HAL
open science

ICA based algorithms for computing optimal 1-D linear block transforms in variable high-rate source coding

Michel Narozny, Michel Barret, Dinh-Tuan Pham

To cite this version:

Michel Narozny, Michel Barret, Dinh-Tuan Pham. ICA based algorithms for computing optimal 1-D linear block transforms in variable high-rate source coding. Signal Processing, 2008, 88 (2), pp.268-283. 10.1016/j.sigpro.2007.07.017 . hal-00278351

HAL Id: hal-00278351

https://centralesupelec.hal.science/hal-00278351

Submitted on 12 May 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ICA based algorithms for computing optimal 1-D linear block transforms in variable high-rate source coding

Michel Narozny[†*], *Michel Barret*[†] and *Dinh-Tuan Pham*[‡]

[†] SUPELEC

Information Multimodality & Signal Team

2 rue É. Belin 57070 Metz (France)

e-mail: FirstName.Name@supelec.fr

[‡] Jean Kuntzmann Laboratory

51 rue des Mathématiques, BP 53, 38041 Grenoble Cedex 9 (France)

e-mail: Dinh-Tuan.Pham@imag.fr

This work was supported in part by the French Ministry of Higher Education and Research,

with the "Action Concertée Incitative Masse de Données" ACI^{2}M project,

*and * by the Lorraine Region.*

5th June 2007

Revised version.

Abstract

The Karhunen-Loève Transform (KLT) is optimal for transform coding of Gaussian sources, however it is no more optimal, in general, for non Gaussian sources. Furthermore, under the high resolution quantization hypothesis, nearly everything is known about the performance of a transform coding with entropy constrained scalar quantization and mean square distortion. It is then straightforward to find a criterion that, when minimized, gives the optimal linear transform under the above mentioned conditions. However, the optimal transform computation is generally considered as a difficult task and the Gaussian assumption is then used in order to simplify the calculus. In this paper, we present the above mentioned criterion as a contrast of Independent Component Analysis modified by an additional term which is a penalty to nonorthogonality. Then we adapt the `icainf` algorithm by Pham in order to compute the transform minimizing the criterion either with no constraint or with the orthogonality constraint. Finally, experimental results show that the transforms we introduced can 1) outperform the KLT on synthetic signals, 2) achieve slightly better PSNR for high-rates and better visual quality (preservation of lines and contours) for medium-to-low rates than the KLT and 2-D DCT on grayscale natural images.

1 Introduction

Transform coding has been extensively developed for coding natural signals like sounds, images (e.g., JPEG [1]) and video. The motivating principle of transform coding is that *simple* coding may be more effective in the transform domain than in the original signal space. Generally "simple coding" corresponds to the use of scalar quantization and scalar entropy coding, because they provide a good trade-off between computational complexity and performance [2].

In a conventional transform coder [3, 4], an input vector $\mathbf{X}$ is first transformed into another vector $\mathbf{Y} = \mathbf{TX}$ of the same dimension. The components of that vector are then described to the decoder using independent scalar quantizers on the coefficients. Finally, the decoder reconstructs the quantized transform vector $\widehat{\mathbf{Y}}$ and then uses a linear transformation $\mathbf{U}$ to get an estimate of the original input vector $\widehat{\mathbf{X}} = \mathbf{U}\widehat{\mathbf{Y}}$. In this paper, we consider mean-square distortion : $\mathrm{E}[\|\mathbf{X} - \widehat{\mathbf{X}}\|^2]$, where E denotes

mathematical expectation. When the input vector is Gaussian, the optimal transforms $\mathbf{T}$ and $\mathbf{U}$ satisfy both the conditions $\mathbf{T}$ is an orthogonal matrix (i.e., $\mathbf{T}^{-1} = \mathbf{T}^T$, where ^{T} denotes transposition) that produces uncorrelated transform coefficients and $\mathbf{U} = \mathbf{T}^{-1}$. Such a transform $\mathbf{T}$ is often called Karhunen-Loève (KLT) in coding lessons.

The optimality of the KLT and its inverse is well known for high rates [3], or when optimal fixed-rate quantizers are employed [5], or more generally when the quantizers are scale-invariant [2] for both the fixed-rate and the variable-rate coding models. Moreover, for non Gaussian sources, it is well known that the KLT can be suboptimal in transform coding [6]. Research in transform coding theory, including subband coding, has been constant for several decades. There is a great interest in extending theoretical results on transform optimality towards low bit-rates [7] [2], or in extending the KLT optimality in more general situations by weakening the assumptions that ensure the optimality. In [8], it is shown that under the classical assumptions of high-rate quantization, mean-square distortion and variable-rate coding, the KLT remains optimal for a more general class of signals than Gaussian ones. In [9], the quantization effect on backward adaptive transform coding of Gaussian sources is studied; the perturbation due to quantization on coding gain is computed up to second order for KLT and causal unitary transform.

Furthermore, under the high resolution quantization hypothesis, nearly everything is known about the performance of a transform coding with entropy constrained scalar quantization and mean square distortion. It is then straightforward to find a criterion that, when minimized, gives the optimal linear transform under the above mentioned conditions. Nevertheless, the optimal transform computation is generally considered as a difficult task [7] and the Gaussian assumption is then used in order to simplify the calculus. In this paper we resolve the problem of computing the optimal transform under high-rate entropy constraint scalar quantization hypothesis.

In 1963, Huang and Schultheiss [5] showed that, for scalar Lloyd-Max quantizers, if the vector source $\mathbf{X}$ is Gaussian then the mean square distortion is minimized by choosing $\mathbf{U} = \mathbf{T}^{-1}$ when $\mathbf{T}$ is a

KLT of $\mathbf{X}$. Actually, in their proof, they used in the one hand the fact that, applied to a real random variable Y, a Lloyd-Max scalar quantizer with cells $S_1, \ldots, S_k$ and outputs $\widehat{y}_1, \ldots, \widehat{y}_k$ satisfies the *centroid condition* $\widehat{y}_i = \mathrm{E}[Y \,|\, Y \in S_i]$ for $(1 \leq i \leq k)$; and in the other hand the following property satisfied by Gaussian sources when $\mathbf{T}$ is a KLT: the jth component of the quantization noise $\mathbf{Y} - \widehat{\mathbf{Y}}$ and the ith component of $\widehat{\mathbf{Y}}$ are uncorrelated for $i \neq j$. When the Gaussian assumption does not hold, it is clear that the previous property is satisfied if the components of $\mathbf{Y}$ are statistically independent (indeed $Y_j - \widehat{Y}_j$ and $\widehat{Y}_i$ are deterministic functions of Y_j and Y_i respectively). Hence their proof remains valid under this condition. Moreover, the result by Huang and Schulteiss still holds when the Lloyd-Max scalar quantizers are replaced by uniform scalar quantizers, under the additional assumption of high-rate quantization [3] (since they satisfy the centroid condition). In this paper, we suppose the transform used by the decoder satisfies $\mathbf{U} = \mathbf{T}^{-1}$. This assumption is partially justified by the fact that the linear transforms $\mathbf{T}$ we consider in the following give transformed vectors $\mathbf{Y}$ with minimal mutual information between components, and hence with components generally close to independence. However, the total independence is rarely achieved and it will be interesting to complete our study by the general case of arbitrary invertible matrices $\mathbf{T}$ and $\mathbf{U}$.

In this paper, we first show that under the high-rate entropy constraint scalar quantization hypothesis, for mean-square distortion and the condition $\mathbf{U} = \mathbf{T}^{-1}$, the criterion that gives — when minimized — the optimal linear transform $\mathbf{T}$ can be expressed as a classical contrast of Independent Component Analysis (ICA) (actually the opposite of such a contrast), modified by an additional term which can be explained as a pseudo distance to orthogonality. This new presentation of a classical result gives an interesting point of view covering ICA and data compression. Indeed, it is well known that images of natural scenes are well modeled with ICA [10], this underlies the potential usefulness of ICA to compression [11, 12, 13]. Then, we present two variants of the `icainf` algorithm by Pham [14] that permit to compute the optimal transform $\mathbf{T}$ with 1) no constraint and 2) the orthogonality constraint. Finally, experimental results on both synthetic data and natural images are given in Section 4, they

show that the transforms returned by the new algorithms can 1) outperform the KLT when used with synthetic signals, and 2) achieve slightly better performance for high-rates and better visual quality (preservation of lines and contours) than the KLT and 2D DCT when applied to the medium-to-low bit rate compression of natural images. This last result is unexpected, since no optimality is ensured at low bit-rates. The paper begins with a brief review of high bit-rate transform coding. The results presented here have been partially published in [15] and [16].

2 High bit-rate transform coding

The general structure of a transform coding scheme is shown in Fig. 1. The class of signals to be encoded is represented by a random vector $\mathbf{X} = [X_1, \ldots, X_N]^T$ of size N. Although these signals may be multidimensional like images, they are indexed by an integer to simplify notations: $\mathbf{X}(m)$. The components of $\mathbf{X}(m)$ are successive samples from a source signal $(x(n))_n$. A conventional transform coder applies a linear invertible transform $\mathbf{T}\colon\ \mathbb{R}^N \to \mathbb{R}^N$ to $\mathbf{X}$ in order to obtain a random vector $\mathbf{Y} = [Y_1, \ldots, Y_N]^T$ better suited to coding than $\mathbf{X}$. To construct a finite code, each coefficient Y_i is first approximated by a quantized value $\widehat{Y}_i$. We concentrate only on scalar quantizers. The quantized coefficients are then entropy coded. The coded representation is stored or communicated over an error-corrected (lossless) channel. The receiver (decoder) provides an approximation $\widehat{\mathbf{X}} = [\widehat{X}_1, \ldots, \widehat{X}_N]^T$ of the original signal $\mathbf{X}$ by applying a linear transform $\mathbf{U}$ to the quantized signal $\widehat{\mathbf{Y}}$. In this paper we assume $\mathbf{U} = \mathbf{T}^{-1}$.

