Similarity measures between feature maps Application to texture comparison

Stéphane HERBIN

Abstract— **This paper presents a framework for the comparison of patterns based on the statistical analysis of spatial relations between feature values. The main contribution is the application of an asymptotic behavior study of the likelihood ratio Test on an Active Process of Exploration ("TAPE" approach). The similarity measure is defined as the error speed of convergence towards zero which can be computed exactly thanks to techniques of large deviation theory.**

A special attention is given to an application of this approach to texture comparison and to a comprehensive evaluation on Brodatz patches.

Keywords— **texture comparison, image retrieval, pattern similarity measure, large deviation theory, Markov chain.**

I. Introduction

This paper is devoted to the definition of a similarity measure between textures considered as feature maps, i.e. sets of points on a regular lattice labelled with a local measure — a feature.

A first example of a feature map object used in pattern recognition or image processing is the patch obtained after image quantization or filtering. The feature values are local measures of the physical signal and usually weigh some quantity in relation to a given scale.

Another example of a feature map is the labelled mosaic used in image segmentation. The segmentation process reveals the underlying structure of an image in terms of regions of similar value: the local feature is an index.

In the following, no assumption on any ordinal relation between the feature values will be made. This elementary setting will allow us to handle in a very general way the structural distribution of feature values on a map, be they indices or quantities. We use quantization as a way of assigning an index to numerical values.

A. The "TAPE" approach

A pattern will be understood as a source of localized feature values. It is analyzed by a random spatial process of exploration collecting a series of values. The dissimilarity measure is based on the statistical differences of the series collected on two different patterns by the same process of exploration. What will be characterized is the difference of the process and pattern *interaction* behaviors when exploring two patterns.

To build the dissimilarity measure, we construct a *virtual* empirical comparison of two patterns by the computation of a test aiming at distinguishing the statistical differences between the collected feature values. The measure is defined as the logarithmic speed of convergence of this test as the time of exploration goes to infinity.

The acronym TAPE (for "Test of an Active Process of Exploration") is defined to designate this approach.

B. Results

The main contribution of this paper is a general framework for the construction of a similarity measure between feature maps based on techniques of large deviation theory of empirical processes. The measure is a rate of convergence of a likelihood ratio Testing procedure of an Active Process of Exploration ("TAPE" approach).

Texture patches are interpreted as a special case of feature maps where one may consider pixel grey level as the extracted features. In this context, co-occurence matrices are interpreted as controlled Markov chain probability transitions; the TAPE approach provides a natural setting for combining them in a global measure.

Based on a comprehensive experimental evaluation on Brodatz textures [1], the simplest form of the approach compares favorably with other schemes such as the popular wavelet coefficient histogram comparison.

C. Related work

Texture classification have been extensively studied in image retrieval problems [2]. Most of the approaches are based on a first feature extraction step followed by a comparison using some pre-defined metric.

The statistical study of co-occuring features for texture analysis has been conducted in various directions. Co-occurence matrices [3] are among the most popular. Gray-level difference histograms have been studied in [4] and compared to other approaches. [5] studies more complex feature maps applied to global texture orientation detection.

Techniques of large deviation theory of empirical processes have been used in few papers in the computer vision litterature. [6] analyses random walks on the view sphere modelled as Markov chains as a basis for 3D object recognition. [7] defines a parameter computed from local statistics to assess the detectability of curves in an image. [8] proposes an interpretation of visual pop-out phenomena using Kullback-Leibler divergence rate between empirical histograms of filtered textures.

D. Organization of the paper

Section II describes the principles of the similarity measure construction. Section III applies the technique to tex-

ONERA, Département Traitement de l'Information et Modélisation, 29, avenue de la Division Leclerc, BP 72, 92322 Chatillon Cedex, France, Email: Stephane.Herbin@onera.fr

ture comparison, leading to a probabilistic interpretation of co-occurence matrices. Experimental results on texture classification and image retrieval are shown in section IV.

II. Measure construction

This section describes the formal steps leading to the construction of a similarity measure between feature maps. The specific application to texture classification is delayed until the next section.

The main steps are the following:

1. Define a random process dedicated to the partial exploration of the object to be analyzed. The process generates a sequence of local feature values.
2. Calculate the likelihood of the sequence of feature values so gathered, based on a probabilistic model.
3. Compute an index describing the asymptotic behavior of the likelihood ratio test.

