Internal Statistics of a Single Natural Image

Maria Zontak and Michal Irani
Dept. of Computer Science and Applied Mathemathics
The Weizmann Institute of Science, ISRAEL

Abstract

*Statistics of 'natural images' provides useful priors for solving under-constrained problems in Computer Vision. Such statistics is usually obtained from large collections of natural images. We claim that the substantial internal data redundancy within a single natural image (*e.g.*, recurrence of small image patches), gives rise to powerful internal statistics, obtained directly from the image itself. While internal patch recurrence has been used in various applications, we provide a parametric quantification of this property. We show that the likelihood of an image patch to recur at another image location can be expressed parametricly as a function of the spatial distance from the patch, and its gradient content. This "internal parametric prior" is used to improve existing algorithms that rely on patch recurrence. Moreover, we show that internal image-specific statistics is often more powerful than general external statistics, giving rise to more powerful image-specific priors. In particular: (i) Patches tend to recur much more frequently (densely) inside the same image, than in any random external collection of natural images. (ii) To find an equally good external representative patch for all the patches of an image, requires an external database of hundreds of natural images. (iii) Internal statistics often has stronger predictive power than external statistics, indicating that it may potentially give rise to more powerful image-specific priors.*

1. Introduction

Visual reconstruction problems (e.g., denoising, inpainting, super-resolution, etc.), are often under-constrained and ill-posed, thus rely on having good image priors. Such priors range from naive and simple "smoothness" priors, to more sophisticated statistical priors learned from large collections of natural images. To date, natural image statistics are mostly based on models extensively trained on wide external databases of 'natural images'. For example, parametric models (*e.g.*, [16, 13, 15]) impose parametric distribution on natural image responses to local filters. The filters and other parameters of these models are learned using a large database of natural image examples. Although the space of all natural images is sparse [11], trying to capture its wide variety of features with only few parameters is impossible. As a result the learned models reduce to the lowest common denominator of all natural images.

We claim that the substantial internal data redundancy (*e.g.*, recurrence of small 5×5 image patches) gives rise to powerful *internal statistics*, learned directly from the image itself. Internal patch recurrence has been used in various applications, *e.g.* texture synthesis [6], denoising [4, 7], super-resolution [10, 9]. However, the full extent and behavior of these internal patch recurrences and their power relative to external statistics, have never been studied or quantified.

In this paper we parametricly quantify the degree of recurrence of small image patches. We empirically show that the patch density decays rapidly as the spatial distance from the patch location grows, and as its gradient content increases. We further demonstrate that incorporating such parametric knowledge into existing algorithms (*e.g.*, the Non-Local Means denoising [4]) provides improved results.

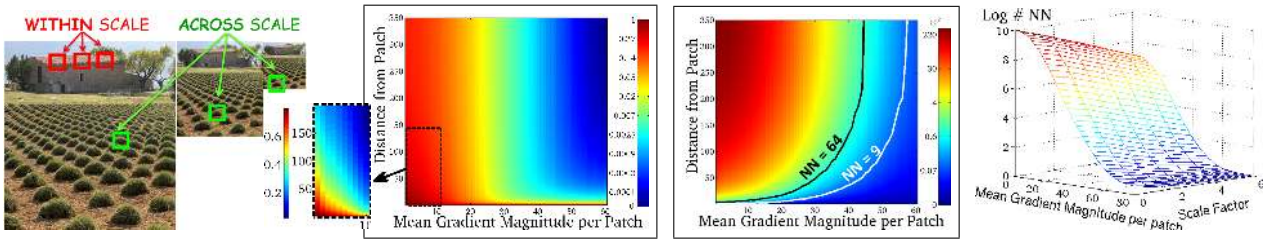
Besides the obvious advantages of internal statistics in terms of low memory and computation demands, we show that the internal image-specific statistics is often *more powerful* than general external image statistics - an observation not necessarily intuitive. Given a patch extracted from an image, it will almost surely recur again in the same image. However, it may not appear in another image. In fact, we show that in order to find equally good external representatives for all image patches of a single image, an external database of *hundreds* of images is required (which may be computationally infeasible to search). Moreover, patches extracted from a natural image, tend to recur much more frequently (densely) inside the same image, than in any random collection of natural images. We further demonstrate that these observations are particularly true for very detailed patches (of high gradient content), which usually contain the most important image details. In addition, we show that the *predictive power* of internal image-specific statistics is often stronger than that of the general external statistics.

Finally, we observe that the patch recurrences within a single image are characterized by a long-tailed distribution. Therefore, compact representations of patches (such as K-SVD [1], Epitome [10] etc.) cannot capture well the full richness of single-image statistics.

The rest of the paper is organized as follows. Sec. 2 presents the quantification of internal patch recurrences. Sec. 3 demonstrates how these quantifications can be incorporated into existing algorithms to obtain improved results. Sec. 4 compares the descriptive and predictive properties of the internal vs. external statistics. Sec. 5 discusses the limitations of sparse representation of patches.

2. Quantifying Internal Statistics

Fig. 1.a (courtesy of [9]) schematically illustrates the notion of "patch recurrence" within a single natural image (similar patches are marked by same colors). For illustra-



(a) Patch Recurrence (Courtesy of [9]) (b) Patch Density (c) $NN(dist, |grad|)$ (d) $NN(n, |grad|)$

Figure 1. **Internal Statistics of a Single Natural Image**. *(a) Small image patches tend to recur within the source image, and across its coarser scales; (b)-(c) The empirical* $Density(dist, |grad|)$ *of patches and the number of "similar" patches,* $NN(dist, |grad|)$*, as a function of the mean gradient magnitude* $|grad|$ *in the patch, and the spatial distance* dist *from the patch location (Red signifies high values, Blue signifies low values); (d) The* $\log NN(n, |grad|)$ *shown for various image scales* ($n = 0, .., 6$).

tion purpose, the patches were chosen large and on clearly repetitive structures. However, when much smaller (5×5) image patches are used, such patch repetitions occur abundantly within and across image scales, even when we do not visually perceive any obvious repetitive structure in the image. Glasner *et al*. [9] empirically showed that most of the patches in a natural image have many similar patches at the same image scale, and at coarser images scales. In this section we provide a formal *parametric* quantification of the degree of internal recurrence of small 5×5 patches.