A relevant problem for a system designer is to minimize the reconstruction error under bit rate constraint. Here, the optimization criterion is the mean-square distortion $D = \frac{1}{N}\,\mathrm{E}[\|\mathbf{X} - \widehat{\mathbf{X}}\|^2]$ which satisfies

$$D = \frac{1}{N}\sum_{i=1}^{N} w_i E\left[(Y_i - \widehat{Y}_i)^2\right] + \frac{1}{N}\sum_{i=2}^{N}\sum_{j=1}^{i-1} w_{ij} E\left[(Y_i - \widehat{Y}_i)(Y_j - \widehat{Y}_j)\right], \tag{1}$$

where $[\mathbf{A}]_{i,j}$ denotes the element localized on the i-th row and the j-th column of $\mathbf{A}$, $w_i = \sum_{j=1}^{N} [\mathbf{T}^{-1}]_{j,i}^2$

$(1 \leq i \leq N)$ and $w_{ij} = \sum_{k=1}^{N} 2[\mathbf{T}^{-1}]_{k,i}[\mathbf{T}^{-1}]_{k,j}$ $(1 \leq j < i \leq N)$. The last term in (1) vanishes when 1) $w_{ij} = 0$ for $i \neq j$, i.e., the $\mathbf{T}$ transform is orthogonal — or more generally when the column vectors of $\mathbf{T}^{-1}$ are pairwise orthogonal — or 2) when the quantization noises of different quantizers are uncorrelated and centered. The last condition 2) is satisfied when i) the vector $\mathbf{X}$ is Gaussian and the transform $\mathbf{T}$ is a KLT[1] or when ii) the transform coefficients are statistically independent — the Gaussian assumption is then useless — and the quantizers satisfy the centroid condition. Moreover, experimental tests show that condition 2) approximatively holds under high-rate quantization hypothesis for any kind of signal.

In the following, we assume that the end-to-end distortion D can be well approximated by a weighting sum of the distortion $D_i = \mathrm{E}[(Y_i - \widehat{Y}_i)^2]$ of each transform coefficient as

$$D \approx \frac{1}{N} \sum_{i=1}^{N} w_i D_i. \tag{2}$$

Approximation (2) is an equality when $\mathbf{T}$ is orthogonal or when $\mathbf{Y}$ has independent components; it is a good approximation for any distortion when the transform coefficients are close to independence and last it is a good approximation for any $\mathbf{T}$ under the high-rate quantization hypothesis.

2.1 Entropy-constrained scalar quantization

A scalar quantizer Q approximates a random variable Y by a quantized variable $\widehat{Y} = Q(Y)$. Q is a mapping from a source alphabet $\mathbb{R}$ to a reproduction codebook $\mathcal{C} = \{\widehat{y}_i\}_{i \in \mathcal{K}} \subset \mathbb{R}$, where $\mathcal{K}$ is an arbitrary countable index set. We denote $p_i = P\{\widehat{Y} = \widehat{y}_i\}$. The Shannon theorem [3] proves that the entropy $H(\widehat{Y}) = -\sum_i p_i \log_2 p_i$ is a lower bound of the average number of bits per symbol used to encode the values of $\widehat{Y}$. Arithmetic entropy coding [17, 18] achieves an average bit rate that can be arbitrarily close to the entropy lower bound; therefore, we shall consider that this lower bound

[1] Indeed, the transform coefficients $(Y_1, \ldots, Y_N)$ are then independent and hence the quantization noises $(Y_1 - \widehat{Y}_1, \ldots, Y_N - \widehat{Y}_N)$ which are deterministic functions of the transform coefficients are uncorrelated, moreover they are centered when the quantizers satisfy the centroid condition.

is reached. An *entropy constrained scalar quantizer* is designed to minimize $H(\widehat{Y})$ for a fixed mean square distortion $D = \mathrm{E}[(Y - \widehat{Y})^2]$. It is well known [3] that for a fixed distortion D, under the high-resolution quantization hypothesis and if we assume that the random variable Y admits a probability density fonction (pdf) $p(y)$, then the minimum average bit rate $R = H(\widehat{Y})$ is achieved by a uniform quantizer, and $R \approx h(Y) - \frac{1}{2}\log_2(12D)$, where $h(Y) = \int \log_2[p(y)]\, p(y)\, dy$ is the differential entropy of Y. Generally it is preferable to introduce the variance σ^2 of Y and the differential entropy $h(\tilde{Y})$ of the standardized random variable $\tilde{Y} = (Y - \mathrm{E}[Y])/\sigma$: $h(\tilde{Y}) = h(Y) - \log_2 \sigma$ in order to separate the contribution of the signal power with that of its pdf shape (of course, this is possible only if Y admits finite second order statistics). The distortion rate satisfies then

$$D \approx c\sigma^2 2^{-2R}, \qquad \text{with} \quad c = \frac{2^{2h(\tilde{Y})}}{12}, \tag{3}$$

where the constant c depends only on the pdf shape.

2.2 Optimal bit allocation

Coding (quantizing and entropy coding) each transform coefficient Y_i separately splits the total number of bits among the transform coefficients in some manner. This bit allocation problem can be stated this way: one is given a set of quantizers described by their distortion-rate performances as $D_i \approx c_i\sigma_i^2 2^{-2R_i}$, $R_i \in \mathcal{R}_i$ for $(1 \leq i \leq N)$. Each set of available rates $\mathcal{R}_i$ is a subset of the nonnegative real numbers and may be discrete or continuous. The problem is to minimize the end-to-end distortion D in eq. (2) given a maximum average rate $R = N^{-1}\sum_{i=1}^{N} R_i$.

It results of the mean theorem (i.e., the arithmetic mean of the w_iD_is is not smaller than their geometric mean, with equality if and only if all the terms are equal) that, under the constraint of a given average rate R, the distortion D is minimum if and only if all the w_iD_is are equal, in which case

$$D_{\mathbf{T}}(R) \approx \left(\prod_{i=1}^{N} w_i c_i\right)^{1/N} \left(\prod_{i=1}^{N} \sigma_i^2\right)^{1/N} 2^{-2R}, \tag{4}$$

where σ_i^2 is the variance of Y_i and c_i is the constant associated with the standardized variable of Y_i according to the relation (3). When no transform $\mathbf{T}$ is applied to $\mathbf{X}$, or equivalently when $\mathbf{T}$ is the identity $\mathbf{I}$, the minimum distortion associated with the same maximum average rate R is given by $D_{\mathbf{I}}(R) \approx \left(\prod_{i=1}^N c_i^\star\right)^{1/N} \left(\prod_{i=1}^N \sigma_i^{\star 2}\right)^{1/N} 2^{-2R}$, where $\sigma_i^{\star 2}$ is the variance of X_i, and $c_i^\star$ is the constant associated with the standardized variable of X_i according to the relation (3).

2.3 Generalized coding gain and maximum reducible bits

In this paragraph we present a criterion called the generalized coding gain (resp. the generalized maximum reducible bits) which is a generalization of the coding gain [3] (resp. the maximum reducible bits) to non-Gaussian signals and non orthogonal linear transforms.

The distortion rate (4) can be used to define a figure of merit that we call the *generalized coding gain*

$$G^\star = \frac{D_{\mathbf{I}}(R)}{D_{\mathbf{T}}(R)} \approx \frac{\left(\prod_{i=1}^N c_i^\star\right)^{1/N} \left(\prod_{i=1}^N \sigma_i^{\star 2}\right)^{1/N}}{\left(\prod_{i=1}^N w_i c_i\right)^{1/N} \left(\prod_{i=1}^N \sigma_i^2\right)^{1/N}}. \tag{5}$$

It is the factor by which the distortion is reduced because of the linear transform $\mathbf{T}$, assuming high rate quantization and optimal bit allocation. Taking the inverse of (4), we obtain the optimal rate distortion: $R_{\mathbf{T}}(D) \approx \frac{1}{N}\sum_{i=1}^N h(Y_i) + \frac{1}{2N}\sum_{i=1}^N \log_2 w_i - \frac{1}{2}\log_2[12D]$, from which we can define the *generalized maximum reducible bits* $R^\star = R_{\mathbf{I}}(D) - R_{\mathbf{T}}(D) = \frac{1}{2}\log_2 G^\star$. It is the quantity by which the average number of bits to code $\mathbf{X}$ is reduced because of the transform $\mathbf{T}$.

If we assume that $\mathbf{X}$ is Gaussian, then $\mathbf{Y}$ is Gaussian and we have $c_i = c_i^\star = \frac{\pi e}{6}$ $(1 \leq i \leq N)$, hence the generalized coding gain becomes $G^\star = \left(\prod_{i=1}^N \sigma_i^{\star 2}\right)^{1/N} \Big/ \left(\prod_{i=1}^N w_i \sigma_i^2\right)^{1/N}$. Furthermore, if we suppose that $\mathbf{T}$ is orthogonal, then the w_is are all equal to one and the generalized coding gain is identical to the well-known coding gain G appearing in texts and scholarly books, which is maximized for any Karhunen-Loève basis of $\mathbf{X}$ (see, for example, [3]).

