Statistical Structure of Natural 4×4 Image Patches

Kostadin Koroutchev and José R. Dorronsoro*

Depto. de Ingeniería Informática and Instituto de Ingeniería del Conocimiento
Universidad Autónoma de Madrid, 28049 Madrid, Spain

Abstract. The direct computation of natural image block statistics is unfeasible due to the huge domain space. In this paper we shall propose a procedure to collect block statistics on compressed versions of natural 4×4 patches. If the reconstructed blocks are close enough to the original ones, these statistics can clearly be quite representative of the true natural patch statistics. We shall work with a fractal image compression–inspired codebook scheme, in which we will compute for each block B its contrast σ, brightness μ and a normalized codebook approximation D^B of $(B - \mu)/\sigma$. Entropy and mutual information estimates suggest a first order approximation $p(B) \simeq p(D^B)p(\mu)p(\sigma)$ of the probability $p(B)$ of a given natural block, while a more precise approximation can be written as $p(B) \simeq p(D^B)p(\mu)p(\sigma)\Phi(||\nabla B||)$. We shall also study the structure of $p(\sigma)$ and $p(D)$, the more relevant probability components. The first one presents an exponential behavior for non flat patches, while $p(D)$ behaves uniformly with respecto to volume in patch space.

1 Introduction

Natural images, that is, those derived from natural scenes, have a distinctive nature that makes them far from random. In particular, they convey "information" that allows their processing by the human visual system. In fact, natural image information is not distributed uniformly over the image: there are parts that are most relevant to the human visual system, while other are far less relevant. It is therefore clear that the undestanding of their statistical behavior it is extremely important, not only for basic human visual processing but also for a number of everyday visual information processing tasks such as for instance efficient static and dynamic image compression. A large effort has been undertaken in that direction [7, 2, 9]; a thorough and recent survey is in [8]. In any case, it can be quite easily seen that direct statitistics computation for 4×4 natural image blocks is not currently possible. In fact, state of the art lossless image compression (see e.g. [10]) can achieve for 8 bit gray level images compression rates down to 2.5 bits per pixel. Thus, a block of size 4×4 would requires in average about $16 \times 2.5 = 40$ bits. Assuming that this representation is close to the informational limit of the lossless representation of the image, it follows that

* With partial support of Spain's CICyT, TIC 01–572.

A. Fred et al. (Eds.): SSPR&SPR 2004, LNCS 3138, pp. 452–460, 2004.

Fig. 1. Statistics computed from other codebooks are similar, provided the source image is "rich" enough, as exemplified here.

natural block statistics require at least 2^{40} 4×4 blocks or, in other words, about $2^{40-16} \simeq 16 \times 10^6$ natural 1024×1024 images. Direct statistics computation is clearly not possible today because of, among other things, the lack of so many machine readable raw images.

We shall work here with the equivalent representation $(\tilde{B}, \sigma, \mu)$ of a natural block B, with σ, μ the standard deviation and mean of B and $\tilde{B}$ its normalization $\tilde{B} = (B - \mu)/\sigma$. Direct statistics are clearly possible for the 1–dimensional σ and μ, while this is not so for $\tilde{B}$. To estimate them we shall approximate natural patches $\tilde{B}$ by normalized blocks D^B extracted from a given codebook. Since an approximation $\tilde{B} \simeq D^B$ immediately translates into the affine approximation $B \simeq \sigma D^B + \mu$, it is natural to try to derive D through fractal image compression (FIC) techniques [1]. If a good reconstruction quality is obtained, the (D_B, σ, μ) statistics should provide meaningful approximations to those of $(\tilde{B}, \sigma, \mu)$. This will be done in section 2, where we shall introduce a concrete FIC codebook, namely, 4×4 domains extracted from the well known Lena image, and shall use it to approximate about 280 million natural 4×4 patches extracted from the well known Van Hateren database [2]. Although the concrete codebook used certainly influences the resulting statistics, we have obtained similar results using codebooks derived from other images, provided they are "rich" enough. This is the case, for instance, of figure 1, derived from the Van Hateren database. That section also describes the approximating technique used to collect the raw frequency data for the approximations $B \simeq (D^B, \sigma, \mu)$, which are analyzed in section 3. The main results of that section are, first, several entropy and mutual information estimates for the joint (D^B, σ, μ) distribution and the marginals of D^B, σ and μ. These estimates suggest as a first approximation that the marginals may be taken to be independent, that is, that $p(D^B, \sigma, \mu) \simeq p(D^B)p(\sigma)p(\mu)$, decomposition that still leaves about 1.5 bits of mutual information between D^B and (σ, μ) to be explained. To do so we shall refine the previous first order ap-

