Spatio-Chromatic Decorrelation for Color Image Compression

Mark S. Drew, [a],[*],[1] Steven Bergner [a]

[a]*School of Computing Science, Simon Fraser University, Vancouver, Canada V5A 1S6*

Abstract

We investigate the implications of a unified spatio-chromatic basis for image compression and reconstruction. Different adaptive and general methods (PCA, ICA, and DCT) are applied to generate bases. While typically such bases with spatial extent are investigated in terms of their correspondence to human visual perception, we are interested in their applicability to multimedia encoding. The performance of the extracted spatio-chromatic spatial patch bases is evaluated in terms of quality of reconstruction with respect to their potential for data compression. Since independent component analysis is not as widely used as it should be, compared to the other decorrelation methods applied here in a new domain, we also provide a review of ICA. The results discussed here are intended to provide another path towards perceptually-based encoding of visual data. This leads to a deeper understanding of the role played by chromatic features in data reduction.

Key words: Color image compression, spatio-chromatic bases, PCA, ICA.

1 Introduction

Decorrelation for redundancy reduction has a long history in image processing. In particular, variants of the Principal Component Analysis (PCA) [1] for orthogonal decorrelation have been part

* Corresponding author.
  Email addresses: `mark@cs.sfu.ca` (Mark S. Drew,), `sbergner@cs.sfu.ca` (Steven Bergner).
[1] Mailing address: School of Computing Science,
Simon Fraser University
8888 University Drive,
Vancouver, B.C. CANADA V5A1S6

Tel: 1-778-782-4277
Fax: 1-778-782-3045

of the arsenal of data reduction for many years. The main idea here is to account for most of the variance in the data using the first several principal axes, and then reduce the influence of further terms either by directly omitting these or by adopting a bit allocation scheme to deprecate their influence.

PCA can tell us how Nature processes vision, if we consider natural images. In particular, we expect to see color-opponent channels arise in a natural fashion, simply by automatic inspection of the data. But as well, we hope to glean evidence of how spatial processing operates. And in fact, in terms of human visual system quantum catches Ruderman et al. [2] found not only such color-opponent structures but also spatial derivative-like filters, operating similarly and independently in each opponent-color and luminance channel.

Ruderman analyzed hyperspectral image sequences of natural scenes, looking for a decorrelated color-space basis. In part, Ruderman et al. built on the work of Webster et al. [3] in arguing for a representation involving a log LMS (long-medium-short) space.

Their work was based on a very simple scheme: first, using natural images derived from measured hyperspectral data and transformed to log space, Ruderman et al. formed small tiles, 3 pixels by 3 pixels, from an assembly of such images. Treating these 9 pixels as a vector, with each pixel containing 3-vector color information, leads to 27-vector data for PCA analysis. Re-assembling the color information, these color patches could also be reconstituted as 3×3 color squares, for viewing. The Principal Component Analysis then produced 27 such 3×3 color checkerboards, ranked in order of singular values from most influential to least important. The answer that arose was that, for the foliage data that was used, these color squares naturally grouped into gray luminance, a blue-yellow axis, and a red-green axis, in that order.

In terms of spatial components, for each of the color channels the spatial structure of the bases resembled the derivative-like and frequency-analysis-like structures arising in a Fourier analysis of grayscale images.

The latter result was not surprising (although the decorrelation from color was): Olshausen and Field's seminal work on receptive field properties [4] implied that the receptive fields in mammalian primary visual cortex simple cells are spatially localized, oriented, and spatially bandpass in the sense of being selective to structure at different spatial scales, for non-color luminance inputs. Visually, these fields resemble a 2-dimensional Discrete Cosine Transform (DCT) basis in an $N \times N$ checkerboard structure (see, e.g., [5], and below), but with diagonal as well as axis-aligned basis images.

PCA has also been applied to non-interpolated, raw Color Filter Array (CFA) data [6], with the result that the recovered basis finds only color information, not luminance information, so is not appropriate for modeling spatial information.

Here we are interested in considering in detail how the use of Independent Component Analysis (ICA) as opposed to PCA affects color still image compression based on such color checkerboard basis image blocks. In the next section, we explicate the ICA approach.

2 Related Work

2.1 ICA for Signal Separation

As opposed to an orthogonal PCA basis, some workers have also considered an Independent Component Analysis (ICA) of natural images [7]. ICA proceeds by producing a minimally redundant set of basis functions. To do so, a set of maximally statistically *independent* basis vectors is found.

To understand what this means let us consider two stochastic variables x_1 and x_2, and also the coefficients s_1, s_2 of our data *projected on two different axes*. Their joint probability density is $p(x_1, x_2)$. The separate probability densities for x_1 or x_2, their marginal probability, can be computed e.g. by $p_1(x_1) = \int p(x_1, x_2) dx_2$. Then x_1 and x_2 are independent if and only if in the new coordinate system, where the new variables are s_1, s_2, we have

$$p(s_1, s_2) = p_1(s_1) p_2(s_2) \tag{1}$$

This leads to the condition

$$E\{h_1(s_1) h_2(s_2)\} = E\{h_1(s_1)\} E\{h_2(s_2)\}, \tag{2}$$

where E means expectation and h_1, h_2 can be essentially any two integrable scalar functions of the s; the above is therefore a very strong condition on the distributions of s_1 and s_2. It says that any nonlinear transforms of the independent components are uncorrelated — the covariance between different independent components is zero. In comparison, PCA decorrelates but does not guarantee independence. That is, projecting the data to the decorrelated axes, the distribution of two resulting coefficients s_1 and s_2 fulfill

$$
\begin{aligned}
cov(s_1, s_2) &= E\{(s_1 - \bar{s}_1) \cdot (s_2 - \bar{s}_2)\} \\
&= E\{s_1 \cdot s_2\} - E\{s_1\} E\{s_2\} \\
&= 0
\end{aligned}
\tag{3}
$$