2.4 Criterion satisfied by an optimal transform

Using the following relations (see, e.g., [17] for notions of information theory) $h(\mathbf{X}) = \sum_{i=1}^{N} h(X_i) - I(X_1;\ldots;X_N)$, $h(\mathbf{Y}) = \sum_{i=1}^{N} h(Y_i) - I(Y_1;\ldots;Y_N)$, $h(\mathbf{Y}) = h(\mathbf{X}) + \log_2 |\det \mathbf{T}|$, where $I(X_1;\ldots;X_N)$ and $I(Y_1;\ldots;Y_N)$ denote respectively the mutual information between the components of $\mathbf{X}$ and of $\mathbf{Y}$, $h(\mathbf{X})$ and $h(\mathbf{Y})$ the differential entropy of vectors $\mathbf{X}$ and $\mathbf{Y}$ respectively, the generalized maximum reducible bits can be expressed as follows:

$$R^\star = \frac{1}{N} I(X_1;\ldots;X_N) - \frac{1}{N} I(Y_1;\ldots;Y_N) - \frac{1}{N}\log_2 |\det \mathbf{T}| - \frac{1}{2N}\log_2 \prod_{i=1}^{N} w_i. \tag{6}$$

Moreover, according to the expression of the w_is, and noting $\mathrm{Diag}(\mathbf{C})$ for the diagonal matrix having the same main diagonal as $\mathbf{C}$, the last two terms in (6) are equal to $-\frac{1}{2N}\log_2\left(\frac{\det[\mathrm{Diag}(\mathbf{T}^{-T}\mathbf{T}^{-1})]}{\det[\mathbf{T}^{-T}\mathbf{T}^{-1}]}\right)$. Hence we deduce the following expression of the generalized maximum reducible bits:

$$R^\star = \frac{1}{N} I(X_1;\ldots;X_N) - \frac{1}{N} I(Y_1;\ldots;Y_N) - \frac{1}{2N}\log_2\left(\frac{\det[\mathrm{Diag}(\mathbf{T}^{-T}\mathbf{T}^{-1})]}{\det[\mathbf{T}^{-T}\mathbf{T}^{-1}]}\right). \tag{7}$$

Let us remark, according to Hadamard's inequality (see e.g., [17]), that for any definite positive matrix $\mathbf{C}$ of order N, the quantitity $\log_2\left(\frac{\det[\mathrm{Diag}(\mathbf{C})]}{\det \mathbf{C}}\right)$ is greater or equal than zero, with equality if and only if the matrix $\mathbf{C}$ is diagonal.

It is now clear that the problem of finding a linear transform $\mathbf{T}$ which maximizes the generalized coding gain $G^\star$ defined in (5) is the same problem as finding $\mathbf{T}$ which maximizes $R^\star$, or equivalently, finding the linear transform $\mathbf{T}$ which minimizes the contrast

$$\mathcal{C}(\mathbf{T}) = I(Y_1;\ldots;Y_N) + \frac{1}{2}\log_2\left(\frac{\det[\mathrm{Diag}(\mathbf{T}^{-T}\mathbf{T}^{-1})]}{\det[\mathbf{T}^{-T}\mathbf{T}^{-1}]}\right). \tag{8}$$

The first term of (8) is a measure of the statistical dependence between the transform coefficients Y_i. It is always non-negative, and zero if and only if the variables are statistically independent. As for

the second term, it is always non-negative, and zero if and only if the column vectors of $\mathbf{T}^{-1}$ are pairwise orthogonal. Furthermore, if $\mathbf{D}$ is a diagonal matrix, one can verify that $\mathcal{C}(\mathbf{DT}) = \mathcal{C}(\mathbf{T})$, i.e., the contrast is scale invariant. Intuitively, this is not surprising since multiplying each component by a factor λ_i and each quantization step by the same factor has no impact on both the final rate and the end-to-end distortion. We can now state the following theorem.

Theorem 1 *Under the high-rate quantization hypothesis, for entropy-constrained scalar quantizers, mean-square distortion and the condition* $\mathbf{U} = \mathbf{T}^{-1}$*, a linear transform* $\mathbf{T}$ *is optimal in coding if and only if it minimizes the contrast* $\mathcal{C}(\mathbf{T})$ *of relation (8).*

Remark now that the contrast $\mathcal{C}(\mathbf{T})$ is always non-negative, and that it is equal to zero if and only if $\mathbf{T}^{-1}$ is a transform with orthogonal columns which produces independent coefficients.

3 Link between ICA and high-rate transform coding

3.1 Criteria

The criterion (8) may be decomposed into $\mathcal{C}(\mathbf{T}) = \mathcal{C}_{\mathrm{ICA}}(\mathbf{T}) + \mathcal{C}_{\mathrm{O}}(\mathbf{T})$, where $\mathcal{C}_{\mathrm{ICA}}(\mathbf{T}) = I(Y_1; \ldots; Y_N)$ corresponds to the mutual information criterion in ICA, and

$$\mathcal{C}_{\mathrm{O}}(\mathbf{T}) = \frac{1}{2} \log_2 \left[\frac{\det[\mathrm{Diag}(\mathbf{T}^{-T}\mathbf{T}^{-1})]}{\det[\mathbf{T}^{-T}\mathbf{T}^{-1}]} \right]. \tag{9}$$

The second term $\mathcal{C}_{\mathrm{O}}(\mathbf{T})$ measures a pseudo-distance to orthogonality of the transform $\mathbf{T}$. In general, the optimal transform $\mathbf{T}_{\mathrm{opt}}$ in transform coding, i.e., the transform which minimizes the contrast (8), will be different from that $\mathbf{T}_{\mathrm{ICA}}$ which minimizes the first term of (8), i.e., the solution of the ICA problem. It is important to notice here that the classical assumption made in blind source separation problems, that is the observations are obtained from a linear mixing of independent sources, is not really required in the problem of finding the transform that maximizes the generalized coding gain.

The expression of the contrast (8) depends on the definition of the distortion. In this work, we

measure the distortion as mean squared error, therefore it is not surprising that orthogonal transforms are favored over other linear transforms since they are energy-preserving.

3.2 Modified ICA algorithms for coding

In appendix, we propose two variants of the `icainf` algorithm by Pham [14] for the minimization of the contrast (8). The first one, called *Generalized Coding Gain Supremum* (`GCGsup`) gives the transform $\mathbf{T}_{\text{opt}}$ that minimizes the contrast (8), the second, called *Orthogonal Independent Composent Analysis* (`OrthICA`) find the orthogonal matrix $\mathbf{T}_{\text{orth}}$ that minimizes the contrast $\mathcal{C}_{\text{ICA}}(\mathbf{T})$.

The minimization of the criterion (8) can be done through a gradient descent algorithm, but a much faster method is the Newton algorithm (which amounts to using the natural gradient [19]). As in [14], because of the multiplicative structure of our optimization problem, we use multiplicative increment of the parameter $\mathbf{T}$ rather than additive increment. Starting with a current estimator $\widehat{\mathbf{T}}$, it consists of expanding $\mathcal{C}(\widehat{\mathbf{T}} + \mathcal{E}\widehat{\mathbf{T}})$ with respect to the matrix $\mathcal{E}$ up to second order and then minimizing the resulting quadratic form in $\mathcal{E}$ to obtain a new estimate. Note that the parameter $\mathcal{E}$ is a matrix of order N. This method requires the computation of the Hessian[2] of $\mathcal{C}(\widehat{\mathbf{T}} + \mathcal{E}\widehat{\mathbf{T}})$ with respect to $\mathcal{E}$, which is quite involved. For this reason, we will approximate it by the Hessian of $\mathcal{C}(\widehat{\mathbf{T}} + \mathcal{E}\widehat{\mathbf{T}})$, computed under the assumption that the transform coefficients Y_i are independent. The method is then referred to as quasi-Newton. Although those simplifications result in a slower convergence speed towards the solution, they cause the robustness of the algorithm to be improved by reducing the risk of divergence when the initial estimator $\widehat{\mathbf{T}}_0$ is far from the final solution. Note that the final solution is the same as that obtained without simplification since the algorithm consists of cancelling the first order terms in the expansion of $\mathcal{C}(\mathbf{T} + \mathcal{E}\mathbf{T})$.

[2]The Hessian of a function of several variables is the matrix of its second partial derivatives.

4 Experimental results

In this section, we are interested in assessing the performances of $\mathbf{T}_{\text{opt}}$ and $\mathbf{T}_{\text{orth}}$ in transform coding. Results included in this paper show coding performances on two synthetic data sets and a natural image data set. The synthetic data sets are used to show that $\mathbf{T}_{\text{opt}}$ and $\mathbf{T}_{\text{orth}}$ can outperform the KLT when used for high-rate transform coding of non-Gaussian signals. As for the second data set, it consists of well-known grayscale natural images which will be used to evaluate the performances of $\mathbf{T}_{\text{opt}}$ and $\mathbf{T}_{\text{orth}}$ in medium-to-low bit rate image compression.

4.1 Examples of synthetic signals efficiently compressed with `GCGsup` and `OrthICA`.

4.1.1 The tested signals

The first synthetic data set consists of 2^{16} samples of a bidimensional ($N = 2$) random vector $\mathbf{X}$ obtained as follows. First, we produce 2^{16} samples of a white vector $\mathbf{S} = [S_1, S_2]^T$ whose the i-th component is the standardized random variable associated to $S'_i = \text{Sign}(Z_i).|Z_i|^\alpha$, where $[Z_1, Z_2]^T$ is a standardized white Gaussian random vector. The exponent α is an arbitrary positive real number. When $\alpha > 1$ (resp. $\alpha < 1$), S_i is super- (resp. sub-) Gaussian. Then, the vector $\mathbf{X}$ is obtained via a mirror symmetry, according to the operation $\mathbf{X} = \mathbf{BS}$, with $\mathbf{B} = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ an orthogonal matrix. Samples of $\mathbf{S}$ and $\mathbf{X}$ are depicted respectively on the first row and the second row in Fig. (2) for four different values of α. The second synthetic data set consists of 2^{16} samples of a bidimensional random vector $\mathbf{X}$ obtained as follows: $\mathbf{X} = \mathbf{BU}$, where $\mathbf{B}$ is given above and $\mathbf{U} = [U_1, U_2]^T$ with U_1 and U_2 independent and identically distributed uniform random variables on $[-1, 1]$. Samples of $\mathbf{U}$ and $\mathbf{X}$ are depicted respectively in Fig. 3(a) and Fig. 3(b).

4.1.2 Evaluation methodology

Our objective metric to measure the performance of a transform is the generalized coding gain (5). Given both N sources and a transform, the estimation of (5) requires good estimates of the pdfs of each source as well as each transformed component, which may be very difficult to obtain. In this section, we elaborate on a more "practical" way of evaluating (5) which consists in actually coding the transformed components and measuring both the bit-rate—given here by the first order entropy of the quantized data—and the actual end-to-end distortion.

