

Presented in: Proc. 8th Annual Workshop on Very Low Bitrate Video Coding, Visual Content Processing and Representation, Eds. N Garcia, JM Martinez and L Salgado. Madrid Spain, 18-19 Sept 2003. DOI 10.1007/b13938
Published as: Lecture Notes in Computer Science, vol 2849, pp 331-340, Springer, 2003.

Directly Invertible Nonlinear Divisive Normalization Pyramid for Image Representation

Roberto Valerio¹, Eero P. Simoncelli², and Rafael Navarro¹

¹ Instituto de Óptica “Daza de Valdés” - CSIC, Madrid, Spain, 28006
{r.valerio, r.navarro}@io.cfmac.csic.es

² Howard Hughes Medical Institute, Center for Neural Science, and Courant Institute for Mathematical Sciences - New York University,
New York, USA, NY 10003
eero.simoncelli@nyu.edu

Abstract. We present a multiscale nonlinear image representation that permits an efficient coding of natural images. The input image is first decomposed into a set of subbands at multiple scales and orientations using near-orthogonal symmetric quadrature mirror filters. This is followed by a nonlinear “divisive normalization” stage, in which each linear coefficient is divided by a value computed from a small set of neighboring coefficients in space, orientation and scale. This neighborhood is chosen to allow this nonlinear operation to be efficiently inverted. The parameters of the normalization operation are optimized in order to maximize the independence of the normalized responses for natural images. We demonstrate the near-independence of these nonlinear responses, and suggest a number of applications for which this representation should be well suited.

1 Introduction

The choice of an appropriate image representation is often driven by the goal of removing statistical redundancy in the input signal. The general problem is extremely difficult, and thus one typically must restrict it by constraining the form of the decomposition and/or by simplifying the description of the input statistics. A classical solution is to consider only linear decompositions, and the second-order (i.e. covariance) properties of the input signal. This technique, known as Principal Components Analysis (PCA), has several drawbacks. First, the solution is not unique if one does not impose additional constraints. Moreover, although PCA can be used to recover a set of statistically independent axes for representing Gaussian data, the technique often fails when the data are non-Gaussian (as is the case of natural images [1]). More recently, a number of authors have shown that one may use higher-order statistics to uniquely constrain the choice of linear decomposition. These procedures are commonly known as Independent Components Analysis (ICA). The resulting basis functions of such decompositions are similar to cortical receptive fields [2, 3] and the associated coefficients are generally more independent than principal components.

Nevertheless, linear decompositions cannot completely eliminate higher-order statistical dependencies [e.g. 4, 5], basically due to the fact that natural images are not

formed as sums of independent components. Large-magnitude coefficients tend to lie along ridges with orientation matching that of the subband, and also tend to occur at the same spatial locations in subbands at adjacent scales and orientations [5]. Large number of recent “context-based” algorithms in image processing take advantage of this, often implicitly.

In recent years, Simoncelli and co-workers [6, 7, 8] have shown that a nonlinear divisive normalization can significantly reduce statistical dependencies between adjacent responses. In this nonlinear stage, the linear inputs are squared and then divided by a weighted sum of squared neighboring responses in space, orientation and scale, plus a regularizing constant. Divisive normalization not only reduces dependency, but also can be used to describe the nonlinear response properties of neurons in visual cortex [9, 10] and yields image descriptors more relevant from a perceptual point of view [11]. However, using of the divisive normalization in image processing applications is not straightforward since it is not easily invertible [12].

A number of authors have proposed nonlinear extensions of multiscale decompositions for use in image processing. For example, nonlinear pyramid schemes can be obtained by replacing linear filters in their linear counterparts by median, morphological or rank order based filters. There are also many nonlinear decompositions based on nonredundant (critically sampled) linear decompositions, such as morphological subband [e.g. 13, 14, 15], order statistics based subband [e.g. 16, 17, 18], and morphological wavelet [e.g. 19, 20, 21] decompositions.