This corresponds to eq.(2) with h_1, h_2 being linear and therefore constitutes a much weaker statement than does eq. (2). However, PCA ("whitening") is still useful as a pre-processing step for ICA, and we use it that way here. The vocabulary used for ICA is somewhat different than that used for PCA. ICA is one way of solving the Blind Source Separation problem. Just as for PCA, if we have a k-dimensional feature-vector $\boldsymbol{x}$, then if there are k basis vectors we can approximate $\boldsymbol{x}$ as

$$\boldsymbol{x} \simeq \boldsymbol{A} \, \boldsymbol{s} \tag{4}$$

where the k columns of $\boldsymbol{A}$ hold our ICA basis vectors, and $\boldsymbol{s}$ is the set of weights. In ICA, the matrix $\boldsymbol{A}$ is called the *mixing matrix* and $\boldsymbol{s}$ is a multidimensional stochastic variable of independent *sources*. E.g., in PCA, vector $\boldsymbol{x}$ could be a grayscale image, and $\boldsymbol{A}$ would consist of orthogonal eigenimages (cf. [8]). However, for ICA the columns of $\boldsymbol{A}$ are not orthogonal. Therefore to find the weights $\boldsymbol{s}$ we must use the Moore-Penrose pseudoinverse $\boldsymbol{A}^+$ of $\boldsymbol{A}$, applied to the target image.
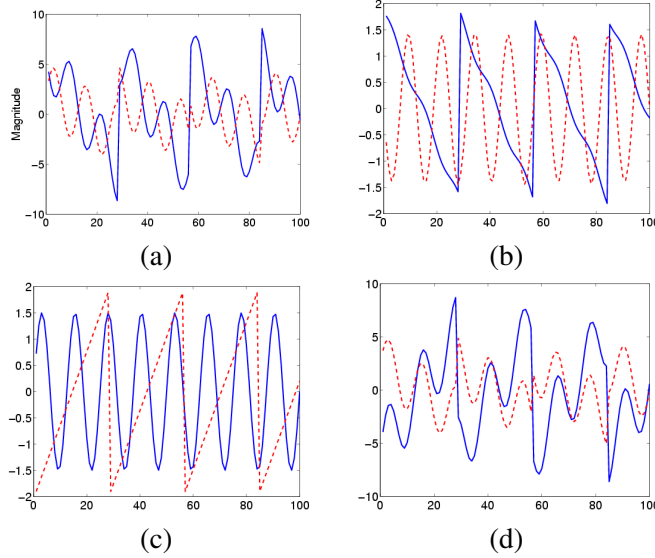
Fig. 1. (a): Mixed signals as measured. (b): Source signals as recovered. (c): Actual source signals. (d): PCA basis.

The pseudoinverse is referred to as the set of ICA *filters*. Matrix $\boldsymbol{A}$ is called the mixing matrix since it produces signals mixed from the sources; whereas $\boldsymbol{W} = \boldsymbol{A}^+$ is called the *separating matrix*:

$$\boldsymbol{s} = \boldsymbol{W}\boldsymbol{x} = \boldsymbol{A}^+\boldsymbol{x} \qquad (5)$$

ICA has been used for multimedia data fusion as well [9,10] — in this case, ICA recovers common latent subspaces for combined media. Typically, however, PCA is used [11,12], applied to much smaller feature spaces than whole images.

The process of finding these independent ICA vectors is based on the Central Limit Theorem, which states that a sum of non-Gaussian random variables is more like a Gaussian than are its individual components. But the independent sources sought can be written as sums of the observed data. Thus, we can move toward independent sources by trying to find a sum of the observed data over vectors which have maximum non-Gaussianity. This property can be measured in terms of higher order statistics, e.g. kurtosis or negentropy (see below).

ICA can be viewed as a linear generative model with non-Gaussian priors for the hidden variables.

To see the power of ICA, consider for the moment 1D signals (these could be audio signals, for example, after [13]). Suppose two audio sources generate signals which are then measured by two microphones placed at random locations near the sources: in a linear model, each microphone generates a signal given by a linear combination of the two original *source* signals. Since the sources do not in principle affect each other in any way, they are *independent*. Recovery of the original sources is a classical problem in blind source separation; it is termed the "cocktail party problem".

Suppose the two signals $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ as measured are as shown in Fig. 1(a). Comparing the actual (independent) sources, in Fig. 1(c), with the ICA recovered versions, in Fig. 1(b), we note that ICA

does not necessarily produce the correct order or sign for the recovered sources — but this will turn out not to matter.

Clearly, if we can indeed identify correct sources in a set of measured signals, then we can do a better job in compression by allocating more bits to the actual sources, and fewer to vectors deemed not truly independent. If we have k-dimensional data, and can obtain an expressive, sparse set of underlying mechanisms for image generation, then coefficients in terms of this underlying set will effectively be of reduced rank. Therefore we can describe the image data well with a small basis, and can as well assign bits from a bit-budget as needed to the surviving coefficients. After entropy-based coding, since the entropy is greatly reduced by focusing on the correct, underlying features of the data, we arrive at a substantial compression.

PCA does do a good job of identifying the most-variance to least-variance set of orthogonal directions from a measured dataset, but the underlying data may not indeed actually be orthogonal. Using PCA, however, we are constrained to the orthogonality assumption, yielding results as in Fig. 1(d) that may not be as useful as they could be if the correct, non-orthogonal sources were determined using ICA.

Since the independent sources are modeled as a linear combination of the measured data, we look for an ICA basis such that the nongaussianity of the combination of measured data is maximized. Thus we seek local maxima of nongaussianity of a linear combination

$$s_i \; = \; \sum_{j=1}^{k} w_j \boldsymbol{x}_{\,ji}\,, \;\; i = 1..N \tag{6}$$

where x_{ji} is the observed data, $j = 1..k$, $i = 1..N$ consisting of N observations of k-vectors, under the constraint that the variance is constant. Each local maximum gives one independent component.

The idea in "sparse" coding is to represent data with components such that only a small number of them are "active" at the same time. It turns out that this is equivalent, in most situations, to finding components that are maximally nongaussian. The latter property can be characterized by the kurtosis (the fourth-order cumulant), which is zero for a Gaussian. However, a more robust measure is formed by the negentropy, the difference between the entropy for a Gaussian and that for the current basis, using the observation that for a given standard deviation σ, a Gaussian G_σ has maximum entropy compared to other probability distributions.

Thus one arrives at a gradient descent method for determining the ICA basis, and since this can be phrased as a fixed-point problem, mechanisms similar to the Contraction Mapping Theorem can be brought to bear for existence, uniqueness, and convergence rate. The method we use here is the FastICA algorithm [14].