Most of the patches in a natural image are rather smooth, and only a small percent contain important image details (edges, corners, etc.) These differences are expressed in different spatial gradient magnitudes in patches. We observe that smooth patches recur much more frequently in the image than detailed patches. We further observe that an image patch is much more likely to recur near itself than far away. Therefore, our experiments (and plotted graphs) are expressed in terms of the "mean gradient magnitude" $|grad|$ of a patch, and the "spatial distance" $dist$ to the patch.

Our experiments were conducted on the 300 images of the *Berkeley Segmentation Database*[1] (**BSD**). For each image patch p, we estimated its *empirical density* within an image neighborhood $\mathcal{N}_{dist}$ of radius "$dist$" around the patch, using Parzen window estimation [12]: $density(p; dist) = \sum_{p_i \in \mathcal{N}_{dist}} \mathcal{K}_h(\|p - p_i\|_2^2)/area(\mathcal{N}_{dist})$, where p_i are all the image patches within a spatial neighborhood $\mathcal{N}_{dist}$, and $\mathcal{K}_h(\cdot)$ is a Gaussian kernel[2]. Averaging these individually-computed patch densities over the set of all patches with the same gradient magnitude $|grad|$, we obtain the following average density:

$$Density(dist, |grad|) = Mean_{p_j \, of \, |grad|} density(p_j, dist).$$

The average number of "good Nearest Neighbors" NN within a distance $dist$ from the patch, is defined as:

$$NN(dist, |grad|) = Density(dist, |grad|) \cdot area(\mathcal{N}_{dist}) \quad (1)$$

Note that the Parzen estimation does not distinguish between 10 perfectly similar patches, and 100 partially similar patches. We loosely refer to these as 10 good NNs. Fig. 1.b displays the empirical density $Density(dist, |grad|)$ and the number of "similar" patches $NN(dist, |grad|)$, both as a function of the mean gradient magnitude $|grad|$ of the patch, and the spatial distance $dist$ from the patch location. (In both maps, red signifies high values, blue signifies low values.) Observing these maps, we note that:

(i) Smooth patches recur very frequently, whereas highly structured patches recur much less frequently.

(ii) A patch tends to recur densely in its closest vicinity (small $dist$), and its frequency of recurrence decays rapidly as the distance from the patch increases (see the zoomed-in part in Fig.1.b). Namely, patches in a natural image are likely to reside in clusters of similar patches. This explains why denoising algorithms, such as Non-Local Means [4] and BM3D [5] work well, despite the fact their patch search is restricted to small neighborhoods around each patch.

(iii) Various patch-based applications require obtaining *enough* similar patches for every image patch (*e.g*., Super-Resolution [9], denoising [4], etc.) From Fig. 1.c we note that for a fixed number of similar patches ($NN = const$), patches of different gradient content need to search for nearest neighbors at different distances. For smooth patches, it suffices to search locally, whereas the higher the gradient magnitude, the larger the search region becomes. In fact, one can observe that the *level-sets* in Fig. 1.c (which correspond to a fixed number of Nearest Neighbors), have *exponential shapes* (*e.g*., see the white and black curves, which corresponds to a level-sets of $NN = 9$ and $NN = 64$). In other words, the distance $dist$ in which the nearest neighbor search should be performed grows *exponentially* with the gradient content of the patch $|grad|$.

By empirically fitting an exponential function to the level-set curves (for many fixed NNs), we obtained the following exponential relation between $dist$ and $|grad|$:

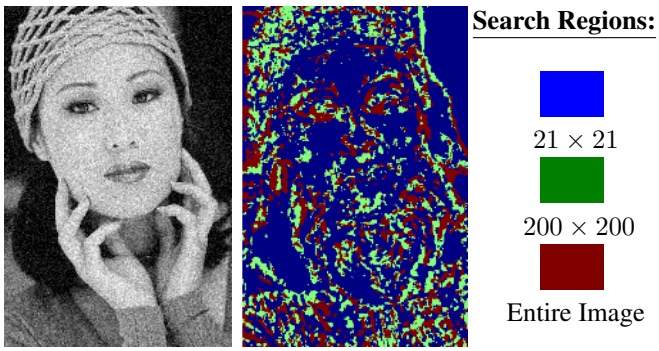
$$dist(|grad|) = \beta_1 + \beta_2 \cdot \exp(|grad|/10) \ ,$$

where β_1 and β_2 depend on the fixed NN (are second order polynomials of $\sqrt{NN}$):

$$\beta_1(NN) = 5 \cdot 10^{-3} NN + 0.09\sqrt{NN} - 0.044$$
$$\beta_2(NN) = 7.3 \cdot 10^{-4} NN + 0.3235\sqrt{NN} - 0.35.$$

[1] www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench

[2] Although L2-norm may not be an optimal measure for patch similarity, it is often used in existing patch-based applications. Since we want to show how quantifying internal patch statistics improves such applications, we follow the L2 convention.