In this paper, we describe a simple nonlinear multiresolution image representation. Starting with a nonredundant linear decomposition, we normalize each coefficient by a value computed from a neighborhood that is suboptimal for dependency reduction, but that allows the transform to be easily inverted. We describe the empirical optimization of the transform parameters, and demonstrate that the redundancy in the resulting coefficients is substantially less than that of the original linear ones.

2 Image Representation Scheme

The scheme proposed here consists of a linear decomposition followed by a nonlinear divisive normalization stage.

2.1 Linear stage

The linear stage is an approximately orthogonal three-level linear decomposition based on symmetric quadrature mirror filters (QMF) with 9 coefficients [22], which are closely related to wavelets (essentially, they are approximate wavelet filters). The basis functions of this linear transform are localized in space, orientation and spatial frequency. This gives rise to 9 subbands (horizontal, vertical and diagonal for each of the 3 scales considered here) plus an additional low-pass channel. Multiscale linear transforms like this are very popular for image representation.

The left panel in Fig. 1 shows a typical conditional histogram of a natural image.

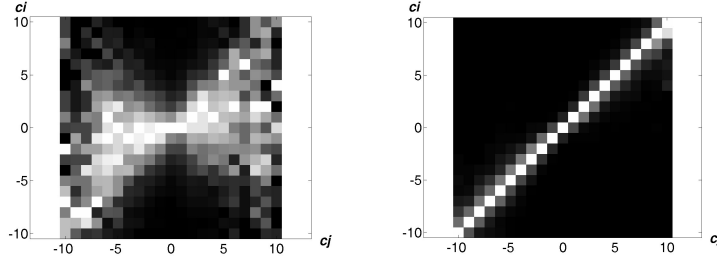


Fig. 1. Conditional histograms of two neighboring coefficients (one is the vertical neighbor of the other) in the lowest (finest) scale vertical subband of the QMF (left), and Gabor (right) pyramid of the “Einstein” standard test image.

As we can see, the QMF coefficients are decorrelated, since the expected value of the ordinate is approximately zero independently of the abscissa and therefore the covariance is close to zero as well. This makes an important difference between orthogonal and non-orthogonal linear transforms, since in the non-orthogonal cases “close” coefficients are correlated and the expected value of the ordinate is not zero but varies linearly with the abscissa. This is what we can see in the right panel in Fig. 1, which corresponds to a non-orthogonal Gabor pyramid [23].

On the other hand, however, the “bowtie” shape of the left histogram reveals that coefficients are not statistically independent. This shape suggests that the variance of a coefficient depends on the value of the neighboring coefficient. The dependence is such that the variance of the ordinate scales with the squared value of the abscissa.

All pairs of coefficients taken either in space, frequency or orientation always show this type of dependence [4, 5], while the strength varies depending on the specific pair chosen. Intuitively, dependence will be stronger for coefficients that are closer, while it will decrease with distance along any axis. The form of the histograms is robust across a wide range of images, and different pairs of coefficients. In addition, this is a property of natural images but is not of the particular basis functions chosen.

Several distributions have been proposed to describe the conditional statistics of the coefficients obtained by projecting natural images onto an orthogonal linear basis [e.g. 7, 8, 24]. Here we will use the Gaussian model of [7].

Assuming a Gaussian model, the conditional probability $p(c_i | \{c_j^2\})$ of an orthogonal linear coefficient c_i of a natural image, given the other squared coefficients $\{c_j^2\}$ ($j \neq i$), can be modeled as a zero-mean Gaussian density with a variance $(a_i^2 + \sum_{j \neq i} b_j c_j^2)$ that depends on a linear combination of the squared coefficients $\{c_j^2\}$ ($j \neq i$) plus a constant a_i^2 :

$$p(c_i | \{c_j^2\}) = \frac{1}{\sqrt{2\pi(a_i^2 + \sum_{j \neq i} b_j c_j^2)}} \exp\left\{-\frac{c_i^2}{2(a_i^2 + \sum_{j \neq i} b_j c_j^2)}\right\} \quad (1)$$