However, a simple explanation using the kurtosis is indeed illuminating: for a zero-mean variable y, the kurtosis is defined as

$$\kappa(s) \; = \; E\{s^4\} \; - \; 3\left(E\{s^2\}\right)^2 \tag{7}$$

and if we re-scale the variance to unity, this reduces to just

$$\kappa(s) = E\{s^4\} - 3 \tag{8}$$

so that clearly we may use just the expectation of the fourth moment of the signal. However, including the -3 means that κ may be negative, so we maximize $|\kappa(s)|$.

An algorithm for finding an independent source combination then proceeds by first "whitening" the data using PCA; i.e., we diagonalize the signal's product-matrix into an orthogonal matrix $\boldsymbol{U}$ and a diagonal one $\boldsymbol{D}$: for a $k \times N$ matrix of signals $\boldsymbol{x}$,

$$\boldsymbol{x}\, \boldsymbol{x}^{\,T} = \boldsymbol{U}\ \boldsymbol{D}\ \boldsymbol{U}^{\,T}, \quad \boldsymbol{U} \text{ and } \boldsymbol{D} \text{ are } k \times k \tag{9}$$

Then in the new coordinate system, $\widetilde{\boldsymbol{x}} = \boldsymbol{x}_{\,whitened}$,

$$\boldsymbol{x}_{\,whitened} = \widetilde{\boldsymbol{x}} \equiv \sqrt{(\boldsymbol{x}\,\boldsymbol{x}^{\,T})^{-1}}\,\boldsymbol{x} = (\boldsymbol{U}\ \boldsymbol{D}^{\,-1/2}\boldsymbol{U}^{\,T})\,\boldsymbol{x} \equiv \boldsymbol{M}\,\boldsymbol{x} \tag{10}$$

so now $\widetilde{\boldsymbol{x}}\,\widetilde{\boldsymbol{x}}^{\,T} = \boldsymbol{I}$. Hence we only need to search for a new, *orthogonal* matrix $\tilde{\boldsymbol{A}}$ such that

$$\widetilde{\boldsymbol{x}} = \widetilde{\boldsymbol{A}}\ \boldsymbol{s} \tag{11}$$

with $\widetilde{\boldsymbol{A}}\,\widetilde{\boldsymbol{A}}^{\,T} = \boldsymbol{I}$. Once $\widetilde{\boldsymbol{A}}$ is determined, we can go back to a mixing matrix in terms of the original, unwhitened $\boldsymbol{x}$ via $\boldsymbol{W} = \widetilde{\boldsymbol{A}}^{\,+}\boldsymbol{M}$.

Thus we first whiten $\boldsymbol{x}$, and then seek a column $\widetilde{\boldsymbol{w}}$ of an orthogonal matrix $\widetilde{\boldsymbol{W}}$, with $\|\widetilde{\boldsymbol{w}}\| = 1$, that maximizes the departure of the kurtosis from zero:

$$\mathcal{I} = \overset{\max}{\widetilde{\boldsymbol{w}}}\ \left| \frac{1}{N} \sum_{i=1}^{N} \left(\widetilde{\boldsymbol{w}}^{\,T}\widetilde{\boldsymbol{x}}\,[i]\right)^4 - 3 \right|\ \text{where } \widetilde{\boldsymbol{x}}\,[i] \text{ is the } i^{th} \text{ column of } \widetilde{\boldsymbol{x}} \tag{12}$$

This leads to an Euler equation as follows:

$$\frac{\partial \mathcal{I}}{\partial \boldsymbol{w}} = \frac{4}{N} \sum_{i=1}^{N} \left[\left(\widetilde{\boldsymbol{w}}^{\,T}\boldsymbol{x}\,[i]\right)^3 \widetilde{\boldsymbol{x}}\,[i] \right] = 0 \tag{13}$$

Sample code for such an algorithm is given in Appendix A. For a more detailed discussion of ICA the interested reader is referred to [15].

2.2   ICA for Imagery

Data reduction for images using ICA has indeed been found useful for sparse coding, i.e., finding underlying sources such that any given image is naturally represented in terms of just a small number of these: ICA [7] in fact reproduces results for optimizing sparseness [4]. Typically, the technique is used for the extraction of hidden sources generating observed data. For example, consider the image in Fig. 2. The RGB values in this image form clusters, as in shown in the left of
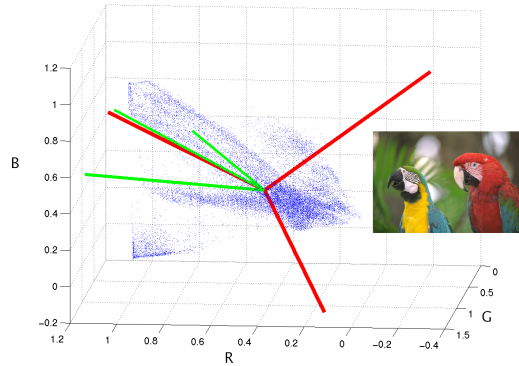
Fig. 2. Basis vectors for a given color distribution from right image as found by PCA (red) and ICA (green).
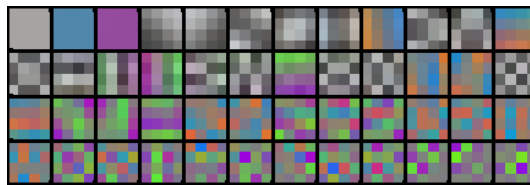

Fig. 3. Spatio-chromatic basis obtained from PCA on $4 \times 4 \times RGB$ image patches of the example in Fig. 2.

Fig. 2, with orthogonal PCA axes shown in red. In contrast, the ICA axes (in green) show that the image is actually comprised of just a few independent sources. It should be noted that ICA is data adaptive: we would like to develop a set of basis vectors that is specific for a certain type of image contents. We could also target at developing a universal basis from a training set, but an adaptive model is bound to be more expressive since, as we see from Fig. 2, the hidden characteristics of content are extracted.

For grayscale imagery [4,7], PCA indicates that a mutually orthogonal spatial basis for imagery consists of bandpass filters similar to 2-dimensional DCT basis images, but with some non-rectangular orientation present (cf. Fig. 3). For grayscale, how one creates such an image is by randomly selecting N-pixel by N-pixel square patches from an image set, vectorizing these N^2 values, and identifying the basis as the eigenvectors of the mean-subtracted covariance matrix. In contrast, ICA of grayscale imagery produces basis functions that are again bandpass, but are more obviously oriented and are similar to Gabor functions — Gaussian-windowed sine waves [4,7].