(a) Noisy Image (b) Preferred Search Regions

Figure 2. **Preferred Search Regions per patch in NLM:** *Smooth patches obtain better denoising results in local search regions; detailed patches benefit from large search regions - See text.*

This can be rewritten as:

$$dist(NN, |grad|) = \beta_1(NN) + \beta_2(NN) \cdot e^{|grad|/10} . \quad (2)$$

Eq. 2 gives an explicit parametric expression to determine the search region needed in order to find a desired number of good nearest neighbors NNs for a patch. In Sec. 3 we show that this parametric expression can serve as a better prior in the Non-Local Means denoising algorithm of [4].

We further note that Eq. 2 is quadratic in $\sqrt{NN}$, of the form:

$$a \cdot \sqrt{NN}^2 + b \cdot \sqrt{NN} + c = 0,$$

$$\text{where : } a = 0.001 \cdot (5 + 0.73 \cdot \exp(|grad|/10))$$
$$b = 0.1 \cdot (0.9 + 3.24 \cdot \exp(|grad|/10))$$
$$c = -0.1 \cdot (0.44 + 3.5 \cdot \exp(|grad|/10) + dist).$$

Solving for its single valid root yields a closed-form expression of NN as a function of $dist$ and $|grad|$:

$$NN(dist, |grad|) = (\frac{-b + \sqrt{b^2 - 4ac}}{2a})^2. \quad (3)$$

Eq. 3 provides an estimate for the expected number of good neighbors a patch has within a predetermined region. This parametric expression provides a good approximation (up to a mean error of 4%) to the empirical function NN computed using Eq. 1 and visually depicted in Fig. 1.c. An equivalent expression can be derived for $Density(dist, |grad|)$ of Fig. 1.b using Eq. 1. A visual comparison of the *empirical* functions in Figs. 1.b,c and their *parametric* approximations can be found in ***www.wisdom.weizmann.ac.il/~vision/SingleImageStatsitics.html***

Finally, we explore the statistics of patch recurrence across coarser scales of a natural image. We say that an image patch recurs in another scale, if it appears "as is" (without down-scaling the patch) in a scaled-down version of the image. For each image I we generated a pyramid of images of decreasing resolutions $\{I_n\}$, scaled down by factors of $s = 0.8^n$, $n = 0, .., 6$, with $I_0 = I$. For each patch in I we measured its recurrence density in $I_0, .., I_6$ (in the entire image). Surprisingly, the patch recurrence density in I and in its coarser pyramid levels is approximately the same. The number of Nearest Neighbors decreases in coarser levels, with the decrease in image area: $NN(I_n, |grad|) \approx s^2 NN(I, |grad|) = 0.8^{2n} NN(I, |grad|)$. This entails that: $\log NN(I_n, |grad|) = -0.223n \cdot \log NN(I, |grad|)$. Indeed, Fig. 1.d displays this linear relation between the log number of Nearest Neighbors and the scale level n.

3. Using Internal Statistics to Improve Priors

We show that the quantifications presented in Eq. 2 can be incorporated into existing algorithms that exploit internal patch redundancy, to improve their results. One such example is the Non-Local Means (**NLM**) denoising algorithm [4]. In that algorithm, the central pixel in each patch is replaced by the mean value obtained from other image patches, weighted by their degree of similarity to the source patch. For computational reasons, the authors restrict the algorithm to a *local* 21×21 search region around each pixel.

Based on our observations in Sec. 2, that a patch density is high within its closest vicinity, it is not surprising that NLM works well, despite its relatively local search. However, *what is surprising*, is that the local search is often *preferable* over a 'global' search in the entire image. A similar observation was also made by [2]. Fig. 2 visualizes these surprising findings. We ran the NLM algorithm on the noisy image of Fig. 2.a, with 3 different search regions: (i) 21×21, (ii) 200×200, (iii) the entire image. For each pixel in the image, we marked which of the 3 search regions gave it the smallest error relative to the ground-truth clean image (Fig. 2.b). Note that smooth patches benefit more from constrained local search, whereas textured patches with high gradients benefit from global search. We will later show that incorporating Eq. 2 into NLM can be used for estimating an 'optimal' search region per patch.

Our interpretation of this surprising phenomenon is the following: Let $p_n = p + n$ be a noisy version of an image patch p. When p is a smooth patch, the noise n dominates p_n, inducing new "patterns". Moreover, although the global mean of the noise is 0, its local mean within small 5×5 patches is often non-zero, inducing a change in the patch mean. Extending the search region to the entire image increases the chance of *over-fitting the noise*, thus preserving effects of the noise n. In contrast, there is very little chance of finding a good match to the noise pattern in a small neighborhood. Moreover, the local vicinity of a smooth patch is sufficient for finding *many* 'correct' nearest neighbors to the signal p (Sec. 2). Local search thus increases the chance of fitting the 'signal' p and not the 'noise' n for such patches.

Unlike smooth patches, high-gradient patches will benefit from a large search region. In such patches, the noisy patch p_n is dominated by the signal p. Therefore, a global search in the entire image is not risky. Moreover, the search region *must be large* in order to find enough nearest neighbors for high-gradient patches (as observed in Sec. 2).

We empirically show that this phenomenon holds for natural images in general. Our experimental setting was the following: We applied the NLM algorithm (using the code of Morel *www.mi.parisdescartes.fr/~buades/recerca.html*

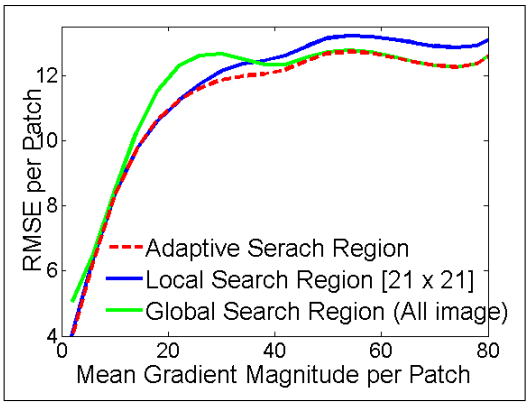


Figure 3. **Non-Local Means applied to different search region:** *Locally, Globally and Adaptively (see text for details).*

to many natural images with added Gaussian noise of std $\sigma = 15$. We applied the algorithm once locally to a 21×21 search region ('local' NLM – the default setting of the code), and once using the entire image as a search region ('global' NLM). Fig. 3 shows the resulting errors relative to the ground truth clean image (averaged over 100 images).