For each tested signal, the vector $\mathbf{X}$ is first linearly transformed by a transform $\mathbf{T}$ to produce the vector $\mathbf{Y} = [Y_1, \ldots, Y_N]^T$ whose components are coded separately. The i-th component Y_i is first high-rate quantized with a uniform scalar quantizer of quantization step q_i. This gives $\widehat{Y}_i$. The bit-rate R_i is then estimated by computing the empirical first order entropy of $\widehat{Y}_i$ and the inverse transform is applied to $\widehat{\mathbf{Y}} = [\widehat{Y}_1, \ldots, \widehat{Y}_N]^T$ in order to reconstruct an approximation $\widehat{\mathbf{X}} = [\widehat{X}_1, \ldots, \widehat{X}_N]^T$ of $\mathbf{X}$. The distortion is the end-to-end one, $D = \frac{1}{N}\,\mathrm{E}[\|\mathbf{X} - \widehat{\mathbf{X}}\|^2]$ and the total average rate is the empirical mean of the N rates R_i $(1 \leq i \leq N)$. The optimal allocation of rates between the transform coefficients results in equal weighted distortions $w_i D_i$ (see section 2). Moreover, using uniform scalar quantizers, bit allocation amounts to choosing a quantization step q_i for each of the N components, and for small q_i the distortion D_i may be well approximated by $q_i^2/12$, [3]. Therefore, the bit allocation follows a simple rule : let c be a constant, then make all the quantization steps $q_1, \ldots, q_N$ such that $w_i D_i = c$. This gives $q_i = \sqrt{12c/w_i}$, for $i = 1, \ldots, N$. When the constant c varies (under the assumption of high resolution quantization for each component) we obtain the classical asymptotic curve (distortion versus bit-rate, or equivalently bit-rate versus distortion). In our tests, we consider that the hypothesis of high resolution quantization is valid when for each component Y_i, the relative deviation between the actual distortion $\mathrm{E}[(Y_i - \widehat{Y}_i)^2]$ (where the expectation is estimated by empirical mean) and $q_i^2/12$ is not greater than 1%. For a given high bit-rate, the ratio between the end-to-end distortion read on the asymptotic curve obtained using the identity transform and that read on the asymptotic curve

obtained using $\mathbf{T}$ yields the generalized coding gain of $\mathbf{T}$.

4.1.3 Results

The first set of synthetic data was designed so that the KLT does nothing on it. Indeed, the random vector $\mathbf{X}$ being white, the generalized coding gain $G^\star$ of the KLT is equal to 0 dB. However, the components of $\mathbf{X}$ are not independent, as can be seen on the second row in Fig. 2, and any algorithm among `GCGsup`, `OrthICA` and `icainf` gives the same result: the components Y_i are independent (see third row in Fig. 2), and the generalized coding gain $G^\star$ is the same. The generalized coding gain associated with using the transform returned by `GCGsup` is presented in Tab. 1 for four different values of α. Except for $\alpha = 1$, i.e., when the components of $\mathbf{X}$ are Gaussian, the generalized coding gain of $\mathbf{T}_{\text{opt}}$ is always greater than zero and becomes more and more important as $|1 - \alpha|$ increases, i.e., as the components of $\mathbf{X}$ depart from Gaussianity.

The second set of synthetic data has recently drawn attention of researchers in transform coding. In [6], the authors have investigated the strengths and limitations of the KLT when it is used for transform coding the vector $\mathbf{X} = \mathbf{B}\mathbf{U}$. While the KLT for $\mathbf{X}$ is not unique, practical implementations of the KLT give a matrix close to the identity. As a consequence, the diamond distribution of $\mathbf{X}$ remains practically unchanged after applying the KLT on $\mathbf{X}$ (see Fig. 3(c)). Yet $\mathbf{X}$ and $\mathbf{U}$ are not equally good sources for scalar quantization as was already pointed out earlier [21]. This is confirmed in our experiments since any algorithm among `GCGsup`, `OrthICA` and `icainf` transforms the diamond distribution back into the square distribution of $\mathbf{U}$ as can be seen in Fig. 3(d). Furthermore, Tab. 2 shows that the generalized coding gain $G^\star$ of the transform returned by `GCGsup` outperforms that obtained with the KLT.

4.2 Application of `GCGsup` and `OrthICA` to the compression of natural images.

The next experimental tests aim at assessing the performances of `GCGsup` and `OrthICA` when applied to the compression of well-known grayscale natural images (*Lenna*, *Goldhill*, *Mandrill*, *Peppers*, *Zelda* and

Boat) each of size 512×512 pixels and coded using 8 bits per pixel (bpp) (see Fig. 4). The image coder used in our tests has been designed for experimentation and is not intended to outperform current state-of-the-art image coders such as JPEG2000 [22]. In natural image compression, we are usually interested in performances at medium-to-low bit rates. In our experiments, we chose to investigate the performances of `GCGsup` and `OrthICA` at bit rates less or equal than 2 bpp. Note that for bit-rates less than about 1.6 bpp, the transforms returned by `GCGsup` and `OrthICA` can no longer be considered as optimal since the high resolution hypothesis is no longer valid[3].

4.2.1 Bases estimation

The modified ICA bases (i.e., the column vectors of $\mathbf{T}$) were estimated according to two learning schemes. The first scheme yields 12 different bases (2 per image): for each test image, the algorithms `GCGsup` and `OrthICA` were applied to a training set consisting of 4 096 non overlapping image blocks each of size 8×8 pixels extracted from the test image. The first and second columns of Fig. 5 displays the estimated modified ICA bases as well as the practically achieved KLT bases obtained from the test images *Boat* and *Peppers*, respectively. Also displayed for comparison are the bases obtained using the ICA algorithm `icainf`. As for the second scheme, it yields only 2 different bases. The modified ICA bases were learned from one training set consisting of 12 288 non overlapping image blocks each of size 8×8 pixels extracted from three test images (*Lenna*, *Goldhill* and *Boat*). The third column of Fig. 5 displays the estimated modified ICA bases as well as the KLT basis (denoted KLT*). The transform whose column vectors were estimated using the algorithm `OrthICA` (resp. `GCGsup`) is denoted $\mathbf{T}^\star_{\text{orth}}$ (resp. $\mathbf{T}^\star_{\text{opt}}$). The ICA basis obtained with the `icainf` algorithm for this training set is also displayed for comparison and denoted $\mathbf{T}^\star_{\text{ICA}}$.

Examining Fig. 5 closely reveals that the features found with `GCGsup` and `OrthICA` are much more localized in space than the checkerboardlike basis vectors obtained with the KLT. Notice also the more

[3] As in Section 4.1.2, we consider that the hypothesis of high resolution quantization is valid when for each component Y_i, the relative deviation between the actual distortion $\mathrm{E}[(Y_i - \widehat{Y}_i)^2]$ (where the expectation is estimated by empirical mean) and $q_i^2/12$ is not greater than 1%.

pronounced edge-like nature of the modified ICA bases, regardless of the learning scheme employed. Other similar experiments with ICA [23] have produced comparable results.

The computation of the pseudo-distance to orthogonality (see eq. (9)) reveals that, unlike the bases obtained with the ICA algorithm `icainf`, those estimated with `GCGsup` are quasi-orthogonal as can be seen in Tab. 3. As mentioned earlier, it is not surprising that orthogonal transforms are favored over other linear transforms since we measure the distortion as mean squared error and orthogonal transforms are energy-preserving.

4.2.2 Medium-to-high bit rates

Tab. 4 shows estimations of the generalized coding gain for each tested transform and each test image. The average generalized coding gain computed over the set of test images is also given. The estimation method used is the same as that described in section 4.1.2. Looking at the average values of the generalized coding gain reveals that, whatever the learning scheme, the modified ICA transforms perform best followed respectively by the 2-D DCT, the KLT and, far behind, the ICA transform. The coding gain of any of the modified ICA transforms relative to the 2-D DCT is about 0.3 dB (resp. 0.1 dB) when the first (resp. second) learning scheme is used suggesting that a transform-based image coder could benefit from using any of the modified ICA transforms. However, these results should be taken with care since they are meaningful only under the high resolution hypothesis (i.e., in general, at medium-to-high bit rates). Yet, in image compression, we are most interested in performances at medium-to-low bit rates, where the high resolution hypothesis is no longer valid. Additional precaution should be taken for the bases estimated according to the first learning scheme. In this case, the basis vectors need to be transmitted (coded) with the image resulting in an extra coding cost which has not been incorporated in the method used for estimating the generalized coding gain.

Another motivation for using the modified ICA bases in compression lies in the transform coefficients distribution which tends to be heavy-tailed, more suited for quantization and entropy coding than the

transform coefficients of the KLT and 2-D DCT, which tend to a normal distribution. Fig. 6 displays for each transform and each image the average kurtosis computed over 63 transformed components (the component which is equivalent to the DC component of the KLT was omitted). The kurtosis was normalized so that it is equal to zero in the case of Gaussian samples. The average of this *average kurtosis* was computed over the set of test images. The result for each transform is displayed over the label "combined" in Fig. 6. Results show that the kurtosis obtained with the modified ICA bases are greater than those obtained with the KLT and 2-D DCT, regardless of the transform and image considered. Note that the ICA bases exhibit even higher kurtosis, yet yielding the worst generalized coding gains among all tested transforms. The explanation for this is straightforward: ICA bases were estimated using the ICA algorithm `icainf` which aims at minimizing the mutual information between the components, i.e., only the first term in eq. (8). This is done without putting any orthogonality constraint on the basis vectors resulting in a transform which is far from being orthogonal (see Tab. 3). This result highlights the importance of trading-off between independence and orthogonality—when the distortion is measured using mean squared error. This is well illustrated by eq. (8) (a discussion about orthogonality and independence can also be found in [4]).

4.2.3 Medium-to-low bit rates

The image coder used in our experiment is a transform coder originally developed by Davis[4]. It is very modular and allows for simple replacements of individual components (quantizer, entropy coder, transform). It was modified so that it ressembles a JPEG-like coder. The image to be coded is first "tiled" into blocks of eight by eight pixels each, then each tile is represented into a new basis using one of the tested transforms. The bases obtained using the first learning scheme are transmitted with the image since they are data-dependent bases. As for the bases estimated using the second learning scheme, they are *not* transmitted with the image. Quantization steps are chosen to minimize the end-to-end distortion (2) subject to bit rate constraint. The bit allocation procedure is based on

[4]http://www.geoffdavis.net/dartmouth/wavelet/wavelet.html

integer programming algorithms described in [24] which provide optimal or near-optimal allocations for the quantizers included here. Entropy coding of the quantizer output is carried out by an adaptive arithmetic coder.

