FAST COMPUTATION OF ENTROPIES AND MUTUAL INFORMATION FOR MULTISPECTRAL IMAGES

Sié Ouattara, Alain Clément and François Chapeau-Blondeau
Laboratoire d'Ingénierie des Systèmes Automatisés (LISA), Université d'Angers
62 avenue Notre Dame du Lac, 49000 Angers, France

Keywords: Multispectral images, Entropy, Mutual information, Multidimensional histogram.

Abstract: This paper describes the fast computation, and some applications, of entropies and mutual information for color and multispectral images. It is based on the compact coding and fast processing of multidimensional histograms for digital images.

1 INTRODUCTION

Entropies and mutual information are important tools for statistical analysis of data in many areas. For image processing, so far, these tools have essentially been applied to scalar or one-component images. The reason is that these tools are usually derived from multidimensional histograms, whose direct handling is feasible only in low dimension due to their memory occupation and related processing time which become prohibitively large as the number of image components increases. Here, we use an appraoch for multidimensional histograms allowing compact coding and fast computation, and show that this approach easily authorizes the computation of entropies and mutual information for multicomponent or multispectral images.

2 A FAST AND COMPACT MULTIDIMENSIONAL HISTOGRAM

We consider multispectral images with D components $X_i(x_1,x_2)$, for $i = 1$ to D, each X_i varying among Q possible values, at each pixel of spatial coordinate (x_1,x_2). A D-dimensional histogram of such an image would comprise Q^D cells. For an image with $N_1 \times N_2 = N$ pixels, only at most N of these Q^D cells can be occupied, meaning that, as D grows, most of the cells of the D-dimensional histogram are in fact empty. For example, for a common 512×512 RGB color image with $D = 3$ and $Q = 256 = 2^8$, there are $Q^D = 2^{24} \approx 16 \times 10^6$ colorimetric cells with at most only $N = 512^2 = 262\,144$ of them which can be occupied. We developed the idea of a compact representation of the D-dimensional histogram (Clément and Vigouroux, 2001; Clément, 2002), where only those cells that are occupied are coded. The D-dimensional histogram is coded as a linear array where the entries are the D-tuples (the colors) present in the image and arranged in lexicographic order of their components $(X_1,X_2,\ldots X_D)$. To each entry (in number $\leq N$) is associated the number of pixels in the image having this D-value (this color). An example of this compact representation of the D-dimensional histogram is shown in Table 1.

The practical calculation of such a compact histogram starts with the lexicographic ordering of the N D-tuples corresponding to the N pixels of the image. The result is a linear array of the N ordered D-tuples. This array is then linearly scanned so as to merge the neighboring identical D-tuples while accumulating their numbers to quantify the corresponding population of pixels. With a dichotomic quick sort algorithm to realize the lexicographic ordering, the whole process of calculating the compact multidimensional histogram can be achieved with an average complexity of $O(N \log N)$, independent of the dimension D. Therefore, both compact representation and its fast calculation are afforded by the process for the

Ouattara S., Clément A. and Chapeau-Blondeau F. (2007).
FAST COMPUTATION OF ENTROPIES AND MUTUAL INFORMATION FOR MULTISPECTRAL IMAGES.
In *Proceedings of the Fourth International Conference on Informatics in Control, Automation and Robotics*, pages 195-199
DOI: 10.5220/0001622801950199

Table 1: An example of compact coding of the 3-dimensional histogram of an RGB color image with $Q = 256$. The entries of the linear array are the components $(X_1, X_2, X_3) =$ (R, G, B) arranged in lexicographic order, for each color present in the image, and associated to the poupulation of pixels having this color.

R	G	B	population
0	0	4	13
0	0	7	18
0	0	23	7
⋮	⋮	⋮	⋮
255	251	250	21
255	251	254	9

multidimensional histogram.

For example, for a 9-component 838×762 satellite image with $Q = 2^8$, the compact histogram was calculated in about 5 s on a standard 1 GHz-clock desktop computer, with a coding volume of 1.89 Moctets, while the classic histogram would take 3.60×10^{16} Moctets completely unmanageable by today's computers.

3 ENTROPIES AND MUTUAL INFORMATION FOR IMAGES

3.1 Fast Computation from Compact Histogram

The multidimensional histogram of an image, after normalization by the number of pixels, can be used for an empirical definition of the probabilities $p(\vec{X})$ associated to the D-values present in image $\vec{X}$. In the compact histogram, coded as in Table 1, only those D-values of $\vec{X}$ with nonzero probability are represented. This is all that is needed to compute any property of the image that is defined as a statistical average over the probabilities $p(\vec{X})$. This will be the case for the statistical moments of the distribution of D-values in the image (Romantan et al., 2002), for the principal axes based on the cross-covariances of the components X_i (Plataniotis and Venetsanopoulos, 2000), and for the entropies, joint entropies and mutual information that we consider in the sequel.

