Using Neighborhood Distributions of Wavelet Coefficients for On-the-Fly, Multiscale-Based Image Retrieval

Sandrine Anthoine, Eric Debreuve, Paolo Piro, Michel Barlaud

Using neighborhood distributions of wavelet coefficients for on-the-fly, multiscale-based image retrieval

Sandrine Anthoine, Eric Debreuve, Paolo Piro, Michel Barlaud
Laboratoire I3S, Université de Nice Sophia-Antipolis / CNRS
2000 Route des Lucioles; 06903, Sophia-Antipolis Cedex, France
{*anthoine, debreuve, piro, barlaud*}*@i3s.unice.fr*

Abstract

In this paper, we define a similarity measure to compare images in the context of (indexing and) retrieval. We use the Kullback-Leibler (KL) divergence to compare sparse multiscale image descriptions in a wavelet domain. The KL divergence between wavelet coefficient distributions has already been used as a similarity measure between images. The novelty here is twofold. Firstly, we consider the dependencies between the coefficients by means of distributions of mixed intra/interscale neighborhoods. Secondly, to cope with the high-dimensionality of the resulting description space, we estimate the KL divergences in the k-th nearest neighbor framework, instead of using classical fixed size kernel methods. Query-by-example experiments are presented.

1. Introduction

Comparing two images in the context of (indexing and) retrieval often relies on global descriptions such as dominant colors or color distribution, or on extracted information such as salient points/regions together with local features or segmentation along with region arrangement [5, 12]. The philosophy here is to use a synthetic, multiscale image description based on the sparse representation in a wavelet domain. Such experiments have been conducted using the marginal distributions of the wavelet coefficients at different scales associated with the Kullback-Leibler (KL) divergence as a similarity measure between distributions [3]. Nevertheless, independence between the coefficients was assumed, preventing from taking into account local image structures such as texture. In contrast, we propose to consider dependency by means of distributions of mixed intra/interscale neighborhoods of coefficients. However, this approach implies to deal with a high-dimensional statistical description space. The number of samples being too small to reasonably fill this space, fixed size kernel options to estimate distributions or divergences fail. Alternatively, we propose to estimate the KL divergence in the k-th nearest neighbor (kNN) framework [2], *i.e.*, adapting to the local sample density and directly from the samples.

2. Similarity between images

A central question in content-based image indexing is to define a similarity measure between images that matches - or at least is close enough to - our perception of the similarity of images. Once this is done, the images in the database can be simply ranked in increasing order of their similarity to the reference (or example) image for a query-by-example task. Perceptual studies to understand how human perceive the similarity between images are still a topic of ongoing research. Therefore, content-based image indexing system relying on such studies may be subjective and hard to implement. Here, we focus on developing an objective and mathematically defined measure that will be easily implementable.

2.1. Neighborhoods of wavelet coefficients

Let us denote by $w(I)_{j,k}$ the wavelet coefficient of image I at scale j and location k. I.e. it is the scalar product $w(I)_{j,k} = \langle \psi_{j,k}, I \rangle$ of I with $\psi_{j,k}$, the mother wavelet ψ translated at location k and dilated at scale j.

The wavelet transform enjoys several properties that have made it quite successful in signal processing and that are relevant for the definition of similarity between images. Indeed, it provides a sparse representation of images, meaning that it concentrates the informational content of an image into few coefficients of large amplitude while the rest of the coefficients are small. This combined with a fast transform is what makes wavelet thresholding methods so powerful: in fact just identifying large coefficients is sufficient to extract where the information lies in the image. Thus

it seems natural to define the feature space in the wavelet domain.