In the model, a_i^2 and $\{b_{ij}\}$ ($i \neq j$) are free parameters and can be determined by maximum-likelihood (ML) estimation. Operating with Eq. 1 we obtain the following ML equation:

$$\{a_i^2, b_{ij}\} = \arg \min_{\{a_i^2, b_{ij}\}} \mathbb{E} \left\{ \frac{c_i^2}{a_i^2 + \sum_{j \neq i} b_{ij} c_j^2} + \log(a_i^2 + \sum_{j \neq i} b_{ij} c_j^2) \right\} \quad (2)$$

where \mathbb{E} denotes expected value. In practice, we can compute \mathbb{E} for each subband, averaging over all spatial positions of a set of natural images.

2.2 Nonlinear stage

The nonlinear stage consists basically of a divisive normalization, in which the responses of the previous linear filtering stage, c_i , are divided by the square root of a weighted sum of squared neighboring responses in space, orientation and scale, $\{c_j^2\}$, plus a constant, d_i^2 :

$$r_i = \frac{c_i}{\sqrt{d_i^2 + \sum_j e_{ij} c_j^2}} \quad (3)$$

Eq. 3 is slightly different than models of cortical neuron responses (in these models the second term of the equality is typically squared) but has the advantage that preserves sign information. We will refer as optimal divisive normalization to the one defined by the values of the parameters (constant d_i^2 and weights $\{e_{ij}\}$) that yields the minimum mutual information, or equivalently minimizes statistical dependence, between normalized responses for a set of natural images. It can be shown that an approximate solution is [7]: $d_i^2 = a_i^2$, $e_{ij} = b_{ij}$ ($i \neq j$) and $e_{ii} = 0$, that is to adopt directly the parameters of the Gaussian model, a_i^2 and b_{ij} ($i \neq j$), as the normalization parameters.

A key feature of the nonlinear stage is the particular neighborhood considered in Eq. 3. As shown in Fig. 2, we have considered 12 coefficients $\{c_j\}$ ($j \neq i$) adjacent to c_i along the four dimensions (9 in a square box in the 2D space, plus 2 neighbors in orientation and 1 in frequency). It is important to note that in this particular choice all neighbors belong to higher levels of the linear pyramid, which permits to invert the nonlinear transform very easily level by level (to recover one level of the linear pyramid we obtain the normalizing values from levels already recovered and multiply them by the corresponding nonlinear coefficients). Obviously, in order to invert the nonlinear transform we need to store the low-pass residue of the linear decomposition.

Therefore, in addition to the nice feature of giving almost statistically independent coefficients, the described scheme is easily invertible. Since both stages, linear and nonlinear, of the image representation scheme are invertible, it is possible to recover the input image from its nonlinear decomposition.

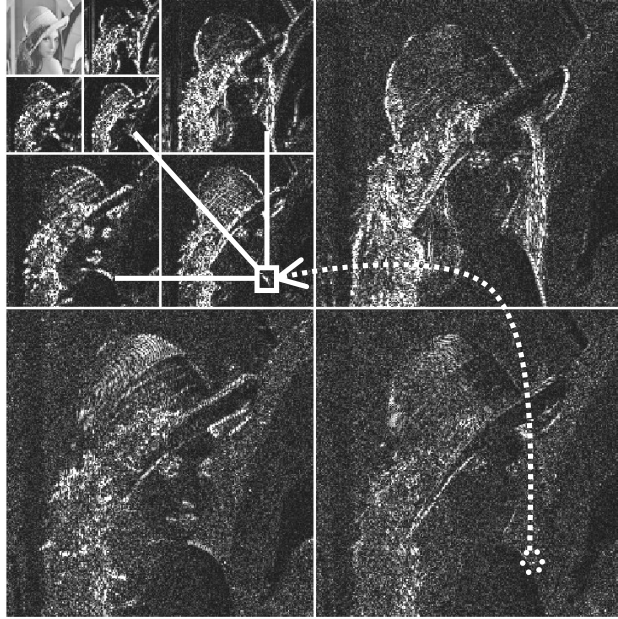


Fig. 2. QMF decomposition of the “Lena” image and neighborhood considered in the nonlinear stage of our image representation scheme.