An entropy $H(\vec{X})$ for image $\vec{X}$ can be defined as (Russ, 1995)

$$H(\vec{X}) = -\sum_{\vec{X}} p(\vec{X}) \log p(\vec{X}) \,. \qquad (1)$$

The computation of $H(\vec{X})$ of Eq. (1) from the normalized compact histogram from Table 1, is realized simply by a linear scan of the array while summing the terms $-p(\vec{X}) \log p(\vec{X})$ with $p(\vec{X})$ read from the last column. This preserves the overall complexity of $O(N \log N)$ for the whole process leading to $H(\vec{X})$.

The compact histogram also allows one to envisage the joint entropy of two multicomponent images $\vec{X}$ and $\vec{Y}$, with dimensions D_X and D_Y respectively. The joint histogram of $(\vec{X}, \vec{Y})$ can be calculated as a compact histogram with dimension $D_X + D_Y$, which after normalization yields the joint probabilities $p(\vec{X}, \vec{Y})$ leading to the joint entropy

$$H(\vec{X}, \vec{Y}) = -\sum_{\vec{X}} \sum_{\vec{Y}} p(\vec{X}, \vec{Y}) \log p(\vec{X}, \vec{Y}) \,. \qquad (2)$$

A mutual information between two multicomponent images follows as

$$I(\vec{X}, \vec{Y}) = H(\vec{X}) + H(\vec{Y}) - H(\vec{X}, \vec{Y}) \,. \qquad (3)$$

And again, the structure of the compact histogram preserves the overall complexity of $O(N \log N)$ for the whole process leading to $H(\vec{X}, \vec{Y})$ or $I(\vec{X}, \vec{Y})$.

So far in image processing, joint entropies and mutual information have essentially been used for scalar or one-component images (Likar and Pernus, 2001; Pluim et al., 2003), because the direct handling of joint histograms is feasible only in low dimension, due to their memory occupation and associated processing time which get prohibitively large as dimension increases. By contrast, the approach of the compact histogram of Section 2 makes it quite tractable to handle histograms with dimensions of 10 or more. By this approach, many applications of entropies and mutual information become readily accessible to color and multispectral images. We sketch a few of them in the sequel.

3.2 Applications of Entropies

The entropy $H(\vec{X})$ can be used as a measure of complexity of the multicomponent image $\vec{X}$, with application for instance to the following purposes:

- An index for characterization / classification of textures, for instance for image segmentation or classification purposes.
- Relation to performance in image compression, especially lossless compression.

For illustration, we use the entropy of Eq. (1) as a scalar parameter to characterize RGB color images carrying textures as shown in Fig. 1. The entropies $H(\vec{X})$ given in Table 2 were calculated from 512×512 three-component RGB images $\vec{X}$ with $Q = 256$. The whole process of computing a 3-dimensional histogram and the entropy took typically less than one second on our standard desktop computer. For comparison, another scalar parameter $\sigma(\vec{X})$ is also given in Table 2, as the square root of the trace of the variance-covariance matrix of the components of image $\vec{X}$. This parameter $\sigma(\vec{X})$ measures the overall average dispersion of the values of multicomponent image $\vec{X}$. For a one-component image $\vec{X}$, this $\sigma(\vec{X})$ would simply be the standard deviation of the gray levels. The results of Table 2 show a specific significance for the entropy $H(\vec{X})$ of Eq. (1), which does not simply mimic the evolution of a common measure like the dispersion $\sigma(\vec{X})$. As a complexity measure, $H(\vec{X})$ of Eq. (1) is low for synthetic images as `Chessboard` and `Wallpaper`, and is higher for natural images in Table 2.

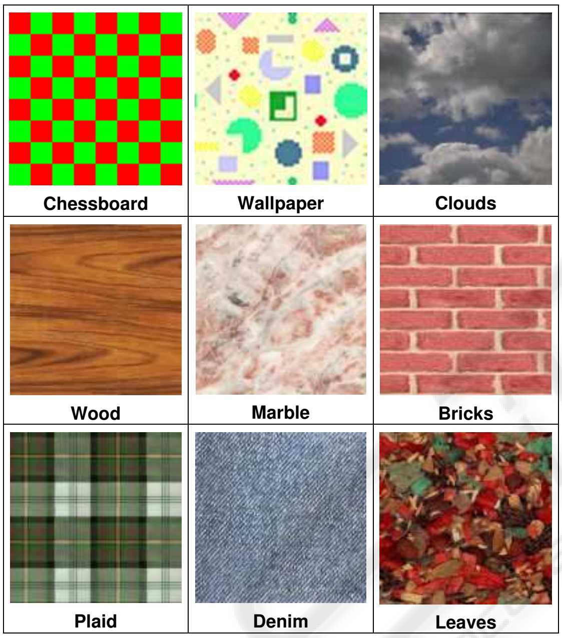

Figure 1: Nine three-component RGB images $\vec{X}$ with $Q = 256$ carrying distinct textures.