Initial thresholding wavelet methods treated each coefficient separately relying on the decorrelation of these coefficients. However, they are not independent and these dependencies are the signature of structures present in the image. For example, a discontinuity between smooth regions at point k_0 will give large coefficients at this point at all scales j ($w(I)_{j,k_0}$ large for all j). The most significant dependencies are seen between a wavelet coefficient $w(I)_{j,k}$ and its closest neighbors in scale ($w(I)_{j-1,k}$) or space ($w(I)_{j,k\pm(0,1)}$, $w(I)_{j,k\pm(1,0)}$). Several models using these dependencies have been proposed and used in image enhancement [8, 9]. Here we use the concept of wavelet neighborhoods introduced in [8]; these are vectors of wavelets coefficients of the form:

$$\mathbf{w}(I)_{j,k} = \begin{array}{l}\big(w(I)_{j,k}, w(I)_{j-1,k}, \\ \quad w(I)_{j,k\pm(1,0)}, w(I)_{j,k\pm(0,1)}\big)\end{array} . \quad (1)$$

It was shown that the probability density function (pdf) of such neighborhoods allow to characterize and estimate fine spatial structures in images [8, 7]. Hence we will define our feature space on the set of neighborhoods of wavelet coefficients of the form of (1).

Critically sampled tensor wavelet transforms lack of translation and rotation invariance and so would the neighborhoods made of such coefficients. Since it is desirable to find rotated and translated versions of an image to be similar to the original one, we prefer to use a slightly more redundant transform, namely the steerable pyramid [8]. This is the decomposition on the set of dilated, translated and (Fourier-)rotated version of a mother wavelet. The image I is then represented by a set of wavelet coefficients of the form: $\{w(I)_{j,o,k}\}_{j\in\mathbb{Z},k\in\mathbb{Z}^2,o=1..N_o}$, o indexing the orientations. Subsampling is not performed on the first level of decomposition (but is done subsequently) thus allowing that all subbands are aliasing-free (translation invariance in each scale). Moreover the different orientations allow to get some rotation invariance. Although this transform is redundant (with a factor $4N_o/3$), it is fast and enforces sparsity of image decomposition as do the critically sampled wavelet transforms, but it also enjoys more invariance properties than the latter.

The sampling of orientations is rather coarse (usually $N_o = 4$). Therefore dependencies between coefficients at different orientations are less significant than across scale or space. Thus we confine the neighborhoods to each orientation, i.e the neighborhood of $w(I)_{j,o,k}$ is:

$$\mathbf{w}(I)_{j,o,k} = \begin{array}{l}\big(w(I)_{j,o,k}, w(I)_{j-1,k,o}, \\ \quad w(I)_{j,o,k\pm(1,0)}, w(I)_{j,o,k\pm(0,1)}\big)\end{array} . \quad (2)$$

Hence our feature space is the set of the neighborhoods as in (2) for all scales j, orientations o and locations k. Let us now turn to the measure of similarity on this space.

2.2. Similarity measure between images

Since geometrically modified or slightly degraded versions of the same image as well as images containing similar objects should be close, one cannot define a measure comparing directly the neighborhoods one by one, but rather their probability distribution. More specifically, we consider the pdf of the neighborhoods of (2) for each scale and orientation, i.e. we consider the pdf $p_{\mathbf{w}_{j,o}(I)}$ of the set neighborhoods $\{\mathbf{w}(I)_{j,o,k}\}_k$ for each fixed j and o.

The considered pdf are those of coefficients that carry the informational content of the signal. The natural way to compare such pdf is to use measures derived from information theory. Here we use the KL divergence between pdfs, an approach that has also been successfully taken for other applications [2]. This was also done in [3, 11] in the context of evaluating the similarity between images using the marginal pdf of the wavelet coefficients. We propose to use this measure on the multidimensional pdf of the neighborhoods of coefficients: the similarity between images I_1 and I_2 is a weighted sum over orientations and scales of the KL divergences between the pdf $p_{\mathbf{w}_{j,o}(I_1)}$ and $p_{\mathbf{w}_{j,o}(I_2)}$:

$$S(I_1, I_2) = \sum_{j,o} \alpha_j D_{kl}(p_{\mathbf{w}_{j,o}(I_1)} || p_{\mathbf{w}_{j,o}(I_2)}) \quad (3)$$

with $p_{\mathbf{w}_{j,o}(I_i)}$ the pdf of the wavelet neighborhoods of image I_i at scale j and orientation o and $\alpha_j > 0$ are weights (chosen according to the redundancy of the transform).