A. Feature map exploration

A feature map is a variety of colored graph, where each color is assumed to be a value of a given feature. Any kind of graph topology, as long as it is connected, is suitable to the same kind of approach: we specialize however the problem to regular lattices or pyramids, as it is usual in image processing and computer vision, and to a finite number of feature values.

To explore the map, we define a virtual trajectory of controlled states $\{s_0, u_1, s_1 \dots u_t, s_t\}$ where each state s_t, generated by control u_t is decomposed into a location x_t and a local feature m_t : $s_t = [m_t, x_t]$.

The control variable u_t plays two exclusive roles: it changes either the location, either the parameters of the feature extraction. We assume that each location is defined by a spatial position and a set of feature extraction parameters. The construction of the dissimilarity measure is based on the probabilistic behavior of the sequence of feature values for a given sequence of control variables.

In the case of textured images, the position is a pixel index and the feature a grey level obtained after image filtering and quantization, for instance. Typical values of control variables are moves of a given number of pixels in one direction, or change of a filter index according to a fixed bank of filters. In the following, we will be essentially interested in the spatial arrangement of features, and will consider a single identity filter returning the quantized pixel value.

B. Marginal feature value sequence as Markov chain

The sequences of positions for a given command are defined by an underlying Markov chain, common for all the maps. The feature value sequence behavior, however, is specific to the type of feature map observed: we are therefore interested in qualifying specifically the type of dependencies between the features themselves, the "surface" process.

When assuming 1) that the next x_t location depends only on the current one and on the command (Markov controlled chain hypothesis) and 2) that the choice of the next command depends only the current feature value, not on the underlying location or on the past (Feature-only dependent command), the joint variable $[m_t, u_t]$ can be modelled as a Markov chain with probability transition:

$$\mathrm{P}[m_t, u_t \,|\, m_{t-1}, u_{t-1}] = \mathrm{P}[m_t \,|\, m_{t-1}, u_t].\mu[u_t \,|\, m_{t-1}] \quad (1)$$

where μ is the command law.

C. Discrimination of Markov chains

The definition of a similarity measure between feature maps is now translated into the problem of comparing the probabilistic behavior of control/feature sequences modelled as Markov chains.

A standard way of discriminating the probabilistic behavior of two processes is by computing their likelihood ratio [9]. In the case of Markov chain models, the likelihood $\mathrm{l}_T(H_k)$ of a control/feature sequence of length T for a hypothesis H_k is defined as:

$$\begin{aligned} \mathrm{l}_T(H_k) &= \mathrm{P}[m_T, u_T, m_{T-1}, u_{T-1}, \dots m_0 \,|\, H_k] \\ &= \mathrm{P}_0 \prod_{t=1}^{T} \mathrm{P}[m_t \,|\, u_t, m_{t-1}; H_k] \prod_{t=1}^{T} \mu[u_t \,|\, m_{t-1}] \end{aligned}$$

where it has been made apparent above that it consists of two parts: the sequence of feature values, characteristic of the feature map observed, and the sequence of controls governed by the command μ, common to all the hypothesis.

The likelihood ratio test (LRT) consists in computing the log-ratio

$$\mathrm{L}_{01}^T = \frac{1}{T} \log \frac{\mathrm{l}_T(H_0)}{\mathrm{l}_T(H_1)} \quad (2)$$

for two competing hypotheses (i.e feature maps) H_0 and H_1 and in comparing it to a fixed threshold. A zero threshold means that we do not favor a priori any of the hypotheses. The LRT depends on the two Markov transition matrices describing the measurement transitions (1) for the two feature maps and on the command law μ.

The LRT generates discrimination errors that goes to zero exponentially fast when $T \to \infty$ with a rate that can be explicitly computed given the Markov chain transition matrices using large deviation techniques [10].

The bigger this rate, the easier it is to discriminate empirically the two hypotheses. The LRT's logarithmic speed of convergence characterizes therefore the difficulty of discriminating two feature maps based on Testing an Active Process of Exploration (TAPE), and will be interpreted as the similarity measure studied in this paper.