We next show that incorporating Eq. 2 into NLM can be used for estimating an 'optimal' search region per patch, yielding improved denoising results. Suppose we want at least k good representatives per patch (to be averaged to recover the clean patch p). Eq. 2 provides an explicit expression for the radius of the search region needed to obtain k 'nearest neighbors' per patch. The exponential white and black curves in Fig. 1.c show two such examples, for $k = 9$ and 64. In our experiment we used $k = 64$. (Note: for $|grad| > 40$, the search region is already the entire image.)

Dealing with noisy patches, we approximate their "clean" gradient content by: $|grad|_p^2 = |grad|_{p_n}^2 - \sigma_{\text{noise}^2}$. This follows from the fact that n (the noise in the patch) and p (the clean patch) are independent, therefore the variance of the noisy patch, p_n, is the sum of their variances. We experimentally found that for patches with $|grad| < 50$, their gradient content linearly relates to their variances.

Fig. 3 shows the resulting NLM after incorporating the adaptive search region based on Eq. 2. As can be seen, it provides improved results with respect to both 'local' and 'global' NLM. The purpose of this is not to claim a state-of-the-art denoising algorithm, but rather to show that incorporating quantitative knowledge about internal image statistics can improve existing algorithms that rely on such statistics.

4. Internal vs. External Statistics

Besides the obvious advantages of internal statistics in terms of lower memory and computation demands, the internal *image-specific* statistics is often more powerful than *general external* image statistics. We compare these two types of statistics according to their degree of "expressiveness" and "predictive power" (both to be defined shortly). The internal statistics of an image is based on the collection of all the patches extracted from the image and its multi-scale versions (as explained in Sec. 2). The external statistics is based on all the patches extracted from a database of different natural images (taken from the BSD Train-Set). The size of the external database ranges from 5 images (small database) to 200 images (large database)[3]. Please note that our analysis does not hold for *class-specific* external database (which are extremely useful, even if small). The applicability of such dedicated databases is limited to handling only images from the same specific class. Our analysis assumes *general natural images*.

4.1. Expressiveness

"Expressiveness" measures the degree of similarity of a 5×5 patch to its most similar patches found internally vs. externally. Internally, the patch itself and its immediate local vicinity are excluded from a search. We calculate the L_2 distance between two patches (after removing their DC[4]). Fig. 4 displays the error induced by replacing each patch in the input image (Fig. 4.a) with its most similar patch, either from an external database of images (Fig. 4.b-d), or from the image itself and its multi-scale versions (Fig. 4.e). We observe that smooth patches can be found quite easily in an external database, as well as in the image itself. However, this is not true for detailed patches (edges, corners etc.), which require as many as 200 images to find equally good external representatives to those found internally.

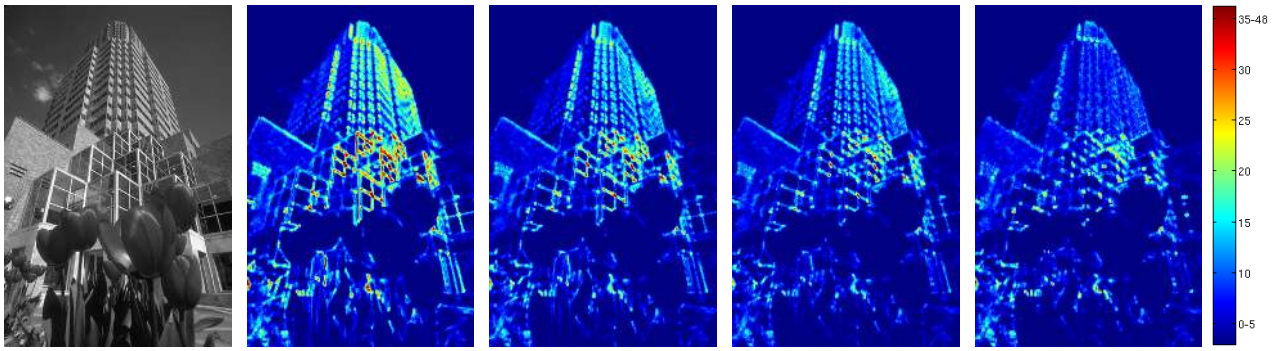
Fig. 5 shows the same analysis as in Fig. 4, empirically conducted over hundreds of images (more than 15 million 5×5 patches). The errors were computed separately for each gradient magnitude (using RMSE, averaged over all patches with the *same gradient magnitude*). It can be observed that for small external databases (up to 40 images), only relatively smooth patches ($|\text{grad}| < 20$) are similarly represented internally and externally. However, patches with higher gradient content require external databases of *hundreds of images* in order to obtain an external patch of similar quality to the one found internally.

Note, that we would fail to see the observation of Figs. 4 and 5, if we were to compute the *mean error* averaged over the *entire image* (which is a widely used measure for evaluating algorithms). More than 80% of the patches in natural images tend to have low mean gradient magnitude (≤ 20). Therefore, any averaging process that does not take into account the uneven distribution of gradient magnitudes, is governed by the errors in the smooth/undetailed regions of the image. Thus, damages in the most important fine details of the image are not reflected in a global RMSE measure.