Previous works on neighborhoods of wavelet coefficients or indexation using marginal pdf of these coefficients all assumed a parametric model for the pdf involved. In the marginal case, efficient models (e.g. generalized Gaussians [3, 11]) lead to an analytic expression of the KL divergence as a function of the model parameter; but they are not easily generalizable to the multidimensional correlated case of wavelet neighborhoods. On the other hand, efficients multidimensional models accounting for correlations (e.g. Gaussian mixtures [7]) fit a wide variety of multidimensional pdf but impose to estimate the KL divergence after estimating the model parameters. Besides the heavy computational cost of the consecutive estimations, the numerical stability of such cascading estimates is difficult to obtain. We prefer to make no hypothesis on the pdf at hand, hence sparing the cost of fitting the model parameters but needing to estimate the KL divergences in this non-parametric case.

2.3. Estimation of the Kullback-Leibler divergences

Let us first remind the reader that the KL divergence between two continuous pdf p_1 and p_2 is:

$$D_{kl}(p_1||p_2) = \int p_1(x) \log \frac{p_1(x)}{p_2(x)} dx = H_x(p_1, p_2) - H(p_1) \quad (4)$$

where H is the differential entropy and H_x is the cross entropy.

The estimation of statistical measures in the multidimensional case is hard. In particular, kernel-based methods such as Parzen estimates become unadapted due to the sparsity of samples in high dimension (curse of dimensionality): the tradeoff between a kernel with a large bandwidth to perform well in low local sample density (which *oversmoothes* the estimator) and a kernel with a smaller bandwidth to preserve local statistical variabilities (which results in an unstable estimator) cannot always be achieved. We use instead the kth nearest neighbor (kNN) framework [10] to compute the KL divergence. Indeed it follows the dual approach to the above fixed size kernel: the bandwidth adapts to the local sample density by letting the kernel contain exactly k neighbors of a given sample. Moreover it allows direct estimation of the divergence without explicitly estimating the pdf.

Assume that ϵ is a set of N_ϵ samples $\mathbf{w}_1$, $\mathbf{w}_2$,..,$\mathbf{w}_{N_\epsilon}$ of pdf p_ϵ. Fix a non-zero integer k. Denote by v_d the volume of the unit sphere in $\mathbb{R}^d$, and ψ the digamma function. Denote by $\mu_\epsilon(g)$ the mean of g over ϵ:

$$\mu_\epsilon(g) = \tfrac{1}{N_\epsilon} \sum_{n=1}^{N_\epsilon} g(\mathbf{w}_n). \tag{5}$$

$\rho_{k,\epsilon}(s)$ is the distance for $s \in \mathbb{R}^d$ to its kth nearest neighbor in $\epsilon - \{s\}$.

kNN balloon estimates are based on the principle that $p_\epsilon(s)$ is inversely proportional to the volume of the sphere containing the k nearest neighbors of s in ϵ [10]:

$$p_\epsilon(s) \sim \frac{k}{v_d\, \rho_{k,\epsilon}^d(s)} \tag{6}$$

An unbiased estimator of the Ahmad-Lin approximation of entropy [1]

$$H_{al}(p_\epsilon) = -\frac{1}{N_\epsilon} \sum_{n=1}^{N_\epsilon} \log\big[p_\epsilon(\mathbf{w}_n)\big] = -\mu_\epsilon\big(\log\big[p_\epsilon\big]\big) \tag{7}$$

in the kNN framework was proposed in [4] by replacing $\log k$ by $\psi(k)$:

$$\widehat{H}(p_\epsilon) = \log[(N_\epsilon - 1)v_d] - \psi(k) + d\, \mu_\epsilon(\log[\rho_{k,\epsilon}]) \tag{8}$$