2.3 ICA basis functions for natural images

A survey of applications of ICA to the processing of different media (image/video, multimodal brain data, audio, text, and combined data) is provided by [16]. However, while ICA has been widely used for classification, implications of ICA for multimedia compression have not been greatly studied, and usually have been discussed in simple terms of dimensional reduction. Studies including a bit allocation scheme have so far considered only audio [17], and grayscale imagery [18–21], with inclusion of color in still imagery only in our previous preliminary papers [22,23].
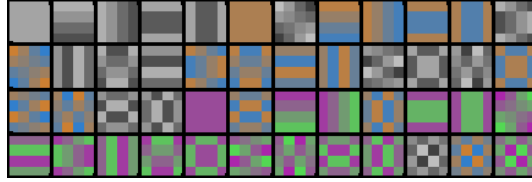
Fig. 4. Basis patches for DCT decomposition of spatio-chromatic $4 \times 4 \times RGB$ patches. Sorted in order of decreasing variance-accounted-for from left to right and top to bottom.

When color is included, our patch vectors become $N \times N \times 3$ structures, for RGB images. PCA proceeds as stated above, but for these longer vectors, and the resulting structure can be visualized as in Fig. 3 (shown for patch size $N = 4$). Note that the PCA basis is adaptive to the image (but in fact does not change much, for natural imagery we have tried).

In [2], Ruderman et al. first extended color PCA, as in the red vectors in Fig. 2, to a spatial patch domain by using 3×3 patches of 3-vector color data. In comparison, in a sense Fig. 2 shows results for 1×1 spatial patches. Ruderman et al. conclude that for foliage images, PCA of log long-medium-short visual color channel data tends to decorrelate spatial processes from chromatic ones, leading to 9 spatial features times 3 color ones. The latter are, in order, luminance, blue-yellow, and red-green. This result was extended by Wachtler et al. [24,25] by replacing PCA with ICA, again for LMS data but now using 7×7 patches.

Color images and stereo vision have also been investigated in [26], which states that the derived independent components again yield a separation of basis vectors into luminance and opponent colors. However, we find this separation is of lesser extent than for PCA (see Fig. 5).

Besides allowing for conclusions regarding human visual perception, these chromatic bases with spatial extent are very interesting from an image compression point of view. In the following we take a closer look at the implications of encoding visual data with respect to these bases.

3 Data specific basis functions

The goal of our analysis is to compare the suitability of different data-adaptive basis functions for compressing color visual data. Therefore, a standard set of color images is chosen that spans a variety of outdoor scenes containing plants, animals, humans and artificial objects. The perspectives of the images range from detail shots to panoramas [2] . In the following we will consider three different sets of bases — two data-adaptive methods, *viz.* PCA and ICA, and one general basis, the discrete cosine transform DCT, moved here into a color domain such that it resembles the PCA.

The DCT basis is a descendant of the Fourier transform. To apply it to the spatio-chromatic setting we simply treat the two spatial patch extents and the index of RGB color components as a three dimensional rectangular prism, e.g., a $3 \times 3 \times 3$ cube. Fig. 4 shows the resulting basis re-assembled as a color picture. Note the similarity of this artificially generated basis to that obtained by PCA

---

[2] We use the 23 publicly available stills from `http://www.cipr.rpi.edu/`
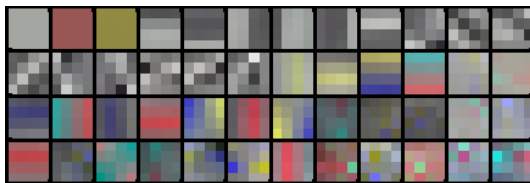
Fig. 5. Basis patches for ICA of spatio-chromatic $4 \times 4 \times RGB$ patches. Sorted in order of decreasing variance.

in Fig. 3. The most prominent difference is a slightly changed orientation of the directional frequencies that in Fig. 4 are axis-aligned . A similar explanation applies to the fact that the main color axes are chosen differently. Also note that the ordering by the variance-accounted-for is quite different because of the different alignment of the basis. In the DCT case the pure color dimensions appear later in the sequence after several luminance frequencies. In the PCA case all three of them appear as the most significant vector. We note that indeed color seems to be fairly separated from luminance, and also that the order is luminance, blue-yellow, and red-green.

Fig. 5 shows the result of ICA performed on the standard image set, for a particular patch size. While the results of DCT and PCA can be interpreted as a frequency decomposition of the data, the functions obtained by ICA exhibit a combined localization in space and frequency [26,22]. This basis seems to somewhat separate opponent color from luminance — but much less so than already illustrated in Fig. 3 for the PCA case: ICA entangles color and luminance. The patterns are not rectangular, but instead resemble Gabor functions.

Besides deciding on a method of basis generation, we have to make a choice about the size of the patches we will operate on. To create a basis, the analysis is performed on square pixel neighborhoods. We randomly sampled a total of 50000 patches over the images of the given set. The resulting basis functions then reflect the statistical properties of the presented data.

As patch size increases, the *sparse* nature of the ICA basis patches becomes more evident. Fig. 6 shows the ICA basis sets for 5×5, 8×8, and 16×16 patches. We see that features captured by the ICA basis are indeed localized in space.

To use the basis for reconstruction, the images are regularly tiled into an arrangement of non-overlapping patches. As mentioned above in the discussion of eq. (4), the coefficients for each patch of the image can be obtained by a linear transform using the filter patches. These are essentially the inverse of the basis patches. Respectively, going back from the coefficients to the actual image data is done by transforming the coefficients in a linear combination of the basis patches.