3.3 Applications of Mutual Information

The mutual information $I(\vec{X},\vec{Y})$, or other measures derived from the joint entropy $H(\vec{X},\vec{Y})$, can be used as an index of similarity or of relationship, between two multicomponent images $\vec{X}$ and $\vec{Y}$, with application for instance to the following purposes:

• Image matching, alignment or registration, especially in multimodality imaging.

• Reference matching, pattern matching, for pattern recognition.

• Image indexing from databases.

Table 2: For the nine distinct texture images of Fig. 1: entropy $H(\vec{X})$ of Eq. (1) in bit/pixel, and overall average dispersion $\sigma(\vec{X})$ of the components.

texture	$H(\vec{X})$	$\sigma(\vec{X})$
Chessboard	1.000	180.313
Wallpaper	7.370	87.765
Clouds	11.243	53.347
Wood	12.515	31.656
Marble	12.964	44.438
Bricks	14.208	57.100
Plaid	14.654	92.284
Denim	15.620	88.076
Leaves	17.307	74.966

• Homogeneity / contrast assessment for segmentation or classification purposes.

• Performance evaluation of image reconstruction, especially in lossy compression.

• Analysis via principal, or independent, component analysis.

For illustration, we consider a lossy compression on an RGB color $N_1 \times N_2$ image $\vec{X}$ via a JPEG-like operation consisting, on the $N_1 \times N_2$ discrete cosine transform of $\vec{X}$, in setting to zero a given fraction $(1 - \mathrm{CR}^{-1})$ of the high frequency coefficients, or equivalently in retaining only the $N_1/\sqrt{\mathrm{CR}} \times N_2/\sqrt{\mathrm{CR}}$ low-frequency coefficients. From this lossy compression of initial image $\vec{X}$, the decompression reconstructs a degraded image $\vec{Y}$. While varying the compression ratio CR, the similarity between images $\vec{X}$ and $\vec{Y}$ is measured here by the mutual information $I(\vec{X},\vec{Y})$ of Eq. (3) based on the 6-dimensional joint histogram for estimating the joint probabilities $p(\vec{X},\vec{Y})$. In addition, for comparison, we also used a more common measure of similarity formed by the cross-correlation coefficient $C(\vec{X},\vec{Y})$ between $\vec{X}$ and $\vec{Y}$, computed as one third of the sum of the cross-correlation coefficient between each marginal scalar component, R, G or B, of $\vec{X}$ and $\vec{Y}$. This $C(\vec{X},\vec{Y}) = 1$ if $\vec{X}$ and $\vec{Y}$ are two identical images, and it goes to zero if $\vec{X}$ and $\vec{Y}$ are two independent unrelated images. For the results presented in Fig. 2, the choice for initial image $\vec{X}$ is a 512×512 RGB `lena.bmp` with $Q = 256$. The whole process of the computation of an instance of the 6-dimensional joint histogram and the mutual information took around 3 s on our standard desktop computer.

Figure 2 shows that, as the compression ratio CR increases, the mutual information $I(\vec{X},\vec{Y})$ and the cross-correlation coefficient $C(\vec{X},\vec{Y})$ do not decrease in the same way. Compared to the mutual information $I(\vec{X},\vec{Y})$, it is known that the cross-correlation

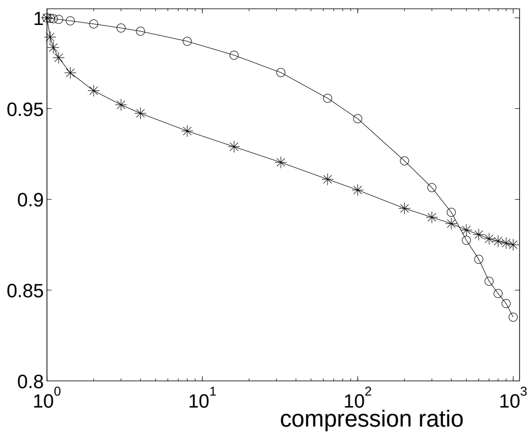

Figure 2: As a function of the compression ratio CR: ($\circ$) cross-correlation coefficient $C(\vec{X},\vec{Y})$ between initial RGB color image $\vec{X}$ and its compressed version $\vec{Y}$; ($*$) normalized mutual information $I(\vec{X},\vec{Y})/H(\vec{X})$ with the entropy $H(\vec{X}) = 16.842$ bits/pixel.