proximation to a second one of the form $p(D^B, \sigma, \mu) \simeq p(D^B)p(\sigma)p(\mu)\Phi(||\nabla B||)$. The structure of the $p(D^B), p(\sigma)$ and $p(\mu)$ marginals is dealt with in section 3. While $p(\mu)$ does not carry significant information, we shall see that $p(\sigma)$ has an exponential structure and that $p(D^B)$ follows a nearly uniform behavior with respect to volume in image space. A final section contains a summary of the paper and pointers to further work.

2 Methods

As the natural patch source, we shall work with 4300 8 bit gray level images of size 1540×1024 from the Van Hateren database. We shall restrict ourselves to their 1024×1024 squared centers and take out flat blocks, that is, those B with $\sigma \leq 3$ (about 20% of all patches). As mentioned above, we shall approximate a normalized natural patch $\tilde{B} = \frac{B-\mu}{\sigma}$ by another normalized domain D^B taken from a codebook derived from a 256×256 version of the well known Lena image as follows: we will first extract all its 4×4 (overlapping) blocks. This gives $(256 - 4 + 1)^2 \simeq 2^{16}$ codebook domains, that become 2^{20} after adding for each block its 8 isometries and its negative (recall that we are reconstructing B using positive σ_B contrast factors). Again, we will exclude flat domains, about 25% of the initial Lena domains.

Thus, we will not estimate the direct distribution $p(B)$ but instead that of the B approximation $p(D^B, \sigma_B, \mu_B)$. To minimize the distortion that this approximation is bound to introduce, we shall take D^B as the codebook domain for which

$$dist(B, D) = ||B - \sigma_B D - \mu_B||_\infty = \sup |B_{ij} - \sigma_B D_{ij} - \mu_B| \tag{1}$$

verifies $dist(B, D) \leq d_M$, taking in what follows $d_M = 8$. Reconstructing a full image $I = \{B_s\}$ by its patches' approximations $\hat{I} = \{D^B_s\}$, our choice of d_M should ensure that $||I - \hat{I}||_2 \leq 8$ and the PSNR of the reconstruction $\hat{I}$ verifies $PSNR(\hat{I}) = 20 \log_{10}(\frac{255}{||I-\hat{I}||_2}) \geq 20 \log_{10}(\frac{255}{8}) \simeq 30$. We shall discard those B for which a matching domain cannot be found. They are about a 1 per 1000 of all non flat domains, which results in a final number of about $232 \times 10^6 \simeq 2^{27.79}$ disjoint 4×4 patches.

Finding matching domains requires to perform at some point the costly full block comparisons in (1), that can make FIC very time consuming. To speed things up, we shall precede full block comparisons with a hash–like pre–comparison. We define a hash function (see [3] for further details)

$$h(D) = \sum_{h=1}^{H} \left(\left\lfloor \frac{D_{i_h j_h}}{\lambda} \right\rfloor \%C + \frac{C}{2} \right) C^{h-1} = \sum_{h=1}^{H} b_h C^{h-1}. \tag{2}$$

with $D_{i_h j_h}$, $1 \leq h \leq H$, adequately chosen points in D, and H, C and λ appropriately chosen parameters. In what follows we shall take the four corner pixels and an extra middle pixel and thus $H = 5$; as the base C we shall take $C = 16$.