3 Results

The following results have been obtained using a “training set” of six B&W natural images with 512x512 pixel format (“Boats”, “Elaine”, “Goldhill”, “Lena”, “Peppers” and “Sailboat”).

First, to model the conditional statistics of the QMF coefficients of the images, we used the Gaussian model in Eq. 1. The free model parameters, a_i^2 and $\{b_{ij}\} (i \neq j)$, were obtained using the mathematical expectation in Eq. 2 over all the QMF coefficients of the 6 images in the “training set”, but independently for each subband of the QMF pyramid. Both in the Gaussian model and in the divisive normalization, we considered the 12-coefficient neighborhood, $\{c_j\} (j \neq i)$, of adjacent coefficients to c_i along the four dimensions described in the previous section (see Fig. 2). A linear search method was used to solve the corresponding minimization problems, using the additional constraint of positivity of the free model parameters to improve convergence. As an example, Fig. 3 shows the values of the Gaussian model parameters for the lowest scale vertical subband.

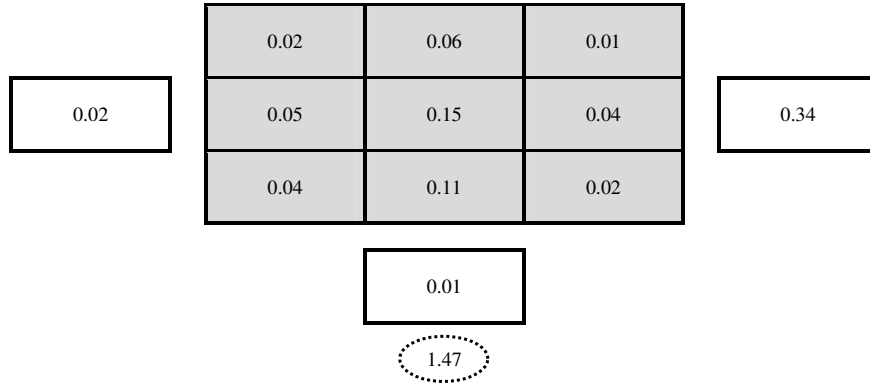


Fig. 3. Parameter values of the Gaussian statistical model for the lowest scale vertical subband. The shadowed values correspond to the 9 spatial parameters. The 2 orientation parameters are horizontally arranged, and vertically the scale parameter. The bottom row contains the value of d_i^2 .

On the other hand, in order to efficiently attain an optimal divisive normalization that minimizes statistical dependence between output responses for natural images, we fixed the divisive normalization parameters to the following values [7]: $d_i^2 = a_i^2$, $e_{ij} = b_{ij}$ ($i \neq j$) and $e_{ii} = 0$, where a_i^2 and b_{ij} ($i \neq j$) are the parameters of the Gaussian model.

When we apply the model described above, we obtain representations similar to that of Fig. 4. Intuitively, the nonlinear transform has the effect of randomizing the image representation in order to reduce statistical dependencies between coefficients belonging to the same structural feature, or in other words, the effect of the divisive normalization is to choose which coefficients are most effective for describing a given image structure.

Fig. 5 represents two conditional histograms of two adjacent samples in space and illustrates the statistical independence achieved by application of the nonlinear (divisive normalization) transform. The left panel shows the conditional histogram of two QMF coefficients c_i and c_j . (c_j is the right down neighbor of c_i). This linear transform does not remove higher-order statistical dependencies, as suggested by the “bowtie” shape of the histogram. The right panel in Fig. 5 shows the conditional histogram between the two corresponding output responses (r_i and r_j). As we can see, after normalization, output statistical dependencies are substantially reduced since the resulting conditional histogram is basically independent on the value of the abscissa.