3.1   *Quantization and entropy encoding of coefficients*

After having projected the image data to the new basis the resulting coefficients have to be reduced in some way. If no reduction takes place no compression will apply. The method we have implemented performs a variance-based quantization. The discretized output can then be efficiently compressed using entropy-based compression.
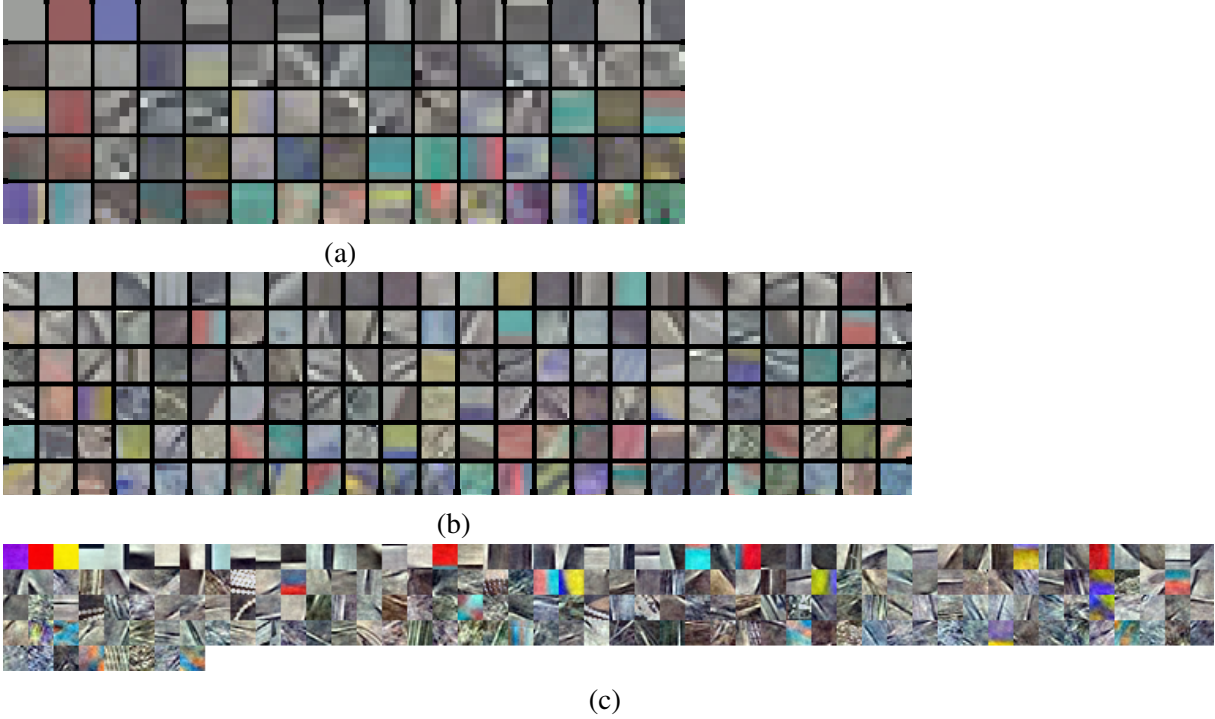
(a)

(b)

(c)

Fig. 6. ICA generated spatio-chromatic basis functions for (a): 5×5; (b): 8×8; and (c): 16×16 patches. As the patch size increases, there are $N \times N \times 3$ patches in each sub-figure, meaning 75 patches, 192 patches, and 768 patches respectively (To save space, just the top-importance basis patches are shown).

The coefficients are given as continuous (floating point) numbers. To apply entropy coding, we have to quantize them first. In our tests we use the method of assigning each basis vector a number of bits proportional to its standard deviation [27]. The proportionality factor is chosen to fulfill a given overall contingent of bits. Given the number of bits, each channel of coefficients is uniformly quantized from its minimum to its maximum occurring value.

Now, having expressed the coefficients as discrete numbers, we can apply an entropy-based encoding, e.g. Huffman coding, or other variable length coding (VLC). The VLC compression indicated in the graphs below is a theoretical limit that can be computed from the sum of entropies for all channels by exponentiation to the base 2. The rate of compression, then, is the factor by which the estimated encoded data is smaller than the original data, which has been stored with 8 bits per channel.

3.2 Quality of reconstruction

Besides considering the size of the data after compression, we most importantly have to look at the quality of the reconstruction obtained from the reduced representations. Here we assess this quality using the peak signal to noise ratio (PSNR), in decibels (dB)

10

$$PSNR = 10 \log_{10} \left(\frac{G^2}{MSE} \right) \tag{14}$$

$$MSE = \frac{\sum_i^N \sum_j^M \sum_{k \in RGB} (p_k(i,j) - o_k(i,j))^2}{3MN} \tag{15}$$

where G is the maximum representable value (e.g. 255) and MSE is the mean squared error of the reconstructed picture, $p_k(i,j)$, with respect to the original, $o_k(i,j)$. A drawback of the PSNR is that it does not actually reflect the distortion as *perceived* by a human observer Nevertheless, perception-based image quality metrics have been found to offer little advantage over PSNR as a measure to evaluate the quality of reconstruction [28], and PSNR is used below for our subsequent evaluation.

4 Evaluation of different bases

The efficiency of a basis is understood as the relation between image quality retained for an achievable rate of compression (or vice versa). Thus, we have conducted a number of tests for different sets of color image bases (ICA, PCA, and DCT). Each basis is generated and applied separately over a range of squared patch sizes from 1×1 to 16×16, with each dimension $\times$ RGB. The first case is similar to just interpreting the pixel colors, as in Fig. 2. As patch size increases, the influence of neighbors is included more and more. Another variable in the comparison is the compression parameter. This is the ratio to the overall maximum number of bits for the stream. Note that the achievable compression entirely depends on the entropy of the data.

4.1 Compression vs. quality using spatio-chromatic bases

The importance of each basis function for representing the data can be determined by looking at the variance of its coefficients. Mostly, these coefficients are centered around zero. The basis vectors found by PCA and ICA each cover most of the variance in the first few basis patches, with a steep falloff.

Fig. 7(a,b) provides a comparison of the entropy-based variable length coding of the spatio-chromatic coefficients of ICA vs. PCA. Here the property of ICA to result in sparsely coded coefficients becomes apparent. The lower entropy of the quantized data results in significantly higher compression rates. Nevertheless, as the error surface for ICA is more bent to the back we note that the PSNR for the mid-range compression rates is lower than for PCA. Both plots show a significant improvement of the compression/error tradeoff as the patch size increases.