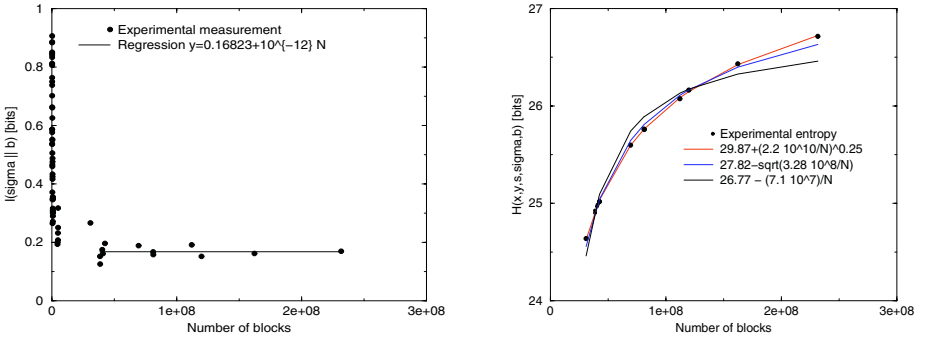

Fig. 2. Large sample behavior of the mutual information $I(\sigma||\mu)$ (left) and of the total entropy $H(i, j, s, \sigma, \mu)$ (right) estimates. While the left picture saturates, this is not the case for the $H(i, j, s, \sigma, \mu)$ estimate.

The choice of λ should ensure that $D_{i_h j_h}$ is distributed more or less uniformly on the interval $[-\lambda\frac{C}{2}, \lambda\frac{C}{2}]$, and, thus, that (2) defines a uniform base C expansion. A good choice for this would be to take λ in the interval 0.5–1, but to streamline our subsequent discussions, we shall take $\lambda = 2$. Once the values $h(D)$ have been computed for all codebook domains, those with the same value are stored in the same linked lists over a hash pointer table. Full block comparisons for a natural block B are performed only over the domains in the linked lists whose h index is contained in the B dependent set $H(B) = \{h_\delta(B)\}$, where

$$h_\delta(B) = \sum_{h=1}^{H} \left(\left\lfloor \frac{B_{i_h j_h} - \mu_B}{\lambda \sigma_B} + \delta_h \right\rfloor \%C + \frac{C}{2} \right) C^{h-1} = \sum_{h=1}^{H} r_h^\delta C^{h-1}, \qquad (3)$$

with the displacement vector $\delta = (\delta_1, \ldots, \delta_H)^t$ verifying $|\delta_h| \leq 1$. It is not difficult to show that this searching procedure will provide the optimal matching domain D^B. Our coding of a block B will then be

$$T(B) = (i, j, s, \sigma, \mu)$$

where (i, j) indicates the position in the Lena image of the left upper corner of the matching domain, and s is an index for the isometry and negative used (notice that the dilations in (1) are positive).

In order to obtain the statistics described in the next section, we shall compute first basic frequency value sets P_i over 9 image batches. All of them contain registers of the form $[i, j, s, \sigma, \mu, c]$, with c counting the number of patches in the file with σ, μ statistics and a matching i, j, s domain; they are sorted in lexicographic order. We then perform a first merge over 9 of these P_i sets to arrive to larger value sets Q_j and then a second merge over 9 of the Q_j sets to arrive to 6 large files S_k of statistical values, 5 of them corresponding to $729 = 9^3$ image batches and a smaller sixth one for the last 650 images. A final merge of these 6 files will give the statistics described in the next section.

Table 1. Different entropy measures (in bits) of image statistics (first column) and limit estimates for them.

Quantity	Value	Limit estimates
N	231511046	—
$\log_2 N$	27.7865	
$H(i,j,s,\sigma,\mu)$	26.7141	29.87
$H(i,j,s)$	17.8156	17.82
$H(\sigma,\mu)$	10.4627	10.46
$I(i,j,s\|\|\sigma,\mu)$	1.5642	0.474
$I(\sigma\|\|\mu)$	0.1698	0.115
$I/H(i,j,s,\sigma,\mu)$	5.86 %	1.70 %
$I(\sigma,\mu)/H(\sigma,\mu)$	1.62 %	1.10 %