Table 1 shows some numerical measures of statistical dependence in terms of mutual information for the 6 images in the “training set” (“Boats”, “Elaine”, “Goldhill”, “Lena”, “Peppers”, and “Sailboat”). Mutual information was calculated from 200 bin joint histograms in the interval (-100, 100) of the corresponding random variables after fixing their standard deviation to 5 and equalizing their histogram in order to compare the results (note that to apply a monotonic nonlinearity to one or the two variables does not modify their mutual information).

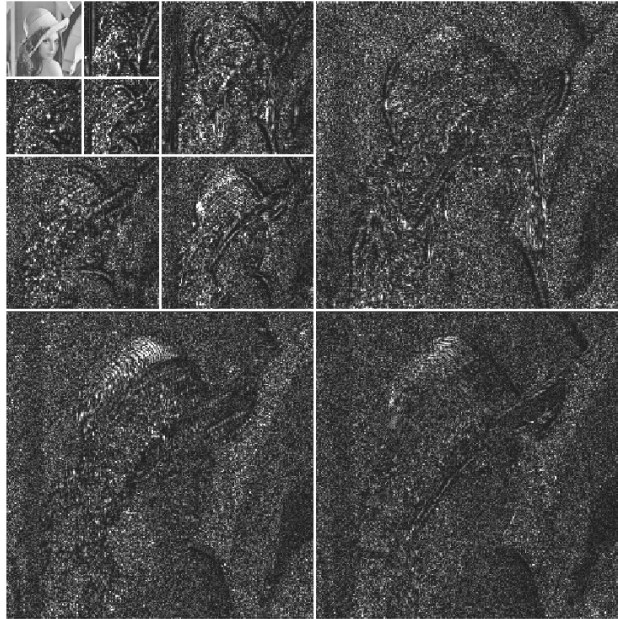


Fig. 4. Nonlinear divisive normalization decomposition of the “Lena” image.

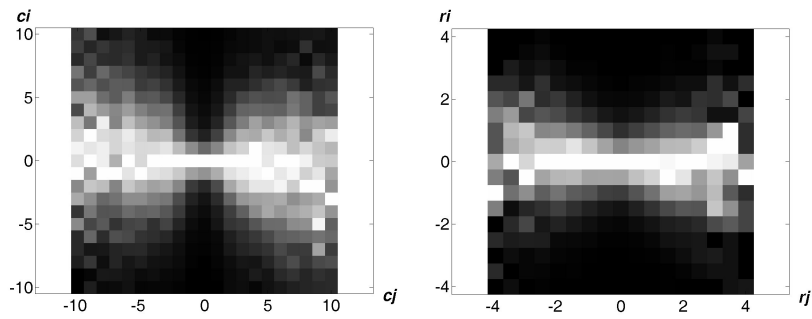


Fig. 5. Conditional histograms of two neighboring QMF coefficients c_i and c_j (c_j is the right down neighbor of c_i), and nonlinear responses r_i and r_j , of the “Sailboat” image. The considered subband is the lowest scale vertical one.

Consistently with Fig. 5, we can see in Table 1 that divisive normalization decreases mutual information, MI, with all values much closer to zero. If we compare our scheme (column A) with a scheme that uses a more general neighborhood (column B), we observe that the former yields results only slightly worse (the resulting nonlinear coefficients are a little more statistically dependent). Nevertheless, this is a very small price to pay for easy invertibility.

Table 1. Mutual information between two neighboring QMF coefficients c_i and c_j (c_j is the right down neighbor of c_i), and between the corresponding normalized coefficients r_i and r_j , of the 6 images in the “training set”. The considered subband is always the lowest scale vertical one. Column A corresponds to our scheme described above and column B corresponds to a scheme that uses a more general neighborhood (a 12-coefficient neighborhood of adjacent coefficients along the four dimensions: 8 in a square box in the 2D space, plus 2 adjacent neighbors in orientation and 2 in frequency) with coefficients belonging not only to higher levels of the QMF pyramid but also to the same and lower levels.

