MODEL-BASED SPARSITY PROJECTION PURSUIT FOR LATTICE VECTOR QUANTIZATION

L. Fonteles, M. Antonini

Laboratoire I3S, UNSA - CNRS
Bât. Euclide B - B.P. 121
2000 Route des Lucioles
06903 Sophia Antipolis Cedex, France

R. Phlypo

MEDISIP - IbiTech, UGent - IBBT
Block B, 185 De Pintelaan
9000 Ghent, Belgium

ABSTRACT

In this work we present an efficient coding scheme suitable for lossy image compression using a lattice vector quantizer (LVQ) based on statistically independent data projections. The independence of these components guarantees the optimality of the quantizer. However, this introduces an overload in coding since the projection matrix rendering the components independent needs to be transmitted to the decoder. This issue is tackled by modeling the data such that the projection matrix can be recovered at the decoder side based solely on the model parameters. The original data can thus be recovered based on a reduced descriptive data model and the statistically independent components. Results show that the coding of independent components with a lattice vector quantizer is highly efficient compared with scalar or simple LVQ. Furthermore, the independent data obtained by a model-based projection shows better efficiency without the penalizing coding load of the projection matrix.

***Index Terms*—** Image compression, Lattice Vector Quantization (LVQ), Independent Component Analysis (ICA), product code, data modeling.

1. INTRODUCTION

Vector quantization (VQ) is known to have the potential to achieve the optimal theoretical performance if the vector dimension is arbitrarily high. Unfortunately, the computational complexity of optimal unstructured VQs such as LBG [1] increases exponentially with dimension. In addition, the storage requirements can be very large. One solution to overcome this problem of dimensionality is to use some constrained VQ such as lattice vector quantization (LVQ) [2].

The LVQ approach leads to the design of a structured dictionary whose codevectors are regularly distributed in the space. Therefore, instead of optimizing the position of the vectors in space, one can fit the source by indexing the lattice vectors according to the shape of its distribution. For most of the real data sources this can be done in an efficient way using a product code [3, 4, 5], leading to an optimal rate-distortion trade-off for symmetric unimodal source distributions [3]. Indeed, one can interpret such distributions as a set of concentric hypersurfaces with the same shape depending on the source distribution.

Now, we can index the lattice codewords by assigning a first index (prefix) corresponding to the norm (radius) of the respective surface and by assigning a unique index (suffix) corresponding to the enumeration of the vectors belonging to a same surface. Figure 1 illustrate an example for a Laplacian source with a $\mathbb{Z}^2$ lattice codebook.

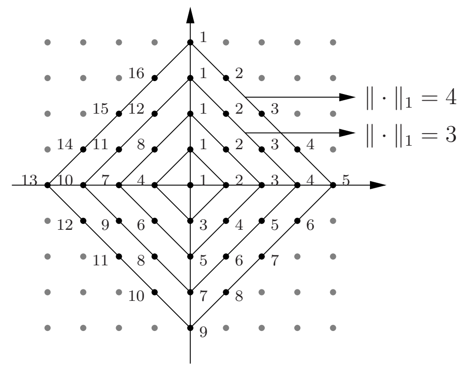

Fig. 1. Example of a product coding scheme for a 2-dimensional Laplacian source with a $\mathbb{Z}^2$ lattice codebook.

In the context of multiresolution image coding, the subband coefficients obtained from wavelet decomposition can be modeled by the generalized Gaussian distributions. This family of distributions is parameterized by a single shape factor p (GG(p)) for a univariate stochastic variable. An interesting property of the GG(p) distributions is that shells of l_p-norm correspond to surfaces of constant probability. This allows for the development of effective product codes [6].

Nevertheless, one may note that the performance of the LVQ associated with the product code strongly depends on how well the shape of the source is fitted. In other words, any mismatch of the data support with respect to the dictionary indexing model degrades the performance in the sense of the rate-distortion trade-off.

In the context of wavelet transformation for image data, the coefficients resulting from the transform are decorrelated but not completely independent, implying that the support does not match the support of an independent GG(p) source. Therefore, the performance of such a coding scheme will be compromised even using indexing techniques capable of indexing GG sources for any shape parameter p [6, 7].