The cross entropy estimate is then [2]:

$$\widehat{H_x}(p_{\epsilon_1}, p_{\epsilon_2}) = \log[N_{\epsilon_2} v_d] - \psi(k) + d\mu_{\epsilon_2}(\log[\rho_{k,\epsilon_1}]). \tag{9}$$

And the KL divergence estimate is:

$$\begin{aligned}\widehat{D_{\mathrm{kl}}}(p_{\epsilon_1}||p_{\epsilon_2}) = \quad & \log\Big[\tfrac{N_{\epsilon_2}}{N_{\epsilon_1}-1}\Big] + d\, \mu_{\epsilon_2}(\log[\rho_{k,\epsilon_1}]) \\ & - d\, \mu_{\epsilon_1}(\log[\rho_{k,\epsilon_1}])\end{aligned} \tag{10}$$

This expression is valid in any dimension and it is robust to the choice of k.

3. Numerical experiments

3.1. Setting

The database used in our numerical experiments contains twenty five 128x128 color images from the VisTex database (available at [6]). Given the small size of the images, only two levels of the decomposition with the steerable pyramid were computed. The number of orientations is fixed to four and the number of neighbors in the kNN procedure to ten.

So far, we have described the feature space and similarity measure considering implicitly single channel images (like gray level images). To extend them to the multichannel case, we consider the luminance/chrominance space (Y, Cb, Cr). Since the luminance and chrominance channels are fairly well decorrelated, one can in first approximation consider them independent. Hence, we simply sum the KL divergences obtained for each channel separately.

3.2. Retrieval results

The results for 5 of the images in the database are displayed in Fig. 1. In this figure, each row displays the retrieval result for the example (or reference) image shown on the leftmost column. From the second column on, one can see the first three images in the database ranked by our similarity distance (the leftmost, the most similar), excluding the example image (which is always at a distance of zero).

In general, our method seems to perform very well. In particular, images coming from the same scene (see rows 2 to 4 in Fig. 1) are ranked first. In this database such images are usually translated versions of one another. Hence this experiment shows that our method is robust to translation. Similar textured images such as trees, grass, and grids are also correctly classified (see the last row of Fig. 1).

3.3. Complexity and computation time

For one query image, the computational cost of the retrieval procedure is the number of image in the database times the cost of computing a similarity between two images. Denoting by N the number of pixel in an image, the similarity computation is made in three steps of complexity:

- $O(\sqrt{N})$ for the steerable transform,
- $O(N)$ for the design of the wavelet neighborhoods,
- $O(N \log N)$ for the evaluation of kNN distances (via a classical KD-tree implementation).

Accordingly, most of the computational effort is put in the evaluation of the kNN distances. To improve this, we reduce our feature space by selecting a small proportion of the neighborhoods to evaluate the KL divergences. We select those with the largest central coefficient, thus exploiting

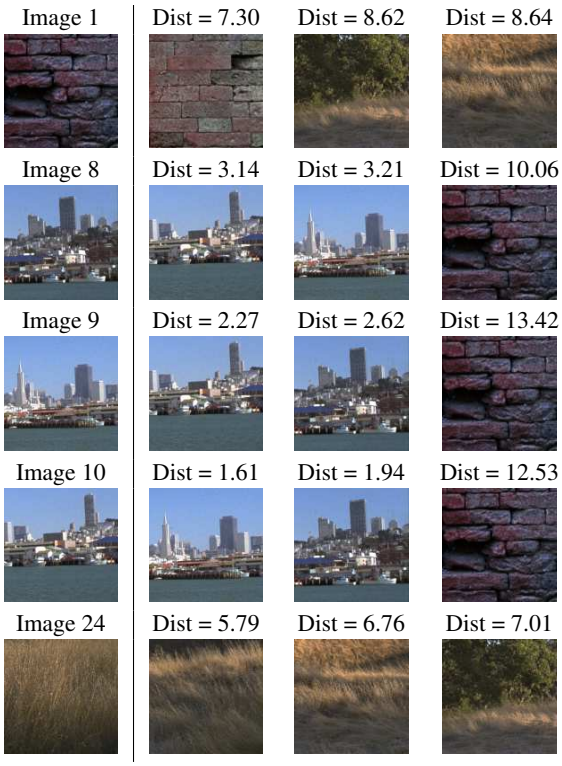

Figure 1. Retrieval results. Left to right: reference image; first 3 ranked images.