	$MI(c_i, c_j)$	$MI(r_i, r_j)$ (A)	$MI(r_i, r_j)$ (B)
“Boats”	0.18	0.05	0.03
“Elaine”	0.05	0.03	0.02
“Goldhill”	0.10	0.04	0.03
“Lena”	0.12	0.03	0.03
“Peppers”	0.09	0.04	0.03
“Sailboat”	0.12	0.04	0.03

4 Conclusions

We have presented a multiscale multiorientation nonlinear image representation scheme. The key feature of this new image representation scheme is that the resulting coefficients are almost statistically independent, much more than those of the orthogonal linear transforms that cannot eliminate higher order dependencies. Such representations are also roughly consistent with the nonlinear properties of neurons in the primary visual cortex of primates, and have been shown relevant to human perception.

One of the main contributions is the particular neighborhood used in the nonlinear stage. Basically, we impose the restriction that the neighboring coefficients considered only belong to higher levels of the linear pyramid. This restriction permits to invert the nonlinear transform very easily and has only little impact on the statistical independence of the resulting nonlinear coefficients.

The scheme is robust in the sense that results do not depend critically on the linear decomposition, the model of the conditional statistics of the linear coefficients, the neighborhood considered, the “training set” of natural images, the computing errors (for example in estimating the parameters), or even the particular input natural images.

Finally, this nonlinear scheme of image representation, which has better statistical properties than the popular orthogonal wavelet transforms, is potentially useful for many image analysis and processing applications, such as restoration, synthesis, fusion, coding and compression, registration, etc., because of its easy invertibility and the great importance of statistical independence in these applications. Similar schemes [25, 26] have already been used very successfully in image analysis and processing applications. In addition, due to its good perceptual properties, our scheme could be useful to define a metric for perceptual image distortion similarly to [26, 27, 28].

Acknowledgments

This research was supported by the Spanish Commission for Research and Technology (CICYT) under grant DPI2002-04370-C02-02. RV was supported by a Madrid Education Council and Social European Fund Scholarship for Training of Research Personnel, and by a City Hall of Madrid Scholarship for Researchers and Artists in the Residencia de Estudiantes.

References

1. D. J. Field, “Relations between the statistics of natural images and the response properties of cortical cells”, *J. Opt. Soc. Am. A*, **4**(12), pp. 2379-2394, 1987.
2. B. A. Olshausen and D. J. Field, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”, *Nature*, **381**, pp. 607-609, 1996.
3. A. J. Bell and T. J. Sejnowski, “The independent components of natural scenes are edge filters”, *Vision Research*, **37**(23), pp. 3327-3338, 1997.
4. B. Wegmann and C. Zetsche, “Statistical dependence between orientation filter outputs used in an human vision based image code”, *Proc. SPIE Vis. Commun. Image Processing*, **1360**, pp. 909-922, 1990.
5. E. P. Simoncelli, “Statistical models for images: compression, restoration and synthesis”, *Asilomar Conf. Signals, Systems, Comput.*, pp. 673-679, 1997.
6. E. P. Simoncelli and O. Schwartz, “Modeling surround suppression in V1 neurons with a statistically-derived normalization model”, *Advances in Neural Information Processing Systems*, **11**, pp. 153-159, 1999.
7. O. Schwartz and E. P. Simoncelli “Natural signal statistics and sensory gain control”, *Nature neuroscience*, **4**(8), pp. 819-825, 2001.
8. M. J. Wainwright, O. Schwartz, and E. P. Simoncelli, “Natural image statistics and divisive normalization: modeling nonlinearities and adaptation in cortical neurons”, *Statistical Theo-*