$C(\vec{X},\vec{Y})$ measures only part of the dependence between $\vec{X}$ and $\vec{Y}$. Figure 2 indicates that when the compression ratio CR starts to rise above unity, $I(\vec{X},\vec{Y})$ decreases faster than $C(\vec{X},\vec{Y})$, meaning that information is first lost at a faster rate than what is captured by the cross-correlation. Meanwhile, for large CR in Fig. 2, $I(\vec{X},\vec{Y})$ comes to decrease slower than $C(\vec{X},\vec{Y})$. This illustrates a specific contribution of the mutual information computed for multicomponent images.

Another application of mutual information between images can be found to assess a principal component analysis. On a D-component image $\vec{X} = (X_1,\dots X_D)$, principal component analysis applies a linear transformation of the X_i's to compute D principal components $(P_1,\dots P_D)$ with vanishing cross-correlation among the P_i's, in such a way that some P_i's can be selected for a condensed parsimonious representation of initial image $\vec{X}$. An interesting quantification is to consider the mutual information $I(\vec{X},P_i)$. From its theoretical properties, the joint entropy $H(\vec{X},P_i)$ reduces to $H(\vec{X})$ because P_i is deterministically deduced from $\vec{X}$, henceforth $I(\vec{X},P_i) = H(P_i)$. This relationship has been checked (on several 512×512 RGB color images with $D = 3$) to be precisely verified by our empirical entropy estimators for $I(\vec{X},P_i)$ based on the computation of $(D+1)$-dimensional histograms for $(\vec{X},P_i)$. This offers a quantification of the relation between $\vec{X}$ and its principal components P_i; a subset of the whole P_i's could be handled in a similar way. Another useful quantification shows that principal component analysis, although it cancels cross-correlation between the components, does not cancel dependence between them, and sometimes it may even increase it in some sense, as illustrated by the behavior of the mutual information in Table 3, with $I(P_1,P_2)$ larger than $I(X_1,X_2)$ for image (2). The mutual information can serve as a measure to base other separation or selection schemes of the components from an initial multispectral image $\vec{X}$.

Table 3: For a 512×512 RGB color image $\vec{X}$ with $D = 3$ and $Q = 256$: cross-correlation coefficient $C(\cdot,\cdot)$ and mutual information $I(\cdot,\cdot)$ of Eq. (3), between the two initial components X_1 and X_2 with largest variance, and between the two first principal components P_1 and P_2 after principal component analysis of $\vec{X}$. (1) image $\vec{X}$ is `lena.bmp`. (2) image $\vec{X}$ is `mandrill.bmp`.

	$C(X_1,X_2)$	$I(X_1,X_2)$	$C(P_1,P_2)$	$I(P_1,P_2)$
(1)	0.879	1.698	0.000	0.806
(2)	0.124	0.621	0.000	0.628

4 CONCLUSION

We have reported the fast computation and compact coding of multidimensional histograms and showed that this approach authorizes the estimation of entropies and mutual information for color and multispectral images. Histogram-based estimators of these quantities as used here, become directly accessible with no need of any prior assumption on the images. The performance of such estimators clearly depends on the dimension D and size $N_1 \times N_2$ of the images; we did not go here into performance analysis, especially because this would require to specify statistical models of reference for the measured images. Instead here, more pragmatically, on real multicomponent images, we showed that, for entropies and mutual information, direct histogram-based estimation is feasible and exhibits natural properties expected for such quantities (complexity measure, similarity index, ...). The present approach opens up the way for further application of information-theoretic quantities to multispectral images.

REFERENCES

Clément, A. (2002). *Algorithmes et outils informatiques pour l'analyse d'images couleur. Application à l'étude de coupes histologiques de baies de raisin en microscopie optique.* Ph. D. thesis, University of Angers, France.

Clément, A. and Vigouroux, B. (2001). Un histogramme compact pour l'analyse d'images multi-composantes.

In *Proceedings 18è Colloque GRETSI sur le Traitement du Signal et des Images*, pages 305–307, Toulouse, France, 10–13 Sept. 2001.

Likar, B. and Pernus, F. (2001). A hierarchical approach to elastic registration based on mutual information. *Image and Vision Computing*, 19:33–44.

Plataniotis, K. N. and Venetsanopoulos, A. N. (2000). *Color Image Processing and Applications*. Springer, Berlin.

Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. (2003). Mutual-information-based registration of medical images: A survey. *IEEE Transactions on Medical Imaging*, 22:986–1004.

Romantan, M., Vigouroux, B., Orza, B., and Vlaicu, A. (2002). Image indexing using the general theory of moments. In *Proceedings 3rd COST 276 Workshop on Information and Knowledge Management for Integrated Media Communications*, pages 108–113, Budapest, Hungary, 11–12 Oct. 2002.

Russ, J. C. (1995). *The Image Processing Handbook*. CRC Press, Boca Raton.