4.2.4 Compression results

We now compare the compression performances of the previously mentioned transforms: KLT, $\mathbf{T}_{\text{opt}}$, $\mathbf{T}_{\text{orth}}$, KLT*, $\mathbf{T}^\star_{\text{opt}}$, $\mathbf{T}^\star_{\text{orth}}$ and 2-D DCT. The results are displayed in Tab. 5, in which we present the peak signal to noise ratio (PSNR) as a function of bit-rate for each test image. Whatever the image, the plots have these common characteristics:

1) The transform codes based on KLT*, $\mathbf{T}^\star_{\text{opt}}$, $\mathbf{T}^\star_{\text{orth}}$ and the 2-D DCT perform better than those based on the KLT, $\mathbf{T}_{\text{opt}}$ and $\mathbf{T}_{\text{orth}}$, regardless of the bit-rate. The poor coding performances of the KLT, $\mathbf{T}_{\text{opt}}$ and $\mathbf{T}_{\text{orth}}$ are mainly due to the coding penalty resulting from coding the basis vectors (11 bits were allocated on average to each matrix coefficient resulting in a coding precision of 10^{-3}). In our tests, we observed that the coding penalty is about 1 dB at 2 bpp and 1 bpp, about 2 dB at 0.5 bpp and can reach up to 5 dB at 0.25 bpp. The rate-distortion curves obtained with the KLT, $\mathbf{T}_{\text{opt}}$ and $\mathbf{T}_{\text{orth}}$ without coding the basis vectors (not plotted here for the sake of clarity) were close to those obtained with respectively KLT*, $\mathbf{T}^\star_{\text{opt}}$ and $\mathbf{T}^\star_{\text{orth}}$ (no meaningful performance difference could be observed).

2) At bit-rates greater than 1 bpp, all the rate-distortion curves are lines of slope about $5 \sim 6$ dB per bit. We observed that the high-resolution hypothesis is verified for bit-rates greater than about 1.6 bpp (in this case, the slope of the curve is ~ 6 dB per bits).

3) Whatever the bit-rate, no meaningful performance difference can be observed between the class-adapted transform codes based on $\mathbf{T}^\star_{\text{opt}}$ and those based on $\mathbf{T}^\star_{\text{orth}}$. For medium-to-high bit rates (i.e., greater than about 1.6 bpp), this result can be more or less predicted by looking at the values of the corresponding generalized coding gains (see Tab. 4). For bit-rates greater than 1 bpp, the

performance difference between either $\mathbf{T}^{\star}_{\text{opt}}$ or $\mathbf{T}^{\star}_{\text{orth}}$ and KLT* is equal to about the difference between their corresponding generalized coding gains suggesting that the performance of our entropy coder is close to that of a perfect first-order entropy coder. In the low bit-rate region (< 1 bpp), the performance difference tends to become smaller. As for the 2-D DCT, its performance is comparable with that of any of the class-adapted modified ICA bases. Thus, our approach has made it possible to learn two bases which are competitive with the well-known 2-D DCT basis according to the standard PSNR measure.

Tab. 5 compares the performance in compression of the 2-D DCT, the class-adapted KLT and both class-adapted modified ICA bases. The corresponding values of PSNR and achieved bit-rates (the target bit-rate was set equal to 0.5 bpp) are shown. Although the KLT*-coded images have worse PSNR than the others, no difference in term of visual quality can be seen. However, when looking further into the details of the reconstructed images *Boat* in Fig. 7, some meaningful visual quality difference can be seen. Fig. 7) show the images coded with $\mathbf{T}^{\star}_{\text{orth}}$. Black arrows point towards details which are not present or blurred on the corresponding images coded with the 2-D DCT. These details represent lines (e.g., some ropes in the case of the image *Boat*) which are well preserved with $\mathbf{T}^{\star}_{\text{orth}}$. These results suggest that the class-adapted modified ICA bases are better suited to coding fine details such as lines and edges compared to the 2-D DCT. This is not quite surprising given the more pronounced edge-like nature of the modified ICA bases (see Fig. 5).

5 Conclusion

This paper addresses the problem of finding optimal 1-D linear transforms in transform coding without the classical assumption of Gaussianity. This paper emphasizes a new point of view in variable rate transform coding by showing, under the high resolution hypothesis, that the problem of finding the optimal 1-D linear transform may be recast as a modified independent component analysis (ICA) problem. Two new modified ICA algorithms, called `GCGsup` and `OrthICA`, are introduced for computing

the optimal 1-D linear transform and the optimal 1-D orthogonal transform, respectively.

Experimental results included in this paper show coding performances of `GCGsup` and `OrthICA` on two synthetic data sets and a natural image data set. Experiments carried out with the synthetic data sets show that the new transforms can outperform the KLT when used for high-rate transform coding of non-Gaussian signals. When applied to the compression of some well-known natural images, `GCGsup` and `OrthICA` have proved 1) to be comparable to the classical 2-D DCT according to the PSNR measure and 2) to yield better visual image quality (better preservation of lines and edges) than the 2-D DCT.

Acknowledgment

The authors are grateful to Pierre Duhamel for his very helpful comments as well as Jacques Weidig and Jean-Louis Gutzwiller who took part in the simulation work.

A Appendix

A.1 Algorithm `GCGsup`

Given that $I(Y_1, \ldots, Y_N) = \sum_i h(Y_i) - h(\mathbf{Y})$ and $h(\mathbf{Y}) = h(\mathbf{X}) + \log_2 |\det \mathbf{T}|$, and since the term $h(\mathbf{X})$ does not depend on $\mathbf{T}$, minimizing the contrast (8) is the same as minimizing $\widetilde{\mathcal{C}}(\mathbf{T}) = \mathcal{C}_{\mathrm{O}}(\mathbf{T}) + \widetilde{\mathcal{C}}_{\mathrm{ICA}}(\mathbf{T})$ where $\widetilde{\mathcal{C}}_{\mathrm{ICA}}(\mathbf{T}) = \sum_{i=1}^{N} h(Y_i) - \log_2 |\det \mathbf{T}|$. Using the results of [20] it can be seen that the Taylor expansion of $\widetilde{\mathcal{C}}_{\mathrm{ICA}}(\mathbf{T} + \mathcal{E}\mathbf{T})$ up to second order may be approximated as follows

$$\widetilde{\mathcal{C}}_{\mathrm{ICA}}(\mathbf{T} + \mathcal{E}\mathbf{T}) = \widetilde{\mathcal{C}}_{\mathrm{ICA}}(\mathbf{T}) + \sum_{1 \le i \ne j \le N} \mathrm{E}[\psi_{Y_i}(Y_i) Y_j] \mathcal{E}_{ij} + \frac{1}{2} \sum_{1 \le i \ne j \le N} \{\mathrm{E}[\psi^2_{Y_i}(Y_i)]\, \mathrm{E}[Y_j^2] \mathcal{E}_{ij}^2 + \mathcal{E}_{ij}\mathcal{E}_{ji}\} + \cdots, \tag{10}$$

where the function ψ_{Y_i} is equal to the derivative of $-\log_2 p(y_i)$ — $p(y_i)$ denoting the Y_i pdf — and is known as the score function. This approximation concerns only the second order terms in the expansion, but *not the first order terms*. It relies essentially on the assumption of independent transform

coefficients, which may not be valid if the solution of the ICA problem is far from the solution that minimizes the contrast (8). But it is quite useful since it leads to a decoupling in the quadratic form of the expansion. Let $\mathbf{M} = \mathbf{T}^{-T}\mathbf{T}^{-1}$. One may verify that the Taylor expansion of $\mathcal{C}_{\mathrm{O}}(\mathbf{T}+\mathcal{E}\mathbf{T})$ with respect to $\mathcal{E}$ and around $\mathcal{E} = \mathbf{0}$, up to second order, is given by $\mathcal{C}_{\mathrm{O}}(\mathbf{T}+\mathcal{E}\mathbf{T}) = \mathcal{C}_{\mathrm{O}}(\mathbf{T}) - \sum_{1\leq i\neq j\leq N} \frac{M_{ji}}{M_{ii}}\mathcal{E}_{ji} - \frac{1}{2}\sum_{1\leq i\neq j\leq N}\left[\mathcal{E}_{ij}\mathcal{E}_{ji} - 2\frac{M_{ji}}{M_{jj}}\mathcal{E}_{ii}\mathcal{E}_{ji}\right] + \sum_{i=1}^{N}\sum_{j=1,\, j\neq i}^{N}\sum_{k=1,\, k\neq i}^{N}\left[\left(\frac{M_{jk}}{2M_{ii}} - \frac{M_{ij}M_{ik}}{M_{ii}^2}\right)\mathcal{E}_{ji}\mathcal{E}_{ki} + \frac{M_{kj}}{M_{kk}}\mathcal{E}_{ji}\mathcal{E}_{ik}\right] + \cdots$. The quadratic form associated with the above expansion is quite involved and is not positive. One possible approximation consists in neglecting the non diagonal elements of $\mathbf{M}$, which amounts to assuming that the optimal linear transform is close to an orthogonal transform. Under this hypothesis, one may verify that

$$\mathcal{C}_{\mathrm{O}}(\mathbf{T}+\mathcal{E}\mathbf{T}) \approx \mathcal{C}_{\mathrm{O}}(\mathbf{T}) - \sum_{1\leq i\neq j\leq N}\frac{M_{ji}}{M_{ii}}\mathcal{E}_{ji} + \frac{1}{2}\sum_{1\leq i\neq j\leq N}\left[\frac{M_{jj}}{M_{ii}}\mathcal{E}_{ji}^2 + \mathcal{E}_{ji}\mathcal{E}_{ij}\right] + \cdots . \tag{11}$$