We further compared the *density* of patch recurrence, externally vs. internally. Fig. 6 shows that image patches tend

[3] One could potentially add multi-scale versions of the external images to the external database, but this would come at the expense of the variety of different images (assuming the same number of patches in the external database is maintained). The external databases benefits more from having a larger variety of images than multiple scaled versions of the same images.

[4] When the DC is not removed, the advantage of the internal statistics over the external one is even more pronounced.



(a) Input Image (b) Ext. DB - 5 imgs. (c) Ext. DB - 40 imgs. (d) Ext. DB - 200 imgs. (e) Internally (Error values)

Figure 4. **External vs. Internal "Expressiveness".** *Errors induced by replacing each patch from the input image with its most similar patch found in: (b)-(d) an external database of 5, 40, 200 images, vs. (e) internally, within the input image (excluding the patch itself and its immediate local vicinity). Red signifies high errors, blue signifies low errors. Patches obtain lower error internally than externally.*

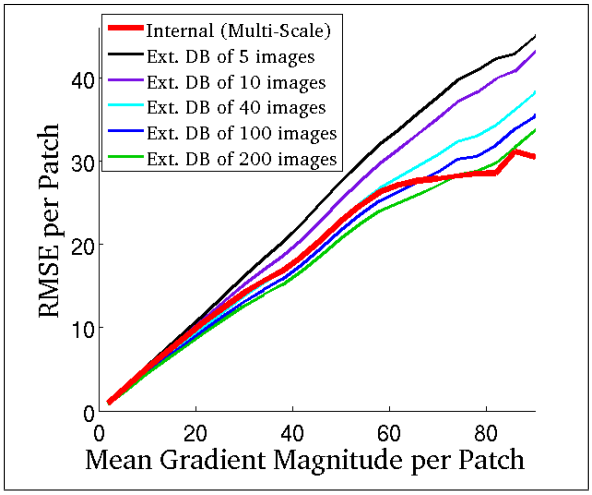


Figure 5. **External vs. Internal "Expressiveness"** *(Statistics on 100 images)***.** *Only relatively smooth patches (*$|\text{grad}| < 20$*) are well represented using small external databases (up to 40 images). Other patches, with higher gradient content, require external databases of* ***hundreds of images*** *in order to obtain external representation of similar quality to the internal representation.*

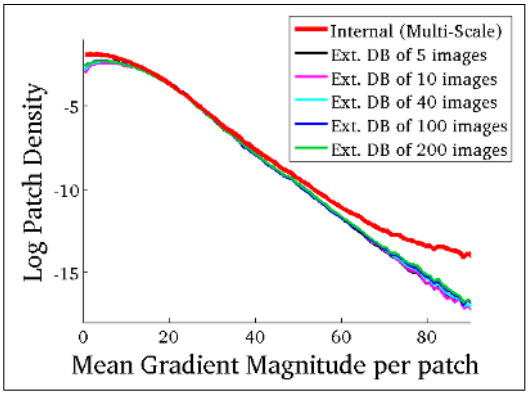


Figure 6. **Density of Patch Recurrence:** *Patches recur more frequently inside the same image than in a random external collection of natural images (regardless of its size). Note that the patch density is displayed in log scale values.*

to recur much more *frequently* inside their own image than in any random collection of natural images, regardless of its size (Note that the patch density is displayed in log scale values). This phenomenon is particularly true for highly detailed patches, which are often the most important ones.

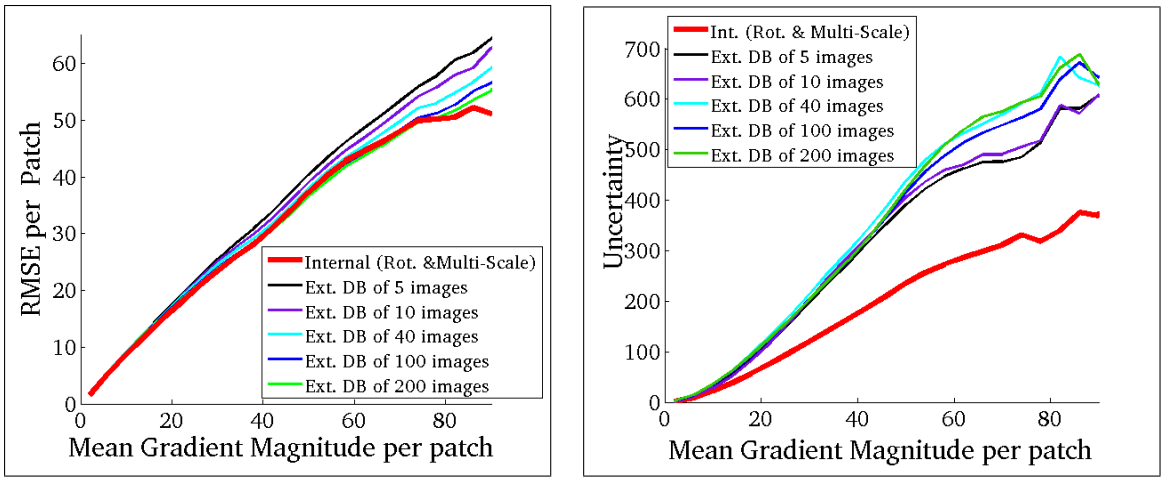
4.2. Predictive Power

Statistical priors are often used to constrain ill-posed problems in Computer Vision. The quality of a prior is determined by *how well it predicts* the 'correct' solution among the infinitely many possible solutions of the underdetermined problem. In this section we compare the "predictive power" of the internal statistics vs. external statistics, when the same prediction method is applied to both of them. As an example test case, we chose the ill-posed problem of *Super-Resolution* (image upsampling), and the prediction method of *"Example-based Super-Resolution"* [8].