DCT results are very similar to those for PCA. The DCT uses the $N \times N \times 3$ generalization of that in the JPEG standard, which applies its algorithm to 3 color planes treated separately and must therefore of necessity give poorer compression than the DCT here since it ignores color correlation. Moreover, we consider a range of patch sizes, not just 8×8.
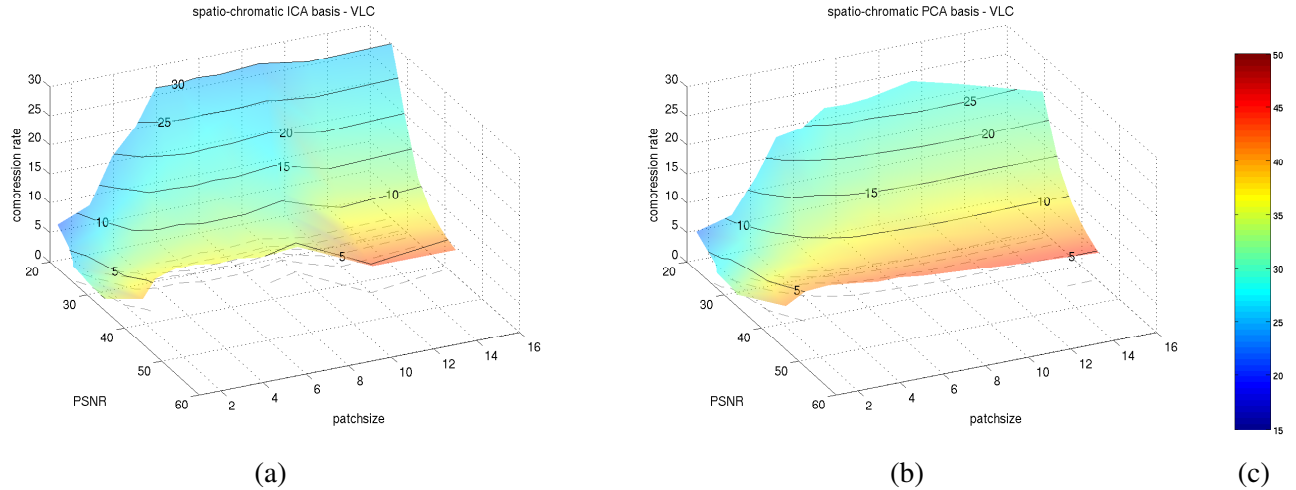
Fig. 7. Entropy-based compression of a set of color images using (a): ICA and (b): PCA generated spatio-chromatic basis functions: color indicates PSNR as per (c). Note that for about equal pseudocolors (equal PSNR), ICA generally has a much better (higher) compression ratio.
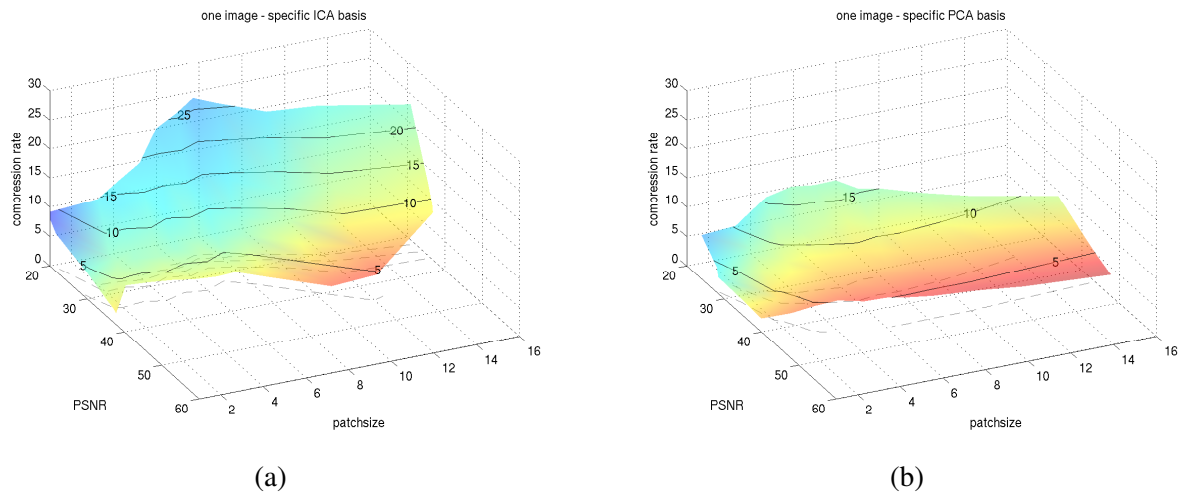

Fig. 8. Entropy-based compression of Fig. 9 a) specific ICA, b) specific PCA. Again, note that for about equal PSNR, ICA has a better compression ratio.

The main result from Fig. 7 is that ICA is considerably better than PCA, particularly at high compression rates, in that it exhibits higher compression for the same PSNR. And for equal compression, quality is increased for increased patch sizes.

4.2 Performance of a specialized basis

The previous examples have shown the quality of reconstruction for general bases that were obtained from the entire set of images using bases obtained from the entire set of images. The following analysis considers just one image (e.g., we display the results for that in Fig. 9). This allows us to tailor a *specialized* basis for a particular image.

In the specific encoding test in Fig. 8, ICA performs significantly better than PCA for large patch sizes. It achieves higher compression for the same quality of reconstruction. Also, a trend can be noted that ICA tends toward increased quality with larger patch sizes while PCA roughly stays constant. The DCT statistics for this image are not included here because they are very similar to the outcome of the PCA for patch sizes larger than 4. These particular plots are for the image in Fig. 9a. A visual comparison in Fig. 9(b,c) shows a detail of the top image. The lower half is the same image compressed to the same projected file size (12:1 compression ratio not including basis data size) using a DCT basis of the same block size (16×16). DCT exhibits strong blocking artifacts while these are hardly noticeable in the ICA version. This is borne out by a comparison of the PSNR (for the same compression ratio in each) of 35.55 for ICA versus 31.97 for PCA. In comparison, baseline sequential JPEG compression at the same compression ratio (quality factor 35) gives a PSNR of 31.94 (without chroma subsampling, the use of which would generate an inappropriate comparison).

Another example is given in Fig. 10. Here, the compression ratio is 7:1. Again, ICA clearly outperforms PCA: for the same compression ratio, ICA yields a PSNR of 39.69, while that for PCA is only 31.40. Standard JPEG compression at the same compression ratio (quality factor 40 with no chroma subsampling) yields a PSNR of only 29.32. The PCA-based compression and JPEG yield blocky ringing around details, and a change in the pattern on the roof. So PSNR does indeed capture image quality, and the plots capture a large amount of information relating compression to quality. Other images we studied also had similar graphical results.