In order to overcome this problem, in this work we propose to preprocess the data vectors such that their support coincides as close as possible to the one of a GG(p) distribution. Hereto we propose to use the technique of Independent Component Analysis (ICA) which renders the vector data maximally mutually independent and as such transforms the support of the data into independent generalized Gaussians. We will show that the ICA algorithm can be data

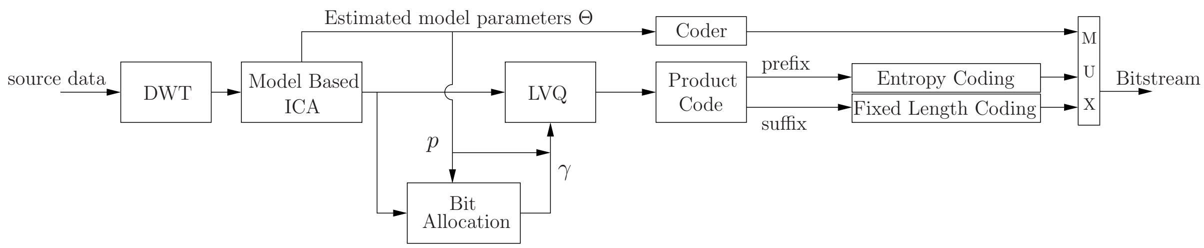

Fig. 2. Global coding scheme. p represents the shape factor of the generalized Gaussian distribution.

driven or model-based. For the latter case we gain in bit-rate *iff* the image data (or wavelet projections) can be characterized with a certain precision by a model-based on only few parameters.

This paper will first briefly discuss the algorithms for LVQ and ICA and will then show that by choosing an appropriate model for the vector data the coding gain of a classical LVQ can easily be increased (Section 3). In the results section we will focus on the rate performance trade-off when the data model covers the data (im)perfectly and subsequent discussion will be developed at Section 5. In Section 6 we conclude the paper and give some perspectives for the future works.

2. LATTICE VECTOR QUANTIZATION

A lattice Λ in $\mathbb{R}^n$ is composed of all integer combinations of a set of linearly independent vectors $\mathbf{a}_i$ (the basis of the lattice) such that:

$$\Lambda = \{\mathbf{x} | \mathbf{x} = u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + ...u_n\mathbf{a}_n\} \quad (1)$$

where the u_i are integers. The partition of the space is hence regular and depends only on the chosen basis vectors $\mathbf{a_i} \in \mathbb{R}^m$ $(m \geq n)$. Note that each set of basis vectors defines a different lattice.

Each vector **v** of a lattice can now be considered as belonging to a surface or hyper-surface containing vectors with constant l_p norm given by:

$$\|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}} .$$

Under these conditions it is possible to encode a given lattice vector using product code. Clearly, if the distribution of the source vectors is Laplacian a good product code consists of a prefix corresponding to the l_1 norm of a vector and a suffix corresponding to its position on the hyper-pyramid[1] with radius equal to the considered l_1 norm. The position of the vector on a hyper-surface can be found using an enumeration algorithm [2, 3, 4, 6]. Moreover, this ensures the uniqueness for decoding.

In the case of sources with generalized gaussian distributions with a shape parameter less than or equal to one, the superiority of the cubic $\mathbb{Z}^n$ lattice over D_4, E_8 or leech lattices has been established [8] and hence we use in this framework the $\mathbb{Z}^n$ LVQ along with product code [6].

LVQ and product code scheme has optimal performance when the vectors to code are independently distributed [3]. However, as discussed before, wavelet transform does not guarantee a complete independence of the coefficients. In order to increase the sparsity of the image vectors and thus the coding gain, we introduce a sparsity pursuit algorithm projecting the wavelet coefficients onto a new basis where maximum sparsity is obtained under some well known preconditions. The algorithm uses a model-based ICA and is presented in the next section.

The overall coding scheme presented in Figure 2 summarizes the coding framework used in this paper.