In Example-Based SR, a Database of 'examples' of high-res/low-res pairs of image patches $\{(h_i, l_i)\}_{i=1}^{n}$ is provided (usually with a relative scale factor of 2). Given an input image L, its high-resolution (upsampled) version H is generated ("hallucinated"), by using the example pairs as 'predictors' (priors) on how to upsample the low-resolution patches of L. This yields the most likely high-resolution image H of L, given the database of examples (predictors).

In order to compare the predictive power of internal vs. external statistics in the above setting, we performed the following experiment (repeated for all 100 images of the BSD Test-set): Given a natural image I (the "ground truth" high-res image denoted also as H_{GT}), we downscale I to half its original resolution, to generate the low-res input L. We generate an *external database* of high-res/low-res examples (from the 200 images in the BSD Train-set), and an *internal database* of high-res/low-res examples (from L and its down-scaled versions). The high-res/low-res pairs were generated both internally and externally by downscaling the available images by a factor of 2:1, and extracting all the corresponding pairs of patches from the two scales.

Note that unlike previous experiments, here we are at an immediate disadvantage, due to the fact that the 'internal image' L is 1/4 of the size (area) of any individual external



(a) **Prediction Error** (b) **Prediction Uncertainty**

Figure 7. **External vs. Internal "Predictive Power", evaluated for "Example-Based Super-Resolution"** – *See text.*

image. A ratio of 1:200 between the internal/external *number of images* now translates to a ratio of 1:800 in the number of examples to learn from (the high-res/low-res pairs of patches). We therefore add to the internal database rotated versions of L (at $\pm 45^\circ$), increasing the space of internal pairs of patches back to its original internal/ external ratio.

Now, for every 5×5 patch, $l \in L$, we search for its $k = 9$ low-res Nearest-Neighbors $\{l_i\}_{i=1}^k$ in the internal/external databases. Their corresponding high-res patches, $\{h_i\}_{i=1}^k$, which serve as individual predictors, are averaged to recover the overall high-res estimate $\hat{h}$ of l:

$$\hat{h} = \frac{\sum_i w_i \cdot h_i}{\sum_i w_i}, \quad \text{where} \quad w_i = \exp - \frac{\|l - l_i\|_2^2}{2\sigma^2}. \qquad (4)$$

For each high-res ground truth patch, h_{GT}, we measure:
(i) The Prediction Error: $\|h_{GT} - \hat{h}\|_2^2$.
(ii) The Prediction Uncertainty: The weighted variance of the predictors $\{h_i\}_{i=1}^k$ is approximated using $trace\,(Cov_W(h_i, h_j))$ (with the same weights as above).

The second measure serves as a reliability measure of the prediction. The high-res predictors, $\{h_i\}_{i=1}^k$, should not only be individually close to the true h_{GT} (low prediction error), but should also be mutually consistent with each other (low uncertainty). High uncertainty (entropy) among all the high-res candidates $\{h_i\}_{i=1}^k$ of a given low-res patch l, indicates high *ambiguity* in the predicted high-res patch. This results in visual artifacts, like 'hallucinations' and blurring (due to multiple *inconsistent* high-res interpretations).

Fig. 7 displays the statistics of these two measures, averaged over all the patches from 100 natural images. Note that it requires hundreds of images to achieve external prediction error similar to the internal one. Moreover, in the high gradient patches, the internal prediction error is still lower than the external error, even for large external databases. Although these patches are relatively sparse in the image, these are the most critical patches in Super-Resolution (the edges, corners, and high-detailed image parts). This is where the increase in resolution is observed.

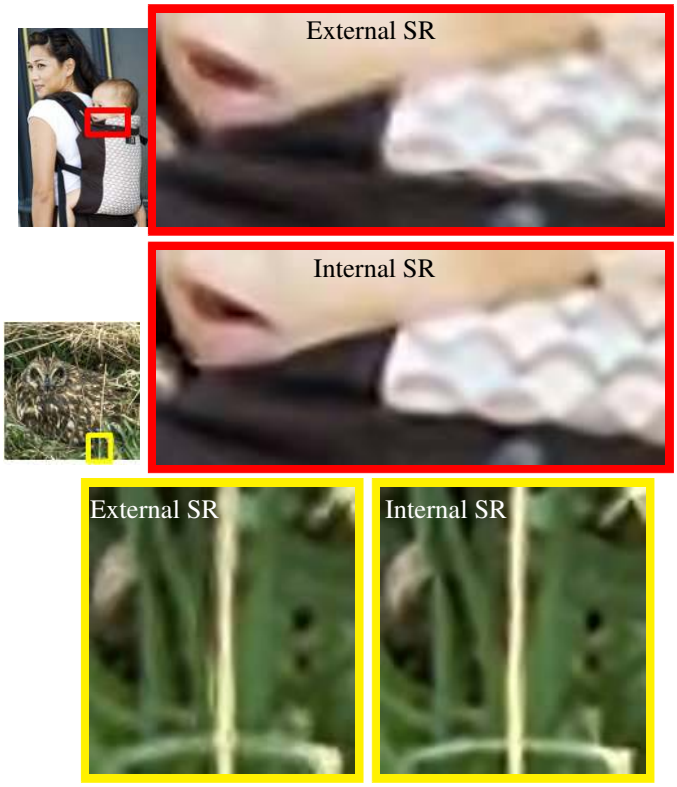


Figure 8. **Super Resolution using External vs. Internal examples.** *External super-resolution exhibits more hallucinations and blur artifacts.* ***More results can be found in*** *www.wisdom.weizmann.ac.il/~vision/SingleImageStatsitics.html*

Moreover, the prediction uncertainty is much higher externally than internally (for any database size), alluding to the fact that general external statistics is more prone to 'hallucinations' than internal image-specific statistics. Fig. 8 illustrates this, showing visual examples of internal vs. external example-based prediction. The external database of 200 general images produces inferior results, displaying hallucination of details and more blurriness.