5 Summary

The computation of individual bases for restricted sets of images is interesting from both vision and image processing points of view. While the first point has been subject of previous work targeting analogies to human perception, we have tried to illuminate the latter. The results indicate a significant difference comparing the compressibility of coefficients from ICA and PCA. The sparse coding property of ICA bases has been shown to have a noticeable impact on the efficiency of subsequent entropy compression.

In previous work, we showed that color produces more efficient compression than simply using grayscale [22]. As well, we see here that ICA entangles color and luminance. Generally, ICA bases certainly outperform DCT and PCA for large patch sizes and low rate encoding. We see that the PSNR is a good indicator for the superior quality of the ICA basis in the chosen compression configuration.

Since we found that the spatial and spectral dimensions are almost independent in a PCA basis, we should ask what advantage in general the spatio-chromatic approach (i.e., the use of 3D basis functions) has over using space and color basis functions that are actually independent. [3] The combined encoding of the channels can in fact only improve compression of the coefficients. One
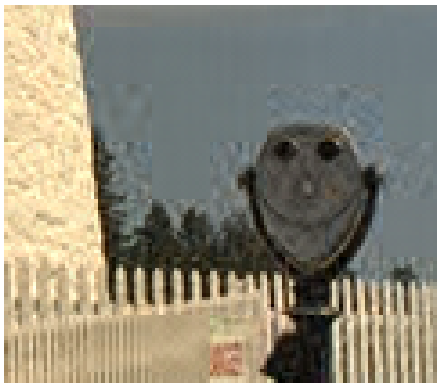
---

[3] We are indebted to an anonymous referee who asked for this issue to be explicitly addressed.
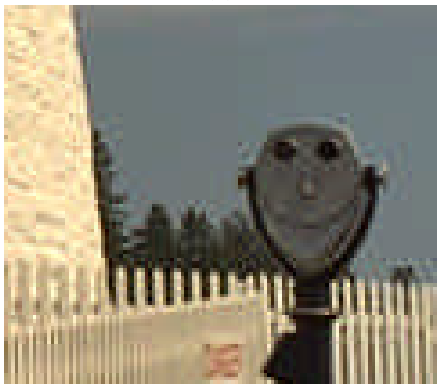
(a)
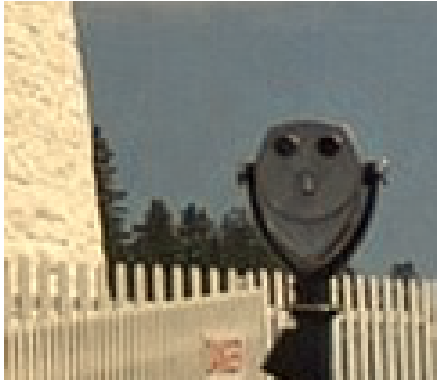
(b)

(c)

(d)

Fig. 9. (a): Original image. (b): Compression ratio 12:1 using an image specific ICA basis (PSNR 35.55), and (c): Same compression ratio using DCT compression (PSNR 31.97). Both are for 16×16 patches. (d): JPEG compression result (no chroma subsampling) — PSNR 31.94.

Fig. 10. (a): Original image. (b): Compression ratio 7:1, using ICA basis (PSNR 39.69), and detail of roof. (c): Using PCA basis (PSNR 31.40), and detail. Both are for 12×12 patches. (d) Using JPEG (no chroma subsampling) gives PSNR 29.32. Details are better preserved using ICA.

can make a sphere packing argument to justify this: a Cartesian-lattice (corresponding to an independent encoding) is increasingly inferior to the densest possible packing of vector quantization sites as the dimension (number of jointly encoded channels) rises. While this argument is based on signal processing alone, it is further the high correlation among the color channels that provides the main reason for our approach. Our comparison with standard JPEG compression, which treats each color plane separately, indeed shows the advantage of linking color and spatial dimensionality.

A problem inherent in the approach of adaptive bases is that they first have to be generated in a computationally expensive preprocessing. Furthermore, a basis specific to one data set would have to be stored along with the coefficients to allow for decoding. This would certainly add overhead to the compressed data. Nevertheless, in a constrained domain it is possible to prepare basis functions that can be re-used.

Because of the proximity of the outcome of independent component analysis to receptive fields of simple cells in the V1 visual cortex, it could be possible to derive a more perceptually-based error metric for evaluation of the quality of visual representations. Advances of research in the human perceptual system may lead the way to an error metric that more closely corresponds to the assessment by a human observer.

Another interesting property of the ICA basis is that it resembles expressive features of the data. This property also hints at the relationship between ICA filters and wavelet analysis. Taking this into account, it seems worthwhile to consider the compressed coefficients as a higher-level feature description of the visual data. In terms of video analysis these features might be useful for object tracking. The inclusion of motion results in a temporal spatio-chromatic basis [23]. As well as revealing implications for video compression, this may indeed also bear relation to human perception.

Appendix A

Using eq. (13), the algorithm for ICA based on maximizing kurtosis may be phrased as follows:

(1) Initialize $\boldsymbol{w}$ randomly, with $\|\boldsymbol{w}\| = 1$
(2) Gradient ascent:

$$\boldsymbol{w} \leftarrow \boldsymbol{w} + \frac{1}{N} \sum_{i=1}^{N} \left[\left(\boldsymbol{w}^{T} \boldsymbol{x}\left[i\right] \right)^{3} \boldsymbol{x}\left[i\right] \right]$$

(3) Re-normalize: $\boldsymbol{w} \leftarrow \boldsymbol{w} / \|\boldsymbol{w}\|$
(4) Stop when $\boldsymbol{w}_{old} \cdot \boldsymbol{w}_{new} \simeq 1$