The quadratic form associated with the above expansion is now positive, but not positive definite. However, this is sufficient for the matrix associated with the quadratic form of the Taylor expansion of $\widetilde{\mathcal{C}}(\mathbf{T})$ to be positive definite, which ensures the stability of the iterative algorithm. Finally, by adding equation (10) and approximation (11) we obtain an approximation of $\widetilde{\mathcal{C}}(\mathbf{T}+\mathcal{E}\mathbf{T})$ up to second order of $\mathcal{E}_{ij}$ and the iteration consists explicitly of solving the linear equations

$$\begin{bmatrix} \mathrm{E}[\psi_{Y_i}^2(Y_i)]\mathrm{E}[Y_j^2] + \frac{M_{ii}}{M_{jj}} & 2 \\ 2 & \mathrm{E}[\psi_{Y_j}^2(Y_j)]\mathrm{E}[Y_i^2] + \frac{M_{jj}}{M_{ii}} \end{bmatrix} \begin{bmatrix} \mathcal{E}_{ij} \\ \mathcal{E}_{ji} \end{bmatrix} = \begin{bmatrix} \frac{M_{ij}}{M_{jj}} - \mathrm{E}[\psi_{Y_i}(Y_i)Y_j] \\ \frac{M_{ji}}{M_{ii}} - \mathrm{E}[\psi_{Y_j}(Y_j)Y_i] \end{bmatrix}.$$

The indeterminate diagonal terms $\mathcal{E}_{ii}$ are arbitrarily fixed to zero. Then the estimator $\widehat{\mathbf{T}}$ is left multiplied by $\mathbf{I}+\mathcal{E}$ in order to update it. In this expression, the probability density functions being unknown, the score function $\psi_{Y_i}(y_i)$ is replaced by an estimation (see [14]) and the expectations are estimated by empirical means.

A.2 Algorithm `OrthICA`

In this section, we propose to find the orthogonal transform that minimizes the contrast (8). Since the second term of (8) vanishes for any othogonal matrix $\mathbf{T}$, this amounts to finding the orthogonal transform which minimizes the first term of (8), or equivalently, which minimizes $\widetilde{\mathcal{C}}_{\mathrm{ICA}}(\mathbf{T})$. If the matrix $\mathbf{T}$ is orthogonal, so is $\mathbf{T}+\mathcal{E}\mathbf{T}$, providing that $\mathbf{I}+\mathcal{E}$ be orthogonal. This last condition will be satisfied up to second order if $\mathcal{E}$ is anti-symmetric, since $(\mathbf{I}+\mathcal{E})^T(\mathbf{I}+\mathcal{E}) = \mathbf{I}+\mathcal{E}^T\mathcal{E}$ differs from the identity only by second order terms. Let $\mathcal{E}$ be anti-symmetric. The Taylor expansion of $\widetilde{\mathcal{C}}_{\mathrm{ICA}}(\mathbf{T}+\mathcal{E}\mathbf{T})$ becomes $\widetilde{\mathcal{C}}_{\mathrm{ICA}}(\mathbf{T}+\mathcal{E}\mathbf{T}) = \widetilde{\mathcal{C}}_{\mathrm{ICA}}(\mathbf{T}) + \sum_{1\le i<j\le N}\left\{\mathrm{E}[\psi_{Y_i}(Y_i)Y_j] - \mathrm{E}[\psi_{Y_j}(Y_j)Y_i]\right\}\mathcal{E}_{ij} + \frac{1}{2}\sum_{1\le i<j\le N}\mathcal{E}_{ij}^2\left[\mathrm{E}[\psi_{S_i}^2(Y_i)]\,\mathrm{E}[Y_j^2] + \mathrm{E}[\psi_{Y_j}^2(Y_j)]\,\mathrm{E}[Y_i^2] - 2\right] + \cdots$, and the minimization of the second term in the above expansion yields

$$\mathcal{E}_{ij} = \frac{\mathrm{E}[\psi_{Y_j}(Y_j)Y_i] - \mathrm{E}[\psi_{Y_i}(Y_i)Y_j]}{\mathrm{E}[\psi_{Y_i}^2(Y_i)]\,\mathrm{E}[Y_j^2] + \mathrm{E}[\psi_{Y_j}^2(Y_j)]\,\mathrm{E}[Y_i^2] - 2}. \tag{12}$$

Actually, $\mathbf{T}+\mathcal{E}\mathbf{T}$ is not a true orthogonal transform. This may be overcome by replacing $\mathbf{T}+\mathcal{E}\mathbf{T}$ with $e^{\mathcal{E}}\mathbf{T} = (\mathbf{I}+\mathcal{E}+\mathcal{E}^2/2!+\cdots)\mathbf{T}$, which is an orthogonal matrix differing from $\mathbf{T}+\mathcal{E}\mathbf{T}$ only by second order terms.

References

[1] G. Wallace, "Overview of JPEG (ISO/CCITT) still image compression standard," *Commun. ACM*, vol. 4, no. 4, pp. 30–40, 1991.

[2] V. K. Goyal, J. Zhuang, and M. Vetterli, "Transform coding with backward adaptive updates," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1623–1633, July 2000.

[3] A. Gersho and R. Gray, *Vector Quantization and Signal Compression.* Kluwer, 1992.

[4] V. K. Goyal, "Theoretical foudations of transform coding," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 9–21, 2001.

[5] J.-Y. Huang and P. M. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Trans. Commun.*, vol. COM-11, pp. 289–296, Sept. 1963.

[6] M. Effros, H. Feng, and K. Zeger, "Suboptimality of the Karhunen-Loève transform for transform coding," *IEEE Trans. on Inform. Theory*, vol. 50, no. 8, pp. 1605–1619, 2004.

[7] S. Mallat and F. Falzon, "Analysis of Low Bit Rate Image Transform Coding," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 1027–1042, 1998.

[8] S. Jana and P. Moulin, "Optimality of KLT for high-rate transform coding of Gaussian vector-scale mixtures: application to reconstruction, estimation and classification," *IEEE Trans. Info. Th.* vol. 52, no. 9, pp. 4049–4067, 2006.

[9] D. Mary and D. Slock, "A theoretical high-rate analysis of causal versus unitary online transform coding," *IEEE Trans. Sig. Proc.*, vol. 54, no. 4, pp. 1472–1482.

[10] A. Bell and T. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vision Research*, vol. 37, pp. 3327–3338, 1997.

[11] A. T. Puga and A. P. Alves, "An experiment on comparing PCA and ICA in classical transform image coding," in *Proc. 1st Workshop on Blind Sepeparation and ICA*, pp. 105–108, 1998.

[12] S. Marusic and G. Deng, "ICA-FIR based image redundancy reduction," *Proc. 1st Int. Workshop on ICA and Signal Separation*, pp. 191–196, Aussois, France, 1999.

[13] A. Ferreira and M. Figueiredo, "Class-adapted image compression using independent component analysis," *Proc. IEEE Int. Conf. on Image Processing - ICIP'2003*, Barcelona, Spain, 2003.

[14] D. T. Pham, "Fast algorithms for mutual information based independent component analysis," *IEEE Transaction on Signal Processing*, vol. 52, no. 10, pp. 2690–2700, 2004.

[15] M. Narozny, M. Barret, D. T. Pham, I. P. Akam Bita, "Modified ICA algorithms for finding optimal transforms in transform coding", *Proc. IEEE 4th Int. Symp. on Image and Sig. Proc. and Analysis*, Zagreb (Croatie), pp. 111–116, 2005.

[16] M. Narozny and M. Barret, "ICA-based algorithms applied to image coding", *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Hawaii (USA), April 2007.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley & Sons, 1991.

[18] I. Witten, R. Neal, and J. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, no. 6, pp. 519-540, 1987.

[19] S. Amari, "Natural gradient works eficiently in learning", *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.

[20] D. T. Pham, "Entropy of a variable slightly contaminated with another," *IEEE Signal Processing Letters*, vol. 12, no. 7, pp. 536–539, 2005.

[21] H. Feng and M. Effros, "On the rate-distortion performance and computational efficiency of the Karhunen-Loève transform for lossy data compression," *IEEE Trans. Image Proc.*, vol. 11, no. 2, pp. 113–122, 2002.

[22] M. W. Marcellin, M. Gormish, A. Bilgin, M. Boliek, "An overview of JPEG2000," in *Proc. of Data Compression Conference*, Snowbird, Utah, March 2000.

[23] A. Hyvärinen, J. Tarhunen, and E. Oja, *Independent Component Analysis.* Wiley, 2001.

[24] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, 1988.

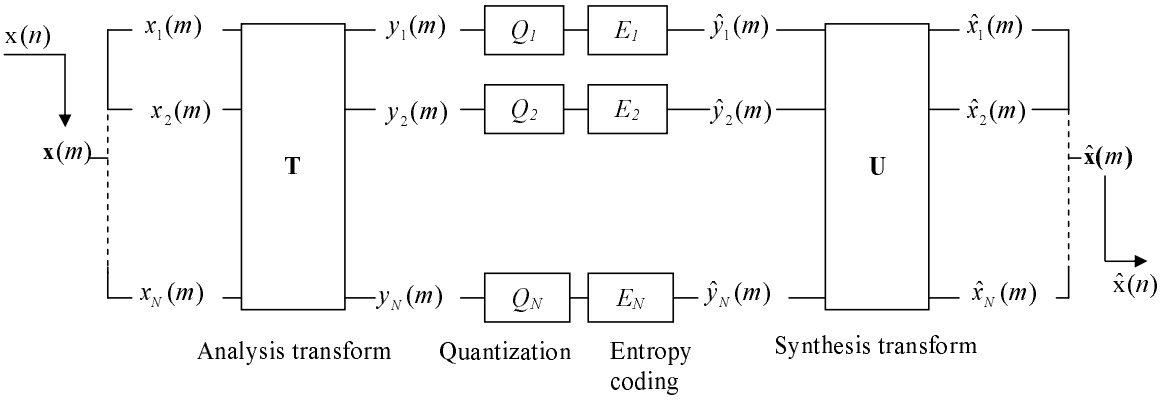
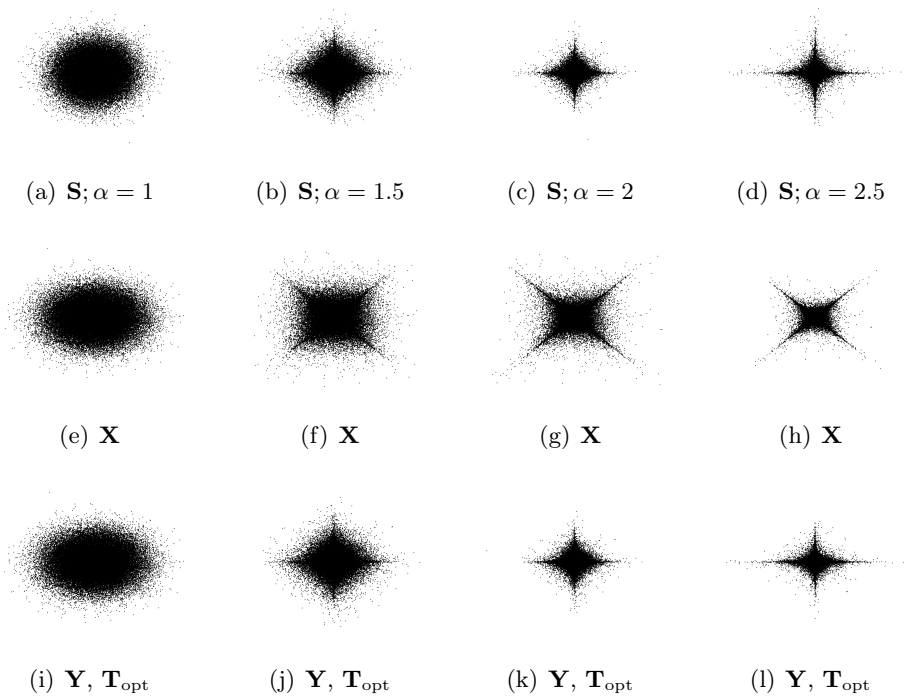