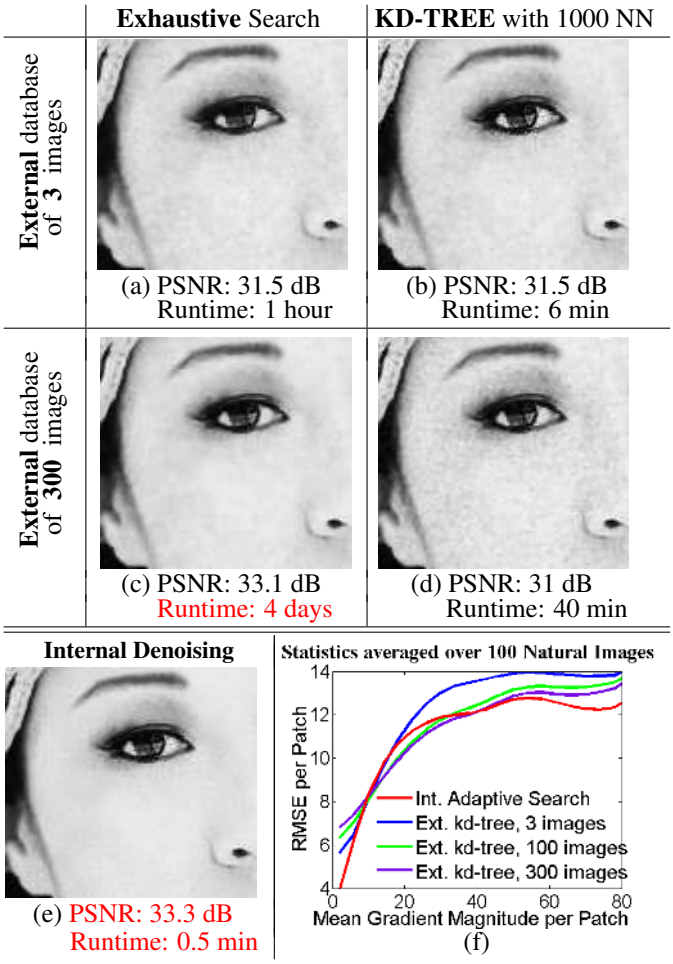


Figure 9. **External vs. Internal Denoising** – *See text.*

Exploiting Huge Databases: What happens if we push the envelop to *huge* external databases, e.g., all the images on the Internet? Patches will undoubtedly find better matches! But there are two major problems with huge databases:
(i) *Predictive power:* Our purpose is not only to find similar patches, but to use them as predicting priors in ill-posed problems. Fig. 7.b shows that *larger* external databases exhibit *lower* predictive power (higher uncertainty). Huge databases are likely to exhibit very low predictive power.
(ii) *Computation:* High gradient patches (which are the most informative ones) are *rare* and have very low density in an external database, *regardless* of its size (see Fig. 6). As such, these patches *cannot* be captured well by any compact quantized representation (K-SVD, PCA, etc. – see Sec. 5). Thus, finding such patches in *huge* external databases requires an *extensive* search. This is computationally infeasible for any practical application. In contrast, internally these patches have sufficiently good Nearest-Neighbors (comparable to hundreds of external images - Figs. 4,5), and their search space is limited to a single image (and even better, to the patch local vicinity - Fig. 1.b,c).

The problem of huge databases is further exemplified in Fig. 9, this time in the context of image denoising. Denoising is performed in the NLM [4] manner, once using internal (noisy) patches, and once using an external database of clean images (averaging over similar external patches). In principal, increasing the number of clean patches in the database should improve the denoising results, as observed in Fig. 9.a,c (denoising using 300 external images yields cleaner results than denoising using 3 external images). Yet, with 300 images, the external denoising is still *inferior* to internal denoising (Fig. 9.e). Moreover, it comes with an enormous run-time of *4 days* (on a Linux 2668 MHz machine), vs. *0.5 minute* run-time for internal denoising.

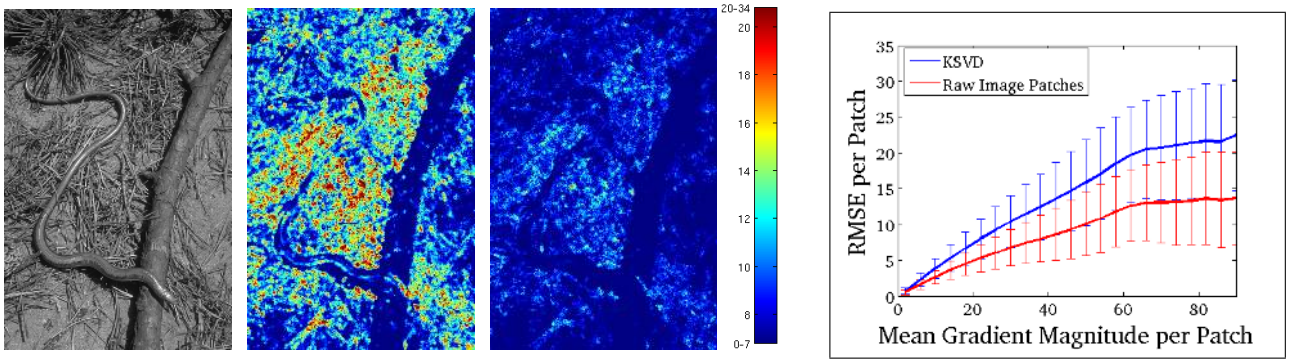
To reduce the external run time, for each patch in the noisy image we limit its external averaging to only 1000 external Nearest Neighbors (using KD-tree). While the reduced run-times are still high, the limited NN-search induces a new problem: *Enlarging* the external database (*e.g.*, from 3 to 300 – Fig. 9.b,d) now yields *worse* (noisier) results! This surprising artifact is due to over-fitting the noise in smooth image patches (as also occurs in global-NLM – see Sec. 3). While denoising of the detailed (infrequent) patches improves as the database grows, denoising of smooth patches (the majority) becomes worse. The graph in Fig. 9f shows that this observation holds in general for noisy natural images. Note that unlike super-resolution, in image-denoising the smooth patches are the most important ones (this is where the noise is most visible). Global PSNR is thus an adequate measure in denoising (larger dB is better).