the sparsity of wavelet representation. Fig. 2. shows how the computational time evolves then in $O(M \log M)$ where M is the number of selected coefficients (green curve with circles) while the similarity measure remains consistent (the similarity between image 9 and its 3 closest matches are displayed). Selecting only 1/32 of the coefficients leaves us with results of the same accuracy as with all coefficients while greatly reducing the computation time.

4. Conclusion

In this paper, we proposed a similarity measure between images based on the KL divergence between multidimensional pdf of wavelet coefficients grouped in coherent sets called neighborhoods. The KL divergence is estimated nonparametrically via a kNN approach.

Experiments on small images show good performances of the proposed measure in the retrieval problem, particularly its robustness to simple geometric tranforms and to the sparsity of the feature space. Future works will focus on dataset with larger images as well as evaluation of the performances through recall-precision curves.

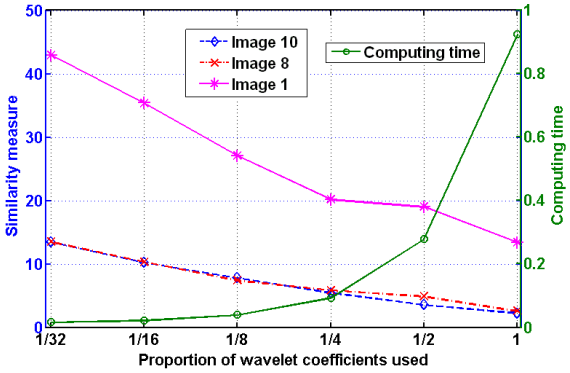

Figure 2. Evolution of similarity and computing time with proportion of coefficients used.

References

[1] I. Ahmad and P.-E. Lin. A nonparametric estimation of the entropy absolutely continuous distributions. *IEEE Trans. Inform. Theory*, 22:372–375, 1976.

[2] S. Boltz, E. Debreuve, and M. Barlaud. High-dimensional kullback-leibler distance for region-of-interest tracking: Application to combining a soft geometric constraint with radiometry. In *CVPR*, Minneapolis, USA, 2007.

[3] M. Do and M. Vetterli. Wavelet based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Trans. Im. Proc.*, 11:146–158, 2002.

[4] M. Goria, N. Leonenko, V. Mergel, and P. Novi Inverardi. A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *J. Nonparametr. Stat.*, 17:277–298, 2005.

[5] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. Object-based mpeg-2 video indexing and retrieval in a collaborative environment. *Multimed. Tools Appl.*, 30:255–272, 2006.

[6] MIT. Vision and modeling group. http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html.

[7] E. Pierpaoli, S. Anthoine, K. Huffenberger, and I. Daubechies. Reconstructing sunyaev-zeldovich clusters in future cmb experiments. *Mon. Not. Roy. Astron. Soc.*, 359:261–271, 2005.

[8] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli. Image denoising using a scale mixture of Gaussians in the wavelet domain. *IEEE Trans. Im. Proc.*, 12:1338–1351, 2003.

[9] J. Romberg, H. Choi, and R. Baraniuk. Bayesian tree-structured image modeling using wavelet-domain hidden markov models. *IEEE Trans. Im. Proc.*, 10:1056–1068, 2001.

[10] D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, 1992.

[11] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik. Quality-aware images. *IEEE Trans. Im. Proc.*, 15:1680–1689, 2006.

[12] Q. Zhang and E. Izquierdo. Optimizing metrics combining low-level visual descriptors for image annotation and retrieval. In *ICASSP, Toulouse, France*, 2006.