Figure 1: Compression system including a linear transform, a quantization and an entropy coding stage.

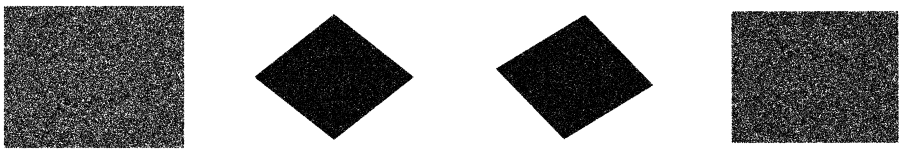


(a) $\mathbf{S}; \alpha = 1$ (b) $\mathbf{S}; \alpha = 1.5$ (c) $\mathbf{S}; \alpha = 2$ (d) $\mathbf{S}; \alpha = 2.5$

(e) $\mathbf{X}$ (f) $\mathbf{X}$ (g) $\mathbf{X}$ (h) $\mathbf{X}$

(i) $\mathbf{Y}$, $\mathbf{T}_{\text{opt}}$ (j) $\mathbf{Y}$, $\mathbf{T}_{\text{opt}}$ (k) $\mathbf{Y}$, $\mathbf{T}_{\text{opt}}$ (l) $\mathbf{Y}$, $\mathbf{T}_{\text{opt}}$

Figure 2: Samples of $\mathbf{S}$, $\mathbf{X}$ and $\mathbf{Y}$ (obtained after applying $\mathbf{T}_{\text{opt}}$ on each sample of $\mathbf{X}$) for four different values of α.



(a) $\mathbf{U}$ (b) $\mathbf{X}$ (c) $\mathbf{Y}$, KLT (d) $\mathbf{Y}$, $\mathbf{T}_{\text{opt}}$

Figure 3: Samples of $\mathbf{U}$, $\mathbf{X}$ and $\mathbf{Y}$ (obtained after applying the KLT and $\mathbf{T}_{\text{opt}}$ on $\mathbf{X}$).

α	1	$1,5$	2	$2,5$
$G^\star$ (dB)	0	$1,02$	$2,80$	$4,78$

Table 1: Generalized coding gain of $\mathbf{T}_{\text{opt}}$ for the first set of synthetic data.

Transform	KLT	$\mathbf{T}_{\text{opt}}$
$G^\star$ (dB)	$0,02$	$1,25$

Table 2: Generalized coding gains of the KLT and $\mathbf{T}_{\text{opt}}$ for the second set of synthetic data.

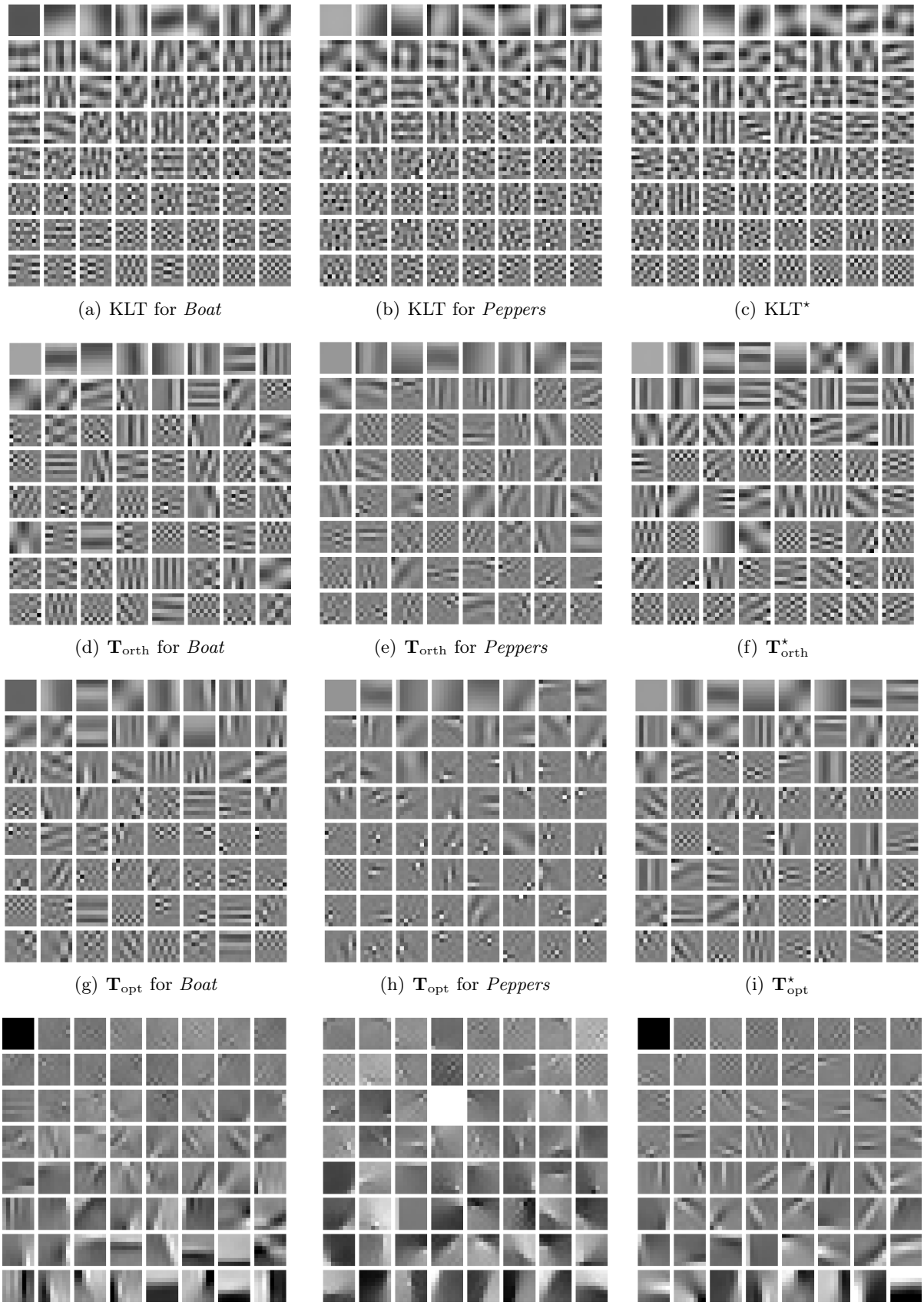
(a) Lenna (b) Goldhill (c) Mandrill

(d) Peppers (e) Zelda (f) Boat

Figure 4: Test images.

	Lenna	*Goldhill*	*Mandrill*	*Peppers*	*Zelda*	*Boat*
$\mathbf{T}_{\text{opt}}$	0.009	0.008	0.008	0.016	0.004	0.012
$\mathbf{T}^\star_{\text{opt}}$	0.006	0.006	0.006	0.006	0.006	0.006
$\mathbf{T}_{\text{ICA}}$	1.955	0.711	0.356	1.108	2.435	0.765
$\mathbf{T}^\star_{\text{ICA}}$	0.305	0.305	0.305	0.305	0.305	0.305

Table 3: Distance to orthogonality (in bpp) of $\mathbf{T}_{\text{opt}}$, $\mathbf{T}^\star_{\text{opt}}$, $\mathbf{T}_{\text{ICA}}$ and $\mathbf{T}^\star_{\text{ICA}}$ for each test image.



(a) KLT for *Boat* (b) KLT for *Peppers* (c) KLT*

(d) $\mathbf{T}_{\text{orth}}$ for *Boat* (e) $\mathbf{T}_{\text{orth}}$ for *Peppers* (f) $\mathbf{T}^\star_{\text{orth}}$

(g) $\mathbf{T}_{\text{opt}}$ for *Boat* (h) $\mathbf{T}_{\text{opt}}$ for *Peppers* (i) $\mathbf{T}^\star_{\text{opt}}$

(j) $\mathbf{T}_{\text{ICA}}$ for *Boat* (k) $\mathbf{T}_{\text{ICA}}$ for *Peppers* (l) $\mathbf{T}^\star_{\text{ICA}}$

Figure 5: KLT, $\mathbf{T}_{\text{orth}}$, $\mathbf{T}_{\text{opt}}$ and $\mathbf{T}_{\text{ICA}}$ basis vectors obtained from boat and peppers (on first and second column, respectively) and KLT*, $\mathbf{T}^\star_{\text{orth}}$, $\mathbf{T}^\star_{\text{opt}}$ and $\mathbf{T}^\star_{\text{ICA}}$ basis vectors obtained from the image class training set.