It can be shown that the log-likelihood ratio test (2) converges as:

$$\lim_{T\to\infty} -\frac{1}{T} \log \mathrm{P}[\mathrm{L}_{01}^T > 0 \,|\, H_0] = \tau_{01} + \rho_{01} \quad (3)$$

where

1. τ_{01} is a rate characterizing the difference between the relative frequencies of transition occurence for two Markov chains with positive probabilities.

2. ρ_{01} is a coefficient measuring the *structural* differences between transitions with probability zero. [1]

We define the TAPE similarity measure as the sum of these two terms which can be explicitly computed given the two Markov matrices. It is a semi-metric: null if and only if the two Markov chains have identical transition probabilities, and strictly positive whenever they differ. The triangular inequality, however, does not hold.

III. Texture discrimination

Co-occurence matrices are formal objects dedicated to summarize the number of grey level transitions between two pixels separated by the same displacement. Given an image $\mathcal{I}(\mathbf{p})$ taking values in the set of grey values $\{g_1, g_2, \dots g_N\}$, the un-normalized co-occurence matrix for displacement $\mathbf{v}$ is a two-dimensional array $C(i, j; \mathbf{v})$ where the coefficient at the i-th row and j-th column is the counting measure of the number of pixel pairs $[\mathbf{p}, \mathbf{p} + \mathbf{v}]$ in the image where $\mathcal{I}(\mathbf{p}) = g_i$ and $\mathcal{I}(\mathbf{p} + \mathbf{v}) = g_j$.

Define a normalized version $\bar{C}(i, j; \mathbf{v})$ of the above matrix by dividing each line by the sum of its coefficients. This new matrix can now be interpreted as the probability of observing a pixel with value g_j when coming from a pixel with value g_i and a displacement of $\mathbf{v}$. The coefficient $\bar{C}(i, j; \mathbf{v})$ can also be interpreted as the Markov probability transition between feature values g_i and g_j for a command value equal to $\mathbf{v}$.

A random sequence of displacements u_t generates a random sequence of feature values m_t specific to the texture. If one chooses the displacements according to a sampling law conditionned by the currently observed feature value $\mu(u_t = \mathbf{v} \,|\, m_t = g_i)$, the joint sequence of feature values and displacements $[m_t, u_t]$ defines a homogeneous Markov chain as described in the previous section. Their probability transition is equal to

$$\mathrm{P}[m_t = g_j, u_t = \mathbf{v} \,|\, m_{t-1} = g_i, u_{t-1} = \mathbf{w}] = \bar{C}(i, j; \mathbf{v})\mu(\mathbf{v} \,|\, g_i) \tag{4}$$

showing the dependence on the choice of the random displacements. The combination of the co-occurence matrices results in a big stochastic matrix with probability transitions defined by equation 4, each finite state being a couple value $\times$ displacement.

It is often claimed that one of the difficult aspect of the practical use of co-occurence matrices lies in the choice of the right displacement and in the selection of the good features computed from it [11]. Indeed, it is common to produce another formal step by computing a series of features that will constitute the operative texture representation used for discrimination [12].

The TAPE scheme applied to co-occurence matrices does not depend on any feature extraction step. It keeps all the information contained in the original signal, and is sensitive only to the features that really “make the difference”. Indeed, the likelihoods filter their contribution to the measure and make neglictable similar features: there is no need here to seek the most informative ones in a preliminary feature selection step.

IV. Experimentation

The TAPE similarity measure is tested on problems of texture classification and image retrieval. In order to compare with other approaches, we use the same experimental protocols as [13] for classification and [14] for retrieval.

A. Classification

The protocol defined in [13] consists in extracting from the Brodatz examples [1] 16 non overlapping sample patches of various sizes (8×8, 16×16, 32×32, 64×64, 128×128) [2] and computing all the dissimilarity measures between patches.

Each of the 87 texture classes is defined by the set of samples extracted from the same image. The classification performance is evaluated by searching for each patch its nearest neighbor according to the TAPE similarity measure in a leave-one-out scheme. The classification succeeds if the nearest neighbor and the patch tested belong to the same original image, and fails otherwise.

The exploration strategy consists in walking randomly from one pixel to another according to a 4-connectedness neighborhood, the law of direction selection being uniform. Its probabilistic behavior is modelled by the calculation of four co-occurence matrices, with displacements $[0, 1]$, $[1, 0]$, $[-1, 0]$, $[0, -1]$. The feature extracted at each visited pixel is the quantized grey level using 8 or 16 quantization levels.