5. Limitations of Compact Representation

Compact representations were proposed to take advantage of the redundancy of patches within an image (e.g., [7, 10]). While these are very useful for many applications, they cannot capture well the full richness of single-image statistics. In fact, as we will see, they harm the most informative (detailed) patches in the image.

As noted in [14, 3], when image descriptors (e.g, SIFT) are divided into fine bins, the bin-density follows a power-law (also known as 'long-tail' or 'heavy-tail' distributions). The long-tail behavior holds also for image patches. This results from the fact that many *different* high-gradient image patches have very low density (i.e., each of them recurs rarely in the image). Such patches are found in low-density regions in the 'space of all image patches', rather isolated. Namely, there are almost no clusters around these patches. Any quantization/clustering process applied to obtain a compact representation, will represent well the most frequent clusterable elements (smoother patches), whereas the infrequent/unclusterable elements (high-gradient patches) will suffer from high quantization errors. This property is inherent to long-tailed distributions, independently of the clustering/quantization method.

To illustrate this we conduct the following experiment. Each image patch is represented in two ways: (i) *Using K-SVD:* Every patch is represented as a linear combination of 3 elements from a 256-element K-SVD dictionary



(a) Input Image (b) K-SVD (c) Raw Patches Error Vals. (d) Statistics over 300 natural images

Figure 10. **Compact Representation vs. Raw Image Information.** *K-SVD introduces high errors in the most informative detailed image patches. In (b) and (c) red corresponds to high errors, blue correspond to low errors; (d) Graph showing mean error and standard deviation of the errors computed on 300 images.*

built from 5×5 patches of the image, plus the mean intensity value (DC) of the patch (using the K-SVD code of *www.cs.technion.ac.il/~ronrubin/software.html*) (ii) *Using raw image patches:* A linear combination of 3 other patches in the same image (*no multi-scale*), plus the patch DC.

Fig. 10.a-c shows a visual result on one image. The 'quantization' error induced by K-SVD dominates in the detailed parts of the image, and is significantly higher than when using raw image patches. Fig. 10.d further shows that this observation holds in general for natural images (statistics accumulated over the 300 images of the BSD).

Finally, we should note that adding those badly represented patches to the compact representation will eliminate its compactness. For example (doing a "back-of-the-envelope" calculations), given a 256×256 image (65536 bytes) and a K-SVD dictionary of 256 elements of 5×5 (6400 bytes), the initial saving in storage space is $1/10$. Let's say we add the 3% most isolated image patches (that are poorly represented by K-SVD). This adds $3\% \cdot 256^2 \cdot 5^2$ = 49152 bytes (almost the original image size!) In other words, patches are already represented quite compactly in the image itself (due to their built-in overlaps with each other), providing the full richness of all image patches. Moreover, the raw image preserves geometric information of where to look for similar patches (Sec. 2), while this information is lost in compact representations.

6. Conclusions

We show that internal patch redundancy within a single natural image rapidly decays with the growth of the spatial distance from the patch, and its gradient content. This yields a new "internal parametric prior", which can be used to improve performance of existing algorithms (e.g., NLM). We further show that besides the obvious advantages of internal image statistics (low memory and low computation costs), it also tends to be *more powerful* than general external statistics (e.g., has better predictive power) – an observation not necessarily intuitive. Finally, we observe that compact representations of patches (e.g., K-SVD) cannot capture well the full richness of single-image statistics; yet, a single raw image is compact enough on its own, rich enough, and preserves geometric information of where to look for patches. These observations open the door to new more powerful image-specific priors.

Acknowledgement: We thank L. Karlinsky for his comments. The work was funded by the Israel Science Foundation.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. Design of dictionaries for sparse representation. In *SPARS*, 2005. 977

[2] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein. The generalized patchmatch correspondence algorithm. In *ECCV*, 2010. 979

[3] O. Boiman. E. Shechtman, M. Irani. In Defense of Nearest-Neighbor Based Image Classification. In *CVPR*, 2008. 983

[4] A. Buadess, B. Coll, and J. Morel. A non local algorithm for image denoising. In *CVPR*, 2005. 977, 978, 979, 983

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3d transform-domain collaborative filtering. *IEEE T-IP*, 16(8), 2007. 978

[6] A. Efros and T. Leung. Texture synthesis by non-paramaetric sampling. In *ICCV*, 1999. 977

[7] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE T-IP*, 54(12), 2006. 977, 983

[8] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2), 2002. 981

[9] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *ICCV*, 2009. 977, 978

[10] N. Jojic, B. Frey, and A. Kannan. Epitomic analysis of appearance and shape. In *ICCV*, 2003. 977, 983

[11] B. A. Olshausen and D. J. Field. Sparce coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37(23), 1997. 977

[12] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33, 1962. 978

[13] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *CVPR*, 2005. 977

[14] T. Tuytelaars and C. Schmid. Vector quantizing feature space with a regular lattice. In *ICCV*, 2007. 983

[15] Y. Weiss and W. T. Freeman. What makes a good model of natural images. In *CVPR*, 2007. 977

[16] S. C. Zhu and D. Mumford. Prior learning and gibbs reaction-diffusion. *PAMI*, 19(11), 1997. 977