3. MODEL-BASED INDEPENDENT COMPONENT ANALYSIS

3.1. Principle of ICA

Consider the stochastic vectors $\mathbf{y} \in \mathbb{R}^N$ and $\mathbf{x} \in \mathbb{R}^M$, where each x_i is independently distributed with respect to $x_{j \neq i}$. We can then state that there exists a transform of variables which make the variables in $\mathbf{y}$ take the properties of the vector $\mathbf{x}$ after transformation. This is a general problem in Blind Source Separation (BSS) and can be solved by Independent Component Analysis (ICA) algorithms. The BSS model considered here is the generative linear mixing model

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \eta \ , \quad (2)$$

where the observed values in $\mathbf{y}$ are believed to be generated by the statistically independent sources $\mathbf{x}$ through a linear mixing $\mathbf{F}$ up to some modelling inaccuracies or noise η. For simplicity, we will consider here only the noiseless case (or perfect modelling) $\eta = \mathbf{0}$, an equal number of variables $(N = M)$ in $\mathbf{y}$ and $\mathbf{x}$ and a reversible mixing matrix $\mathbf{F}$. Whereas the purpose of BSS is generally the reconstruction of the source signals $\mathbf{x}$ by searching an appropriate basis to project the data on, i.e.

$$\hat{\mathbf{x}} = \hat{\mathbf{W}}^T\mathbf{y} \ ,$$

with $\hat{\mathbf{W}}^T$ an estimate of $\mathbf{F}^{-1}$ as in Eq. 2, here we are solely interested in a projection onto a basis which makes the data vectors as independent as possible, optimising the vectors for the subsequent LVQ coding stage. This proves useful, since a lot of attempts are reported in literature to describe natural images by their statistics [11, 12] but as far as we know these statistical models have never been exploited in an ICA framework to gain bandwidth in image coding.

The envisaged statistical independence is defined as the decorrelation of a signal, through all possible function images, i.e.

$$\phi_i(x_i)\,\phi_j(x_j) = \delta_{ij} \ ,$$

with ϕ_i and ϕ_j both functions with their argument and image in $\mathbb{R}$ and δ_{ij} is 1 *iff* $i = j$, 0 otherwise. Independence can thus be seen as a stronger condition than simple decorrelation and as can be seen from Figure 3 it ensures that the projection of x_k onto the k-th canonical basis vector results simply in its marginal distribution. It has been shown that there is a report between these mutual

[1] The hyper-surfaces of constant l_1 norm are called hyper-pyramids

independence (which induces independence of the total set) and negentropy, or the differential entropy of a variable with respect to a Gaussian sharing the same variance. It has also been shown that the Edgeworth expansion of this negentropy can be rewritten as a function of the marginal statistics of the filter output $\hat{\mathbf{x}}$, under minor preassumptions [13]. When truncating this expansion to exclude negligible contributions, it proves sufficient to consider statistics up to fourth order only [14]. Furthermore, we can omit all contributions of the third order cumulants since they vanish for symmetric distributions, which is approximately the case when considering outputs of a wavelet decomposition.

The ICA algorithm considered here is such that the unmixing matrix $\hat{\mathbf{W}}^T$ can be derived from some parameter set inherent to the data, in particular their second and fourth order statistics represented by their cumulants. More specifically, the fourth order cumulants of a stochastic vector $\mathbf{a}$ are given by

$$\begin{aligned} Cum\,(a_i a_j a_k a_l) &= \mathrm{E}\{a_i a_j a_k a_l\} - \mathrm{E}\{a_i a_j\}\,\mathrm{E}\{a_k a_l\} \\ &\quad - \mathrm{E}\{a_i a_k\}\,\mathrm{E}\{a_j a_l\} - \mathrm{E}\{a_i a_k\}\,\mathrm{E}\{a_j a_l\} \quad . \end{aligned} \tag{3}$$

