exponentially with the

patch-size quickly leading to unrealistic sizes. A ”dense-sampling” assumption of the patch space holds true for small patches, while becoming unattainable when working with large patches.

These observation led us to present the con-patch: A low dimensional vector that benefits from the information that large-patches provide, along with preserving the attractive properties of working with small-patches. This is done by concatenating a small patch to its context-feature – a compact representation of the surroundings/context of the central patch. The proposed feature is obtained by measuring the similarity of a small patch to its neighborhood patches, organized as a normalized histogram. As such, it represents the empirical distribution of the co-occurrences of the small central patch in its large surrounding window.

We tested the effect of the con-patches on several very distinct image processing applications that involve a NN search as their core mechanism. More specifically, the robustness of the con-patches to noise was demonstrated via image denoising. The ability to find meaningful NNs despite missing of spatial information is shown by treating the image SR problem, and better estimation of motion-trajectories is evaluated by integrating the context in MC-FRUC.

We note that our choice of features, while shown to be effective, is rather arbitrary. Further work is required to better design those features, perhaps by learning them.

ACKNOWLEDGMENT

This research was supported by the European Research Council under EU's 7th Framework Program, ERC Grant agreement no. 320649, and by the Intel Collaborative Research Institute for Computational Intelligence.

REFERENCES

- [1] A. Levin and B. Nadler, "Natural image denoising: Optimality and inherent bounds," in *CVPR*. IEEE, 2011, pp. 2833–2840.
- [2] A. Levin, B. Nadler, F. Durand, and W. T. Freeman, "Patch complexity, finite pixel correlations and optimal denoising," in *ECCV*. Springer, 2012, pp. 73–86.
- [3] M. Zontak and M. Irani, "Internal statistics of a single natural image," in *CVPR*. IEEE, 2011, pp. 977–984.
- [4] I. Mosseri, M. Zontak, and M. Irani, "Combining the power of internal and external denoising," in *ICCP*. IEEE, 2013, pp. 1–9.
- [5] E. Luo, S. Chan, and T. Nguyen, "Adaptive image denoising by targeted databases," *IEEE Trans. on Image Processing*, vol. 24, no. 7, pp. 2167–2181, 2015.
- [6] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *CVPR*, vol. 2. IEEE, 2005, pp. 60–65.
- [7] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *ICCV*. IEEE, 2009, pp. 349–356.
- [8] O. Mac Aodha, N. D. Campbell, A. Nair, and G. J. Brostow, "Patch based synthesis for single depth image super-resolution," in *ECCV*. Springer, 2012, pp. 71–84.
- [9] A. Kheradmand and P. Milanfar, "A general framework for regularized, similarity-based image restoration," *IEEE Trans. on Image Processing*, vol. 23, no. 12, pp. 5136–5151, 2014.
- [10] Y. Romano, M. Protter, and M. Elad, "Single image interpolation via adaptive nonlocal sparsity-based modeling," *IEEE Trans. on Image Processing*, vol. 23, no. 7, pp. 3085–3098, 2014.
- [11] B. Furlong, J. Greenberg, and R. Westwater, *Motion estimation algorithms for video compression*. Springer Science & Business Media, 2012, vol. 379.
- [12] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [13] Y. Dar and A. Bruckstein, "Motion-compensated coding and frame rate up-conversion: Models and analysis," *IEEE Trans. on Image Processing*, vol. 24, no. 7, pp. 2051–2066, 2015.
- [14] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.
- [15] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *ICCV*. IEEE, 2011, pp. 479–486.