The classification errors are plotted on Fig. 1. There is no clear differences in the two behaviors when using 8 or 16 quantization levels. As expected, the errors decrease with the patch size. When comparing the numerical values with the experiments in [13] based on histogram comparison, we observe that the decrease function of the sample size is steepest. The errors are larger for sample sizes 8×8, 16×16 and smaller for sizes 32×32, 64×64 and 128×128.

The TAPE measure seems to be more sensitive to the estimation of the transition probabilities. It performs well or even better than the dissimilarity measures tested in [13] when a sufficient number of pixel pairs are used for the estimation of the co-occurence matrices, but degrades when the estimation confidence cannot be controlled.

[1] If P_0 and P_1 are the matrices of Markov transition probabilities, the first coefficient is given by

$$\tau_{01} = \sup_{x \in [0,1]} (-\log \rho(P_0^{(1-x)} * P_1^x))$$

where ρ is the operator computing the largest eigen value of a square matrix. The power and $*$ operators are element-wise. The second coefficient is given by

$$\rho_{01} = -\log \rho(P_0 * \mathbf{1}_{(\mathbf{P_1} \neq \mathbf{0})})$$

where $\mathbf{1}_{(\mathbf{P_1} \neq \mathbf{0})}$ is the square matrix containing 1 if the corresponding transition is positive, 0 otherwise. Refer to [6] for a more detailed presentation.

[2] As in [13], we selected from the Brodatz album through a priori visual inspection 87 images essentially characterized by homogeneous micro-pattern textural properties, resulting in $1392 = 16 \times 87$ patches for each size. The selected images are D1, D3-4, D6-13, D15-22, D24, D26, D28-29, D32-38, D46-47, D49-58, D60, D62-75, D77-87, D90, D92-93, D95-96, D98, D100-112.

Another source of error lies in the nature of the textures themselves: Figure 2 shows pairs of textures from which several patches of size 32×32 have low similarity measure values. Those textures exhibit non stationnary behavior at the patch size scale which cannot be captured correctly by a global co-occurence matrix. Some patches look very similar although extracted from different images.

For large patch sizes like those on Fig. 3, several pairs of different textured images may contain areas with low dissimilarity measure values. The reason is that, at this scale of observation, the images are globally undistinguishable.

As shown in most of the experiments on texture databases, the central point of texture comparison is the statistical estimation of feature value spatial dependencies, be they implicitly revealed by spatial filtering or directly evaluated on the image, as in this study.

B. Image retrieval

In the image retrieval application, the goal is to recall, given a sample, all the patches extracted from the same image. The performance of the retrieval process is measured by computing its *precision*, i.e. the number of relevant images compared to the number of retrieved images.

The performances of the TAPE dissimilarity measure are compared with those presented in [14] where textures are analyzed by Gabor filters. The same set of 87 Brodatz textures as above are sampled in 16 non overlapping patches of sizes 8×8, 16×16, 32×32, 64×64 and 128×128. The precision is measured as the ratio to 15 of the number of patches belonging to the same image in the top 15 most similar patches among the $1391 = 16 \times 87 - 1$ other elements in the database. The same 4-neighbor — 8/16-quantization-level strategies are used.

The empirical precisions are plotted on Fig. 1. There is no clear difference between the two strategies based on 8 or 16 quantization levels. As expected, the precision increases with the patch size. The precision for a patch size of 128×128 is equal to 77.7%, which is marginally better than the 74.4% performance claimed in [14].

On patches of size 128×128 with 8 quantized levels, the computation of one similarity measure takes 9.8 ms on a SPARC Ultra, divided into 6 ms for the computation of the two co-occurence matrices and 3.8 ms for the asymptotic rate itself.

V. Conclusion

This paper has presented a general approach for the construction of similarity measure between feature maps. It is based on an application of the theory of large deviation of empirical processes to Markov chain discrimination.

The basic setting presented in this paper already generates good performances compared to other classical approaches such as wavelet or Gabor filter based texture representations, without any preliminary feature selection step or image renormalization. Furthermore, it provides an elegant way of combining several co-occurence matrices by introducing an action law used to sample the displacements.