3 Entropy and Mutual Information Estimates

It is well known that the accuracy of any discrete probability entropy estimate depends on its sample regime, that is, the relationship between the sample number N and the number M of non–empty bins, i.e., of non–zero probabilities. In our case two very different regimes are to be considered. In the first one, we shall have $N \gg M$; this is the case for the joint (σ,μ) distribution and to a smaller extent for the (i,j,s) distribution. In this regime [6] we should expect for large N that the entropy and mutual information estimates closely approach saturation limit points that we can take as the true entropy and mutual information values. Figure 2, left, depicts this situation for the mutual information $I(\sigma||\mu)$. It has been computed for the full sample size range and shows a extremely fast drop for sample sizes below 10^6 followed by a nearly horizontal behavior afterwards. This can be interpreted as showing the sample information $\hat{I}_N$ estimates to saturate at a limiting value $I = 0.17$ which we can take as the actual value of $I(\sigma||\mu)$. On the other hand, the right picture shows a quite different situation for the joint entropy $H(i,j,s,\sigma,\mu)$. Clearly a saturation point has not been achieved, which shows that we are still far from an $N \gg M$ regime. Notice that while we have $\log N \simeq 28$ for the full database sample, the full sample estimate for $\hat{H}_N(i,j,s,\sigma,\mu)$ is 26.7, which is a lower bound for $\log M$. In other words $\log N/M \leq 1.3$, that is, we are in the $N \simeq M$ regime. A similar situation holds for $I(i,j,s||\sigma,\mu)$.

Table 3 collects these empirical entropy and mutual information estimates and also some estimates of possible limit values derived from functional approximations to the empirical data. In most cases they are close to the empirical estimates, while this is not the case for $H(i,j,s,\sigma,\mu)$ and $I(i,j,s||\sigma,\mu)$. We shall take these values as a starting point for our discussion. It can be seen from this table that the mutual information between the σ and μ is very small, about 0.17 bits, less than 2% of the joint entropy; therefore we can assume that σ and μ are independent. On the other hand, the mutual information $I(i,j,s||\sigma,\mu)$ is about 1.56, less than 10% of the joint entropy. Therefore, this suggests the first order approximation

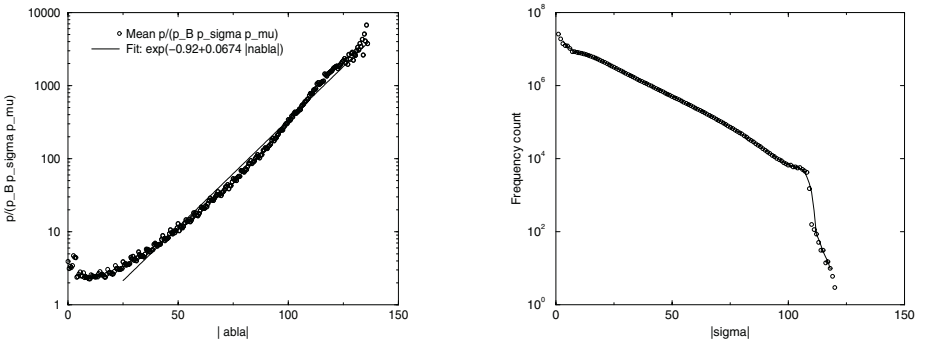

Fig. 3. Left: conditional expectation $\Phi(||\nabla B||)$ in (6) as a function of $||\nabla||$. For lower values of $||\nabla||$, that are the predominant in the distribution, no correction is possible. For larger $||\nabla||$ the approximation results to be approximately exponential. Right: relative frequencies of σ in a logarithmic scale. The linear central behavior suggests an exponential distribution.

$$p(B) = p(\tilde{B}, \sigma_B, \mu_B) \simeq p(i_B, j_B, s_B, \sigma_B, \mu_B) \simeq p(i_B, j_B, s_B)p(\sigma_B)p(\mu_B), \quad (4)$$

which nonetheless leaves more than one mutual information bit to be explained.

To do so, we shall study what we may call the divergence, that is, the average

$$d(i,j) = E_{s,\sigma,\mu}\left[\frac{p(i,j,s,\sigma,\mu)}{p(i,j,s)p(\sigma)p(\mu)},\right]$$

over the joint and (i, j, s) and σ and μ distributions. Values of d different from 1 indicate deviations from the independence assumption and projecting them back over the Lena image suggests that image borders are the main source of the divergence. Thus any correction to (4) should be significantly different from 1 over edge blocks. Other natural assumptions are that it be isotropic, translation invariant and not dependent on the block's absolute brightness, but only on the difference $B - \mu$. Moreover it is natural to look for a lowest order approximation. All this suggests to refine (4) to

$$p(B) \simeq p(D_B)p(\mu_B)p(\sigma_B)\Phi(||\nabla B||). \quad (5)$$

In order to find a reasonable Φ we have compared the conditional average

$$\Phi(||\nabla B||) = E_{||\nabla B||}\left[\log_2 \frac{p(i,j,s,\sigma,\mu)}{(p(i,j,s)p(\sigma)p(\mu)}\right] \quad (6)$$