- [16] Y. Romano and M. Elad, "Boosting of image denoising algorithms," *SIAM Journal on Imaging Sciences*, vol. 8, no. 2, pp. 1187–1219, 2015.
- [17] J. Sulam, B. Ophir, M. Zibulevsky, and M. Elad, "Trainlets: Breaking the dimension barrier in dictionary learning," *IEEE Trans. on Signal Processing*, 2016.
- [18] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *CVPR*. IEEE, 2007, pp. 1–8.
- [19] M. Muja and D. G. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, 2014.
- [20] A. Oliva and A. Torralba, "The role of context in object recognition," *Trends in cognitive sciences*, vol. 11, no. 12, pp. 520–527, 2007.
- [21] J. Sun, J. Zhu, and M. F. Tappen, "Context-constrained hallucination for image super-resolution," in *CVPR*. IEEE, 2010, pp. 231–238.
- [22] L. Sun and J. Hays, "Super-resolution from internet-scale scene matching," in *ICCP*. IEEE, 2012, pp. 1–12.
- [23] T. Ruzic and A. Pizurica, "Context-aware patch-based image inpainting using markov random field modeling," *IEEE Trans. on Image Processing*, vol. 24, no. 1, pp. 444–456, 2015.
- [24] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2005, pp. 886–893.
- [25] A. Foi and G. Boracchi, "Foveated self-similarity in nonlocal image filtering," in *IS&T/SPIE Electronic Imaging*, vol. 8291, no. 829110. International Society for Optics and Photonics, 2012.
- [26] 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 *Proc. 8th Int'l Conf. Computer Vision*, vol. 2, July 2001, pp. 416–423.
- [27] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [28] Y. Romano and M. Elad, "Improving K-SVD denoising by post-processing its method-noise," in *IEEE Int. Conf. on Image Processing*. IEEE, 2013, pp. 435–439.
- [29] —, "Patch-disagreement as a way to improve K-SVD denoising," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, April 2015, pp. 1280–1284.
- [30] G. Ghimpeanu, T. Batard, M. Bertalmio, and S. Levine, "A decomposition framework for image denoising algorithms," *IEEE Trans. on Image Processing*, vol. 25, no. 1, pp. 388–399, 2016.
- [31] V. Pappas and M. Elad, "Multi-scale patch-based image restoration," *IEEE Trans. on Image Processing*, vol. 25, no. 1, pp. 249–261, 2016.
- [32] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.
- [33] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. journal of computer vision*, vol. 47, no. 1-3, pp. 7–42, 2002.
- [34] B.-T. Choi, S.-H. Lee, and S.-J. Ko, "New frame rate up-conversion using bi-directional motion estimation," *IEEE Trans. on Consumer Electronics*, vol. 46, no. 3, pp. 603–609, 2000.
- [35] S.-H. Lee, O. Kwon, and R.-H. Park, "Weighted-adaptive motion-compensated frame rate up-conversion," *IEEE Trans. on Consumer Electronics*, vol. 49, no. 3, pp. 485–492, 2003.
- [36] J. Zhai, K. Yu, J. Li, and S. Li, "A low complexity motion compensated frame interpolation method," in *IEEE Int. Symposium on Circuits and Systems*. IEEE, 2005, pp. 4927–4930.
- [37] T. Q. Vinh, Y.-C. Kim, and S.-H. Hong, "Frame rate up-conversion using forward-backward jointing motion estimation and spatio-temporal motion vector smoothing," in *IEEE Int. Conf. on Computer Engineering & Systems*. IEEE, 2009, pp. 605–609.
- [38] D. Wang, L. Zhang, and A. Vincent, "(motion-compensated frame rate up-conversion part i: Fast multi-frame motion estimation)," *IEEE Trans. on Broadcasting*, vol. 56, no. 2, pp. 133–141, 2010.
- [39] D. Wang, A. Vincent, P. Blanchfield, and R. Klepko, "Motion-compensated frame rate up-conversion Part II: New algorithms for frame interpolation," *IEEE Trans. on Broadcasting*, vol. 56, no. 2, pp. 142–149, 2010.
- [40] J. Dong and Y. Ye, "Adaptive downsampling for high-definition video coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 24, no. 3, pp. 480–488, 2014.