Suppose now that we take a datavector with unit variance and zero mean. As a consequence the first order cumulants (the means) vanish and the second order cumulants (the variances) will all equal 1. This does not harm the generality since any vector $\mathbf{y}$ can linearly be transformed by taking $\mathbf{z} = \mathbf{V}(\mathbf{y} - \mu_{\mathbf{y}})$, where $\mathbf{V}$ is the projection matrix which renders $\mathbf{V}(\mathbf{y} - \mu_{\mathbf{y}})$ uncorrelated and unit variance. This can be achieved by taking $\mathbf{V}$ as the inverse of the covariance matrix of $\mathbf{y}$ or through the numerically more stable Singular Value Decomposition of $\mathbf{y} - \mu_{\mathbf{y}}$. From hereon we will consider $\mathbf{y}$ as a zero mean variable ($\mu_{\mathbf{y}} = 0$) and $\mathbf{z}$ as a vector of zero mean uncorrelated unit variance random variables (Figure 3 left and middle insert). The projection of $\mathbf{y}$ onto $\mathbf{z}$ is generally known as prewhitening. Independence of such a prewhitened vector can now be obtained by applying a rotation to the variable $\mathbf{z}$ i.e. $\mathbf{Q}\hat{\mathbf{x}} = \mathbf{z}$, where $\mathbf{Q}$ is orthonormal. Maximum independence of the set $\hat{\mathbf{x}}$ can then be obtained by minimizing the sum of all cross-cumulants (alleviating the statistical dependencies at order 4) of $\hat{\mathbf{x}}$ over the space of rotation matrices in order to maintain the unit variance, zero-mean and decorrelation obtained in the prewhitening step:

$$\Psi(\hat{\mathbf{x}}) = \sum_{\substack{i,j,k,l=1 \\ \neg i=j=k=l}}^{N} |Cum(\hat{x}_i \hat{x}_j \hat{x}_k \hat{x}_l)| = -\sum_{i=1}^{N} |Cum(\hat{x}_i^4)| \ , \tag{4}$$

where $\hat{\mathbf{x}}$ is the result of a rotation $\mathbf{Q}$ applied to $\mathbf{z}$. The latter equality in Eq. 4 follows from the invariance of the cumulants under rotation. We thus have the following equalities in the system:

$$\hat{\mathbf{x}} = \mathbf{Q}\mathbf{z} = \mathbf{Q}\mathbf{V}\mathbf{y} = \hat{\mathbf{W}}^T(\mathbf{y} - \mu_{\mathbf{y}}) \ . \tag{5}$$

Consequently, the solution $\mathbf{Q}^\star$ rendering the transformed variables $\hat{\mathbf{x}}$ as independent as possible is now given as

$$\mathbf{Q}^\star = \arg\max_{\mathbf{Q}} \Psi\left(\hat{\mathbf{x}}\right) = \arg\max_{\mathbf{Q}} \Psi\left(\mathbf{Q}\mathbf{V}(\mathbf{y} - \mu_{\mathbf{y}})\right) \ .$$

3.2. Introduction of the model

Since the cost of sending the separator matrix $\hat{\mathbf{W}}^T$ or its inverse $\hat{F}$ has a far too high cost when coding, since its structure is suboptimal for the LVQ (a non optimized version occupies N non sparse vectors in $\mathbb{R}^N$), there is a need to model the projection matrix by reducing it to a parameter set. Instead of pursuing a model to directly parameterize the matrix $\hat{\mathbf{W}}^T$, we opt for a model of the random variables $\mathbf{y}$ by reducing them to a parameter set $\mathbf{\Theta}$. $\mathbf{\Theta}$ should now be chosen as such that the second and fourth order statistics of the random vector $\mathbf{y}$ can be deduced from it, since these statistics have proven sufficient to solve for the set of independent generators, see Section 3.1.

In next section highlight the impact of the good choice of the model when preprocessing dependent sources.

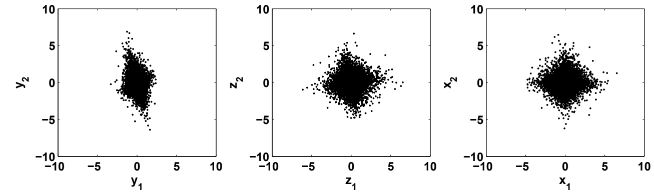

Fig. 3. From left to right: pair of original variables, decorrelated variables and independent variables.