- ries of the Brain, chapter 10, pp. 203-222, eds. R. Rao, B. Olshausen, and M. Lewicki, MIT Press, Cambridge, MA, USA, 2002.
9. D. G. Albrecht and W. S. Geisler, "Motion sensitivity and the contrast-response function of simple cells in the visual cortex", *Visual Neuroscience*, **7**, pp. 531-546, 1991.
 10. D. J. Heeger, "Normalization of cell responses in cat striate cortex", *Visual Neuroscience*, **9**, pp. 181-198, 1992.
 11. J. M. Foley, "Human luminance pattern mechanisms: masking experiments require a new model", *Journal of the Optical Society of America A*, **11**, pp. 1710-1719, 1994.
 12. J. Malo, E. P. Simoncelli, I. Epifanio, and R. Navarro, "Nonlinear image representation for efficient coding", to be submitted, 2003.
 13. O. Egger, W. Li, and M. Kunt, "High compression image coding using an adaptive morphological subband decomposition", *Proceedings of the IEEE*, **83**, pp. 272-287, 1995.
 14. F. J. Hampson and J. C. Pesquet, "M-band nonlinear subband decompositions with perfect reconstruction", *IEEE Transactions on Image Processing*, **7**, pp. 1547-1560, 1998.
 15. R. L. de Queiroz, D. A. F. Florencio, and R. W. Schafer, "Nonexpansive pyramid for image coding using a nonlinear filterbank", *IEEE Transactions on Image Processing*, **7**, pp. 246-252, 1998.
 16. P. Salembier and M. Kunt, "Size-sensitive multiresolution decompositions of images with rank order based filters", *Signal Processing*, **27**, pp. 205-241, 1992.
 17. J. A. Bangham, T. G. Campbell, and R. V. Aldridge, "Multiscale median and morphological filters for 2D pattern recognition", *Signal Processing*, **38**, pp. 387-415, 1994.
 18. G. R. Arce and M. Tian, "Order statistic filter banks", *IEEE Transactions on Image Processing*, **5**, pp. 827-837, 1996.
 19. O. Egger and W. Li, "Very low bit rate image coding using morphological operators and adaptive decompositions", *Proceedings of the IEEE International Conference on Image Processing*, pp. 326-330, 1994.
 20. O. Egger, P. Fleury, T. Ebrahimi, and M. Kunt, "High-performance compression of visual information - A tutorial review: I. Still pictures", *Proceedings of the IEEE*, **87**, pp. 976-1013, 1999.
 21. J. Goutsias and A. M. Heijmans, "Nonlinear multiresolution signal decomposition schemes - Part II: Morphological wavelets", *IEEE Transactions on Image Processing*, **9**, pp. 1862-1913, 2000.
 22. E. P. Simoncelli and E. H. Adelson, "Subband image coding", *Subband Transforms*, chapter 4, pp. 143-192, ed. John W Woods, Kluwer Academic Publishers, Norwell, MA, USA, 1990.
 23. O. Nestares, R. Navarro, J. Portilla, and A. Taberner, "Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions", *Journal of Electronic Imaging*, **7**(1), pp. 166-173, 1998.
 24. M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images", *Advances in Neural Information Processing Systems*, **12**, pp. 855-861, 2000.
 25. R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain", *IEEE Transactions on Image Processing*, **8**(12), pp. 1688-1701, 1999.
 26. J. Malo, F. Ferri, R. Navarro, and R. Valerio, "Perceptually and statistically decorrelated features for image representation: application to transform coding", *Proceedings of the 15th International Conference on Pattern Recognition*, **3**, pp. 242-245, 2000.
 27. P. Teo and D. Heeger, "Perceptual image distortion", *Proceedings of the IEEE International Conference on Image Processing*, **2**, pp. 982-986, 1994.
 28. A. B. Watson and J. A. Solomon, "Model of visual contrast gain control and pattern masking", *Journal of the Optical Society of America A*, **14**(9), pp. 2379-2391, 1997.