with $E_{||\nabla B||}$ denoting the conditional expectation with respect to $||\nabla B||$. This is depicted in figure 3, left, which shows that while for the lower $||\nabla B||$ patches, the most predominant ones in natural images, no correction appears to be possible, $\Phi(||\nabla B||)$ can be quite well linearly approximated for the right side high $||\nabla B||$ values. In turn, this suggests that a natural approximation for $\Phi(s)$ is

$$\begin{aligned}\Phi(s) &= e^{-0.923+0.0337s}, \quad s \geq 25\\ &= 0, \quad 0 < s < 25\end{aligned}$$

with the constant 0.923 being a normalization value so that the second order correction is actually a probability. A natural interpretation of the positive constant 0.0337 in Φ is to see it as a high–contrast patch distribution correction, a high information natural image component, as shown for instance in [4]. To check the impact of this second correction, we have re–computed now the mutual information between the experimentally obtained distribution and the above second order corrected distribution, which turns out to be 0.621 bits, that is, about 1 bit less than the previous estimate. Therefore, just 2.3% of the total information is not covered now by the second approximation. In any case, further alternatives for the correction have to be studied. For instance, there are indications that better edge detectors than simple gradient approximations could give better results.

4 Structure of the $p(i, j, s)$, $p(\sigma)$ and $p(\mu)$ Probabilities

The μ distribution depends on the camera's calibration and can be easily manipulated through, say, histogram equalization. It is thus largely irrelevant. The histogram of the σ distribution is depicted in figure 3, right, in vertical logarithmic scale. The figure shows a central linear behavior, suggesting an exponential distribution, with changes at the boundaries. The drop at around 100 is due to the limited range of brightness levels, with a theoretical maximum below 128 for 256 gray levels. Although seemingly hinting at some underlying structure, the cusp–like peak at 0 is mostly due to the layered structure of natural images, that produces many near flat, small deviation blocks. Anyway, the σ distribution makes clear that σ carries significant structural information. One way to visualize this is to project for each domain D with coordinates i, j in the Lena image the corresponding value of $-\log p(\sigma)$. When done, it shows a clear correlation between edges and large $-\log p(\sigma)$ values that makes clearer the significace of the σ component.

It is more complicated to visualize the structure of $p(i, j, s)$. A possibility is to fix our attention in the (i, j) distribution, and to consider the corresponding Lena domains as bins were the sample patches fall. Defining $N(m)$ as the number of such bins getting m sample patches, it can be seen that $\log N(m)$ shows a near parabolic structure, which is still more clear in figure 4, left, where $N(m)$ has been corrected taking into account the volume surrounding each codebook domain. To perform this correction, we may a priori assume that the hash linked lists cover regions with essentially the same volume, and also that all domains of a given list have the same volume. This implies that the a priori probability of a domain D is thus proportional to $\nu(h(D))$, with $\nu(h)$ the number of codebook domains D' such that $h(D') = h$. We then correct the direct counting estimate m' of the number of patches a certain domain D gets to $m = m' \times \nu(h(D))$. The corrected $N(m)$ values are depicted in figure 4, left. As it can be seen, the parabolic fitting is quite good for the higher patch count right part, which suggests that there is a volume uniform distribution of natural patches between codebook domains, that is, that the probability of a codebook region $\mathcal{R}$ receiving a patch is proportional to its volume $V(\mathcal{R})$.

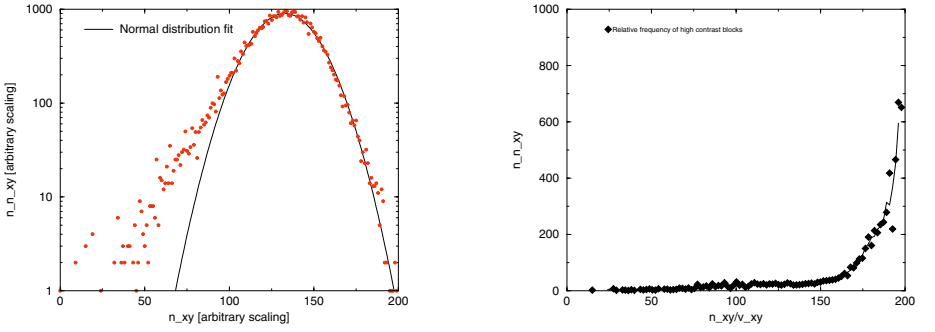

Fig. 4. Volume–corrected (left) values of $\log N(m)$, with $N(m)$ the relative number of domains gettin m natural blocks, that suggest a volume–uniform distribution of all normalized natural patches among codebook domains. This is not true, however, for high contrast patches: the right image shows a markedly higher proportion for them among high count domains.