4. EXAMPLE OF AUTOREGRESSIVE MODEL

4.1. Model design

As an example of a generalized gaussian correlated source we assume that the sources to encode are based on a first order autoregressive model for images and image sequences [11]. For this model the fourth and second order cumulants can be derived from a parameter set $\mathbf{\Theta} = \{\rho, \sigma_y, p\}$ as will be explained below. The regression model AR(1) is given as:

$$y_i(k) = \rho y_{i-1}(k) + u_i(k)^{(p)} \ , \tag{6}$$

where each variable $u_i^{(p)}$ is independently distributed following a generalized gaussian distribution GG(p) and $y_j(m)$ is the m-th realization of the j-th stochastic variable. To compute the matrices $\mathbf{Q}$ and $\mathbf{V}$, we need to retrieve the second and the fourth order cumulants of $\mathbf{z}$. These can be derived from the model in Eq. 6 as (see also [11][2]):

$$\begin{aligned} Cum\,(y_i y_j) &= \rho^{|i-j|}\frac{\sigma_u^2}{1-\rho^2} = \rho^{|i-j|}\sigma_y^2 \\ Cum\,(z_i z_j z_k z_l) &= \rho^{|i-m|+|j-m|+|k-m|+|l-m|}\frac{\kappa_u}{1-\rho^4} \end{aligned} \tag{7}$$

where $m = \min(i,j,k,l)$ and κ_u, σ_u^2 are the fourth and second order autocumulants of u (all u_i are i.i.d. as explained above) and σ_y^2 is the second order marginal cumulant of y. With the parameter set $\mathbf{\Theta} = \{\rho, \sigma_y, p\}$ we have thus all the information that is needed to create the independent components to submit to the LVQ ($\hat{\mathbf{x}} = \hat{\mathbf{W}}^T(\Theta)\mathbf{y}$) and to reconstruct the matrix $\hat{\mathbf{W}}(\mathbf{\Theta})^T$ and thus its inverse together with the original data at the decoder side by

$$\tilde{\mathbf{y}} = (\hat{\mathbf{W}}(\mathbf{\Theta})^T)^{-1}\langle\hat{\mathbf{x}}\rangle \ ,$$

where $\langle\rangle$ denotes the quantization introduced by the LVQ and $\tilde{\mathbf{y}}$ is the reconstruction of the original up to its errors introduced by quantization only. Remark that a mismatch of the model and the data do not introduce any errors, whereas it does have an influence on the efficiency of the LVQ coding.

[2] Only moments were considered in this publication, however an extension to cumulants is straightforward and due to a lack of space we here only give the results.

4.2. Experimental Results

To place the combined ICA-LVQ coder in perspective, the rate-distortion performance will be given with respect to a standard scalar coding scheme and a data-driven ICA preprocessing as described in Section 3.1. For this purpose we choose a source that is artificially generated along the model in Section 3.2 with parameter set $\mathbf{\Theta} = \{0.8, 1, 0.7\}$. The results are given in Figure 4.

The first tested scheme in the simulations joins a uniform scalar quantization without ICA pre-processing and a simple arithmetic coder (SQ). A second scheme uses the $\mathbb{Z}^{64}$ lattice vector quantizer in a product coder framework without ICA (LVQ). A third result combines $\mathbb{Z}^{64}$ LVQ with a non-parametric ICA [14] without the bit cost of the mixing matrix $\hat{\mathbf{W}}$ (LVQ+ICA). Another scheme includes the bit cost (arithmetic coding) of the mixing matrix $\hat{\mathbf{W}}$ (LVQ+ICA+W). The last one applies the proposed scheme, using the $\mathbb{Z}^{64}$ LVQ associated with the model-based ICA with the cost of the vector of model parameters $\mathbf{\Theta}$ estimated from the correlated source (LVQ+MBICA+ $\mathbf{\Theta}$). Results are compared to the well-known bit plane coder of JPEG2000 (kakadu software used).

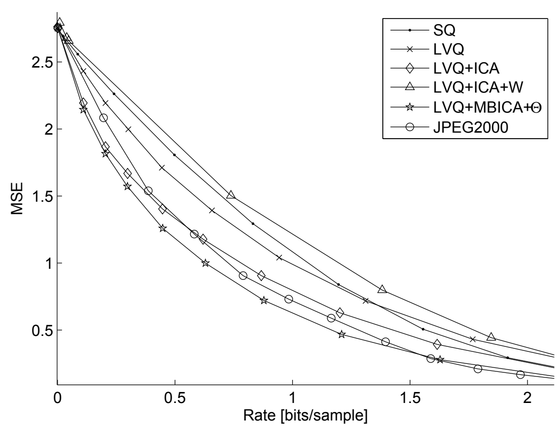

Fig. 4. Rate-distortion results for a synthetic source with $\mathbf{\Theta} = \{0.8, 1, 0.7\}$ and vector dimension 64.