A simple `matlab` program for this procedure is as follows:

```
[k,N] = size(x); % N data observations of k-vectors.

[xmu, xcentered] = centering(x);
```

```
    % subtracts means from each of j=1..k components.
[x_tilde, M] = whitening(xcentered);   % eq. (9)
    % so now is Whitened: x_tilde*x_tilde'=eye(2);
    % and Orthogonal: x_tilde(1,:)*x_tilde(2,:)'==0
% Now use kurtosis:
% For whitened data E(x^2)==1, the kurtosis is the fourth central moment - 3

% initialize a column of W:
maxNumIterations  = 50;
w = 1/sqrt(2)*ones(2,1);
st = 1/sqrt(N);
w = w/st;
epsilon = 1e-7;
alpha = 1.0;
itn = 1;
done = 0;
allkk = [];
while (itn <= maxNumIterations) && (~done)
    tempvec = zeros(n,1);
    for i=1:N
        tempvec = tempvec + (w'*x_tilde(:,i))^3 * x_tilde(:,i);
    end
    tempvec = 4/N*tempvec;
    wnew1 = w + alpha * tempvec;
    wnew2 = w - alpha * tempvec;
    wnew1 = wnew1/norm(wnew1)*(1/st);
    wnew2 = wnew2/norm(wnew2)*(1/st);
    kk1 = kurt(wnew1,x_tilde);
    kk2 = kurt(wnew2,x_tilde);
    if kk1>kk2 % gradient ascent
        kk=kk1; wnew=wnew1;
    else
        kk=kk2; wnew=wnew2;
    end
    allkk = [allkk kk];
    done = (abs((wnew'*w)-N)<=epsilon); % stopping criterion.
    w = wnew;
    itn = itn+1;
end
plot(allkk); disp(w);
% So first ICA component is s = w'*x_tilde.

% ===========================
function kk=kurt(w,x_tilde)
% std(snew)==1 if make norm(w)=sqrt(N), since x_tilde*x_tilde' == eye(k)
```

```
w = w/norm(w)*sqrt(size(x_tilde,2));
snew = w'*x_tilde; % so that snew*snew'=1.0; also mean(snew)==0.0
kk = abs( mean( snew.^4 )-3 );
```

References

[1] H. Fairman, M. Brill, The principal components of reflectances, Color Research and Application 29 (2004) 104–110.

[2] D. L. Ruderman, T. W. Cronin, C.-C. Chiao, Statistics of cone responses to natural images: Implications for visual coding, J. Opt. Soc. Am. 15 (1998) 2036–2045.

[3] M. A. Webster, J. D. Mollon, Adaptation and the color statistics of natural images, Vision Research 37 (1997) 3283–3298.

[4] B. Olshausen, D. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–609.

[5] Z.-N. Li, M. Drew, Fundamentals of Multimedia, Prentice-Hall, 2004.

[6] D. Alleysson, S. Süsstrunk, Spatio-chromatic PCA of a mosaiced color image, in: 2nd European Conference on Color in Graphics, Imaging and Vision, Society for Imaging Science & Technology (IS&T), 2004, pp. 311–314.

[7] J. van Hateren, A. van der Schaaf, Independent component filters of natural images compared with simple cells in primary visual cortex, Proc. Roy. Soc. B 265 (1998) 359–366.

[8] B. Draper, K. Baek, M. Bartlett, J. Beveridge, Recognizing faces with PCA and ICA, Comp. Vis. and Im. Understanding 91 (2003) 115–137.

[9] J.-Y. Pan, C. Faloutsos, Videocubes: A novel tool for video mining and classification, in: ICADL 2002, 2002.

[10] J. Larsen, L. Hansen, T. Kolenda, F. Nielsen, Independent component analysis in multimedia modeling, in: 4th Int. Symp. on Independent Component Analysis and Blind Source Separation (ICA 2003), 2003.

[11] E. Sahouria, A. Zakhor, Content analysis of video using principal components, IEEE Trans. on Circ. and Sys. for Video Tech. 9 (1998) 1290–1298.

[12] M. Drew, J. Au, Clustering of compressed illumination-invariant chromaticity signatures for efficient video summarization, Image and Vision Computing Journal 21 (8, August) (2003) 705–716.

[13] A. Hyvärinen, E. Oja, Independent component analysis: Algorithms and applications, Neural Networks 13 (2000) 411–430.

[14] E. Oja, Convergence of the symmetrical FastICA algorithm, in: 9th Int. Conf. on Neural Information Processing, 2002.

[15] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley and Sons, Inc., New York, 2001.

[16] J. Larsen, L. K. Hansen, T. Kolenda, F. A. Nielsen, Independent component analysis in multimedia modeling, in: Fourth Intl. Symposion on ICA and BSS, Nara, Japan, 2003, pp. 687–696.

[17] A. Ben-Shalom, M. Werman, S. Dubnov, Improved low bit-rate audio compression using reduced rank ICA instead of psychoacoustic modeling, in: ICASSP 2003, 2003.

[18] A. Ferreira, M. Figueiredo, Class-adapted image compression using independent component analysis, in: ICIP 2003, 2003, pp. I: 625–628.

[19] A. Puga, A. Alves, An experiment on comparing PCA and ICA in classical transform image coding, in: ICA 98, 1998, pp. 105–108.

[20] M. Narozny, M. Barret, D.-T. Pham, I. Akambita, Modified ica algorithms for finding optimal transforms in transform coding, in: 4th Int. Symp. on Image and Signal Proc. and Analy. 2005, pp. 111–0–116.

[21] A. Ferreira, M. Figueiredo, On the use of independent component analysis for image compression, Sig. Proc.: Image Commun. 21 (2006) 378–389.

[22] M. Drew, S. Bergner, Analysis of spatio-chromatic decorrelation for colour image reconstruction, in: 12th Color Imaging Conference: Color, Science, Systems and Applications., Society for Imaging Science & Technology (IS&T)/Society for Information Display (SID) joint conference, 2004, pp. 87–92.

[23] S. Bergner, M. Drew, Spatiotemporal-chromatic structure of natural scenes, in: Int. Conf. on Image Processing, IEEE, 2005.

[24] T. Wachtler, T.-W. Lee, T. Sejnowski, Chromatic structure of natural scenes, J. Opt. Soc. Am. A 18 (2001) 65–77.

[25] T.-W. Lee, T. Wachtler, T. Sejnowski, Color opponency is an efficient representation of spectral properties in natural scenes, Vis. Res. 42 (2002) 2095–2103.

[26] P. Hoyer, A. Hyvrinen, Independent component analysis applied to feature extraction from colour and stereo images, Network: Computation in Neural Systems 11 (2000) 191–210.

[27] P. Cosman, R. Gray, M.Vetterli, Vector quantization of image subbands: A survey, IEEE Trans. on Image Proc. 5 (1996) 202–225.

[28] Z. Wang, A. Bovik, L. Lu, Why is image quality assessment so difficult?, in: IEEE Int. Conf. on Acoustics, Speech & Signal Proc., 2002, pp. IV–3313–IV–3316.