	Lenna	*Goldhill*	*Mandrill*	*Peppers*	*Zelda*	*Boat*	Average
KLT	18.13	15.22	6.99	17.23	19.91	15.35	15.47
$\mathbf{T}_{\text{orth}}$	18.58	15.62	7.42	17.89	20.12	15.96	15.93
$\mathbf{T}_{\text{opt}}$	18.60	15.68	7.49	17.84	20.12	15.94	15.94
KLT*	17.76	15.09	6.98	17.08	19.26	14.79	15.16
$\mathbf{T}^\star_{\text{orth}}$	18.34	15.40	7.27	17.63	19.89	15.74	15.73
$\mathbf{T}^\star_{\text{opt}}$	18.31	15.46	7.28	17.70	19.84	15.72	15.71
$\mathbf{T}_{\text{ICA}}$	6.88	11.63	5.52	11.47	5.36	11.62	8.74
$\mathbf{T}^\star_{\text{ICA}}$	16.50	13.75	5.51	16.03	17.83	13.95	13.92
2-D DCT	18.25	15.37	7.17	17.50	19.87	15.61	15.62

Table 4: Generalized coding gain (in dB) of the KLT, $\mathbf{T}_{\text{orth}}$, $\mathbf{T}_{\text{opt}}$, KLT*, $\mathbf{T}^\star_{\text{orth}}$, $\mathbf{T}^\star_{\text{opt}}$, $\mathbf{T}_{\text{ICA}}$, $\mathbf{T}^\star_{\text{ICA}}$ and 2-D DCT for each test image. Last column yields the average generalized coding gain of each transform computed over the set of test images.

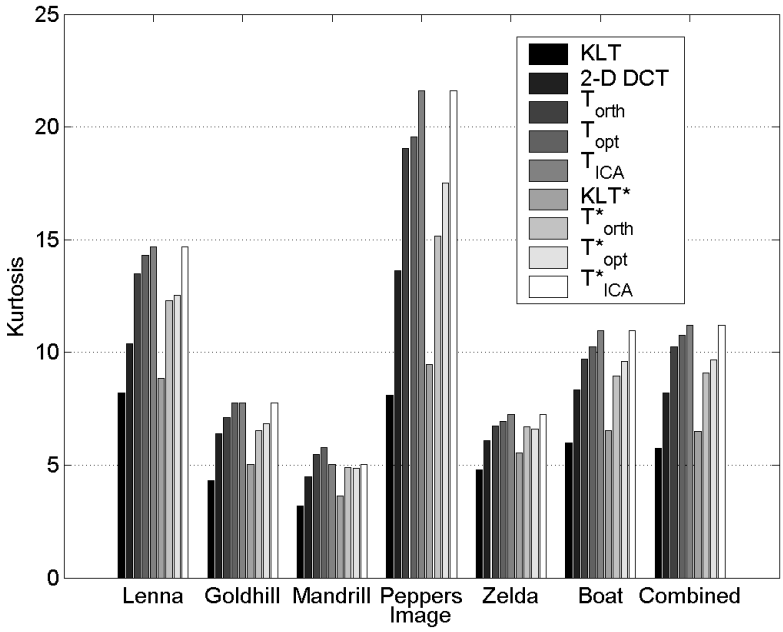
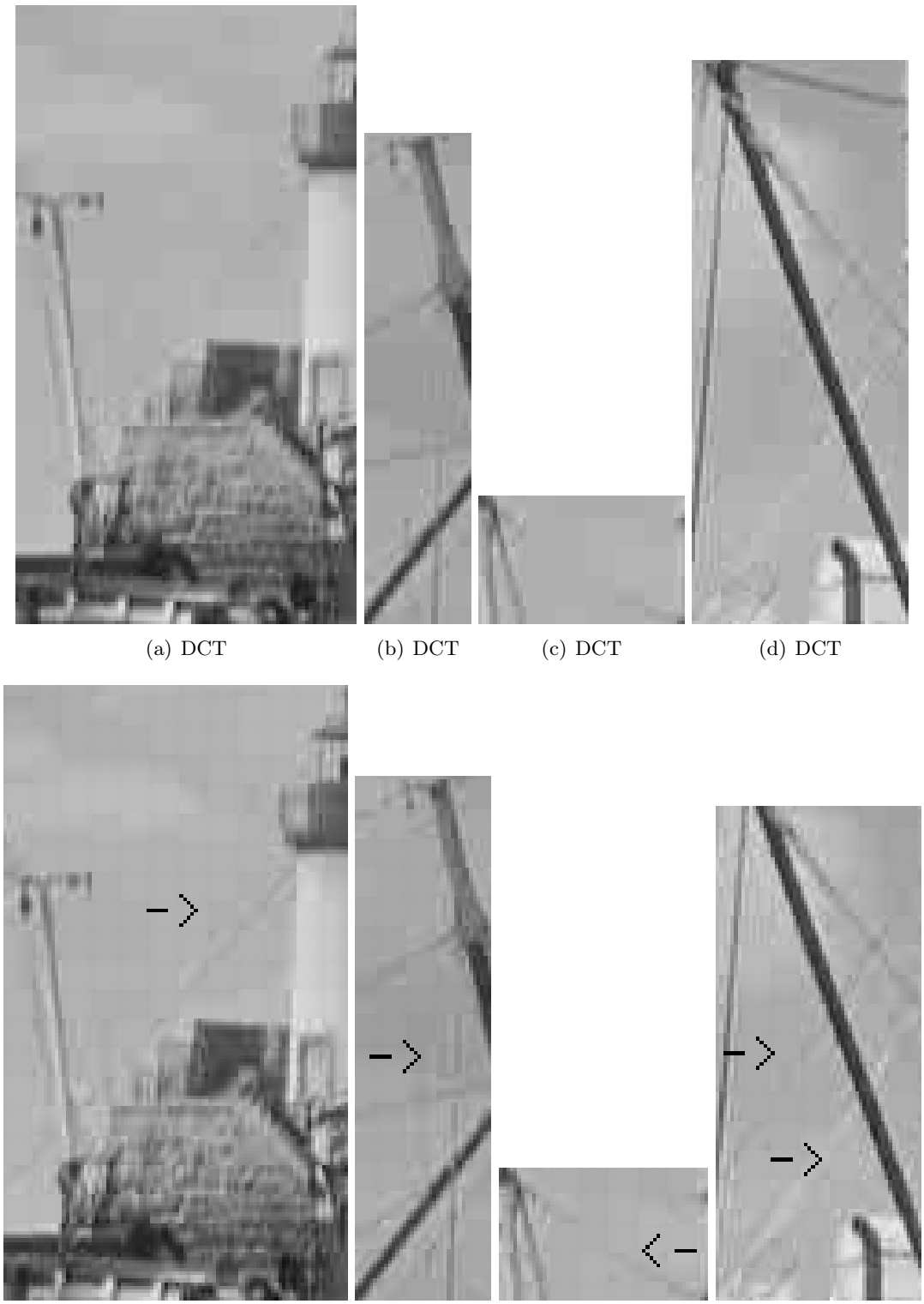


Figure 6: Average kurtosis computed over 63 transformed components (the component which is equivalent to the DC component of the KLT was omitted).

Image	Target bit-rate (bpp)	KLT	$\mathbf{T}_{\text{opt}}$	$\mathbf{T}_{\text{orth}}$	KLT*	2-D DCT	$\mathbf{T}^\star_{\text{opt}}$	$\mathbf{T}^\star_{\text{orth}}$
Lenna	2	42,37	42,72	42,76	42,92	43,21	43,30	43,40
	1	37,32	37,77	37,82	37,95	38,44	38,54	38,55
	0,5	32,56	32,86	33,05	34,31	34,83	34,95	35,00
	0,25	26,15	25,34	26,40	30,72	31,18	31,30	31,32
Goldhill	2	39,45	39,77	39,82	40,17	40,80	40,50	40,48
	1	34,06	34,29	34,34	34,93	35,41	35,24	35,22
	0,5	30,03	30,13	30,34	31,50	31,92	31,85	31,81
	0,25	25,57	24,76	25,66	28,79	29,20	29,15	29,09
Mandrill	2	32,23	32,79	32,69	33,20	33,55	33,48	33,43
	1	26,55	27,03	27,02	27,56	27,87	27,81	27,81
	0,5	23,02	23,21	23,22	24,31	24,49	24,48	24,48
	0,25	20,29	20,37	20,35	22,14	22,34	22,30	22,29
Peppers	2	40,43	40,86	40,90	41,13	41,49	41,70	41,68
	1	35,71	36,11	36,15	36,36	36,71	36,84	36,79
	0,5	31,85	32,23	32,27	33,44	33,91	34,09	33,99
	0,25	25,53	24,24	24,18	30,18	30,74	30,88	30,87
Zelda	2	45,21	45,41	45,39	45,13	45,85	45,86	45,96
	1	40,16	40,30	40,33	40,61	40,94	41,02	41,01
	0,5	35,97	36,31	36,28	37,47	37,75	37,93	37,90
	0,25	29,09	29,14	29,17	34,03	34,49	34,57	34,60
Boat	2	41,67	42,15	42,25	42,04	43,27	42,75	42,87
	1	35,39	35,85	36,09	36,25	37,51	37,14	37,13
	0,5	30,16	30,36	30,65	31,91	32,88	32,73	32,62
	0,25	24,37	23,88	24,56	28,29	29,23	29,09	29,01

Table 5: PSNR (dB) versus target bit-rate for each test image and each transform.



(a) DCT (b) DCT (c) DCT (d) DCT

(e) $\mathbf{T}^{\star}_{\text{orth}}$ (f) $\mathbf{T}^{\star}_{\text{orth}}$ (g) $\mathbf{T}^{\star}_{\text{orth}}$ (h) $\mathbf{T}^{\star}_{\text{orth}}$

Figure 7: Zoom on parts of image *Boat* coded at about 0.5 bpp. Black arrows point towards details which are not present or blurred on the corresponding image coded with the 2-D DCT.