5. DISCUSSION

From Figure 4 one can note the rate-distortion performance improvement by coding the source by LVQ ($\mathbb{Z}^{64}$) with respect to ordinary SQ scheme. Moreover, an additional gain over LVQ coding is obtained by introducing an appropriate preprocessing, rendering our data in concordance with the coding dictionary. However, the performance gain introduced by the LVQ+ICA scheme is deflated by the coding effort needed to send $\hat{\mathbf{W}}^T$ through the decoder, leading to the result illustrated by the curve LVQ+ICA+W.

For this reasons we introduce a preprocessing based on a parameterized ICA model. Here, the cost of sending the mixing matrix $\hat{\mathbf{W}}^T$ is replaced by the cost of sending the vectors of parameters $\mathbf{\Theta}$, which is usually negligible with respect to the coding cost of the independent data. The parameter vector $\mathbf{\Theta}$ is able to describe the cumulants of the correlated source, and thus, it allows us to estimate $\hat{\mathbf{W}}^T$. The recuperated $\hat{\mathbf{W}}(\mathbf{\Theta})^T$ seems a good approximation to $\hat{\mathbf{W}}^T$ and is able to project the data on a good support, within reasonable distance to the dictionary, as prove the results.

6. CONCLUSIONS AND FUTURE WORK

We introduce the use of a model-based ICA as preprocessing algorithm for LVQ/product code scheme. We have shown that if the model is accurate enough and if it has a fairly small amount of parameters, one can improve the classical performances of LVQ coder by better adapting the support of the source to the indexing algorithm. If the model holds for the data the method acts as a Maximum Likelihood separator and thus does not suffer from the limited number of samples that are taken as realizations of the stochastic data vector. This creates an opportunity to sample the data into higher dimensional vectors and thus to improve the coding gain.

As future works we will investigate the use of different models on the context of image compression in order to improve the results presented in [7].

7. REFERENCES

[1] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, january 1980.

[2] J.H. Conway and N.J.A. Sloane, *Sphere packings, lattices and groups*, Springer Verlag, 1988.

[3] T. R. Fischer, "A pyramid vector quantizer," *IEEE Trans. Inform. Theory*, vol. 32, pp. 568–583, July 1986.

[4] P. Loyer, J.M. Moureaux, and M. Antonini, "Lattice codebook enumeration for generalized gaussian source," *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 521–528, Feb. 2003.

[5] J.M. Moureaux, P. Loyer, and M. Antonini, "Low complexity indexing method for $\mathbb{Z}^n$ and D_n lattice quantizers," *IEEE Trans. Commun.*, vol. 46, no. 12, pp. 1602–1609, december 1998.

[6] L. H. Fonteles and M. Antonini, "Indexing Z^n lattice vectors for generalized gaussian distributions," in *Proc. of IEEE ISIT*, June 2007.

[7] L. H. Fonteles and M. Antonini, "High dimension lattice vector quantizer design for generalized gaussian distributions," in *Proc. of IEEE ICIP*, September 2007.

[8] Z. Gao, F. Chen, B. Belzer, and J. Villasenor, "A comparison of the Z^n, E8, and leech lattices for quantization of low-shape-parameter generalized gaussian sources," *IEEE Signal Processing Letters*, vol. 2, pp. 197–199, October 1995.

[9] Wolfgang Niehsen, "Generalized gaussian modeling of correlated signal sources," *IEEE Trans on Sign Proc*, vol. 47, no. 1, pp. 217–9, 1999.

[10] BA Olshausen and DJ Field, "Natural images statistics and efficient coding," in *proc of the Workshop on Information Theory and the Brain*, Scotland, 4-5 September 1995.

[11] J.F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, jan 1999.

[12] Pierre Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.