We have also computed the distribution over m of the patches' average σ values, which shows that the left area corresponds mostly to low σ patches, with small denominators in (3) and, hence, larger hash values, that may cause them to be assigned to wrong matching domains. This would have a little effect in domains with a large block count, as they would lose about as many blocks as they gain, but a bigger one in domains getting fewer blocks. In any case, the natural interpretation of figure 4 is to assume a volume–uniform distribution for the (i, j) domain statistics, as the near gaussian behavior of $N(m)$ is easiest explained as a large sample aproximation of a binomial distribution. Apparently this may contradict recent results in [4], that show a marked structure of high contrast natural 3×3 blocks. However, notice that figure 4, right, depicting the proportion of the high-contrast codebook domains getting a (normalized) number m of patches, has a very sharp rise at the high patch count area. In other words, the high contrast blocks studied in [4] have an statistical behavior of their own, certainly not following the uniform behavior just described.

5 Summary and Future Directions

In this article we have shown how a representation $B \simeq (D_B, \sigma, \mu)$ of natural image 4×4 blocks B, with σ, μ the block's variance an mean and D_B a codebook approximation to the normalization of B, can be used to obtain significant natural image statistics. In fact, we have shown that these blocks' probabilities can be represented as a product of nearly independent factors. The analysis of these factors allows us to conclude that at least in the scale investigated (about one minute of angle), the information is essentially carried by a block's variance, that roughly correlates with the block's edges. This is in accordance with the well known Marr [5] hypothesis that natural image information is extracted from biological systems using its most singular points, e.g. its edges. However, it may

be of some interest to point out that, here, this conclusion is drawn without any regard to the receiving system, but just by using information theoretical considerations; this may suggest that biological systems have adapted themselves to extract the part of a natural image most relevant in terms of information theory. On the other hand, the above results may have applications in, for instance, fast search engines in image databases, or in transmission of moderate quality images over low bit rate, noisy channels. All these matters are currently being researched.

References

1. Y. Fisher (ed.), Fractal Image Compression: Theory and Application, Springer Verlag, New York, 1995.
2. J.H. van Hateren and A. van der Schaaf, *Independent component filters of natural images compared with simple cells in primary visual cortex.* Proc.R.Soc.Lond. B, 265:359-366, 1998.
3. K. Koroutchev and J. Dorronsoro, *Hash–like Fractal Image Compression with Linear Execution Time*, Iberian Conference on Pattern Recognition and Image Analysis, IbPRIA 2003, Lecture Notes in Computer Science.
4. A.B. Lee, K.S. Pedersen and D. Mumford. *The Complex Statistics of High-Contrast Patches in Natural Images.* In WWW Proceedings of Second International IEEE Workshop on Statistical and Computational Theories of Vision. Vancouver, Canada, July 2001.
5. D. Marr, *Vision,* W.H. Freeman and Co., 1982.
6. L. Paninski, *Estimation of Entropy and Mutual Information*, Neural Computation, 15 (2003) 1191–1253.
7. D.L. Ruderman, *The statistics of natural images*, Network: Computation in Neural Systems , vol. 5, 517-548, 1994.
8. A. Srivastava, A.B. Lee, E.P. Simoncelli and S.C. Zhu, *On Advances in Statistical Modeling of Natural Images*, Journal of Mathematical Imaging and Vision 18(1), 17-33, 2003.
9. A Turiel, N Parga, D Ruderman and T Cronin, *Multiscaling and information content of natural color images,* Phys. Rev. E 62 , 1138-1148 (2000)
10. M. Weinberger, G. Seroussi and G. Sapiro, *LOCO-I: A Low Complexity, Context-Based, Lossless Image Compression Algorithm,* Proc. of the IEEE Data Compression Conference, Snowbird, Utah, March-April 1996.