The undelying goal pursued in the approach presented here was not to design a new set of general features able to represent textures "better", but instead to study how a given set of features can be specifically combined in a way that preserves their spatially arranged nature. In other words, the TAPE approach defended in this paper has to be understood as the exemplar of a tentative paradigm shift, from global universal representation to local purposive dynamic analysis.

References

[1] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, Dover, New York, 1966, http://www.ux.his.no/~tranden/brodatz.html

[2] S. Aksoy and R.M. Haralick, "Using texture in image similarity and retrieval," in *Texture Analysis in Machine Vision*, M. Pietikainen, Ed., vol. 20, pp. 129–149. World Scientific, Singapore, 2000.

[3] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, November 1973.

[4] T. Ojala, K. Valkealahti, E. Oja, and M. Pietikäinen, "Texture discrimination with multidimensional distributions of signed gray level differences," *Pattern Recognition*, vol. 34, no. 3, pp. 727–739, March 2001.

[5] D. Chetverikov, "Texture analysis using feature based pairwise interaction maps," *Pattern Recognition*, vol. 32, pp. 487–502, 1999.

[6] S. Herbin, "Combining geometric and probabilistic structure for active recognition of 3D objects," in *European Conference on Computer Vision*, Berlin, 1998, vol. 1407 of *Lecture Notes in Computer Science*, pp. 748–764, Springer Verlag.

[7] A.L. Yuille and J.M. Coughlan, "Fundamental limits of bayesian inference: Order parameters and phase transitions for road tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 2, pp. 160–173, February 2000.

[8] Y.N. Wu, S.C. Zhu, and X.W. Liu, "Equivalence of julesz ensembles and FRAME models," *International Journal of Computer Vision*, vol. 38, no. 3, pp. 245–261, July 2000.

[9] E. Lehmann, *Testing Statistical Hypotheses*, John Wiley & sons, New York, 1959.

[10] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett Publishers, Boston, 1993.

[11] M. Tuceryan and A.K. Jain, "Texture analysis," in *Handbook of Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and P. S. P. Wang, Eds., chapter 2.1, pp. 207–248. World Scientific Publishing, 2nd edition, 1998.

[12] R. Haralick, "Statistical image texture analysis," in *Handbook of Pattern Recognition and Image Processing*, vol. 86, pp. 247–279. Academic Press, 1986.

[13] J. Puzicha, Y. Rubner, C. Tomasi, and J. Buhmann, "Empirical evaluation of dissimilarity measures for color and texture," in *International Conference on Computer Vision*, September 1999, pp. 1165–1173.

[14] B.S. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–42, August 1996.

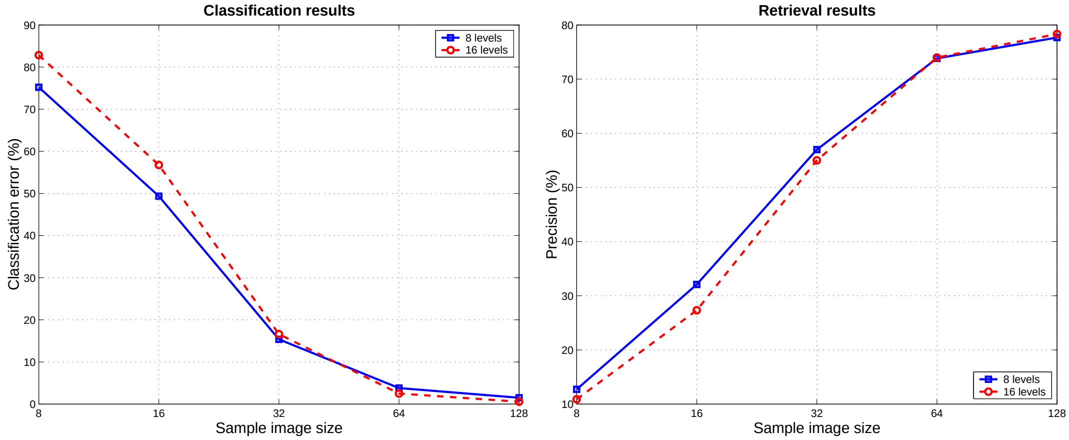

Fig. 1. Texture classification (left) and image retrieval (right) results for 8 and 16 quantization levels.

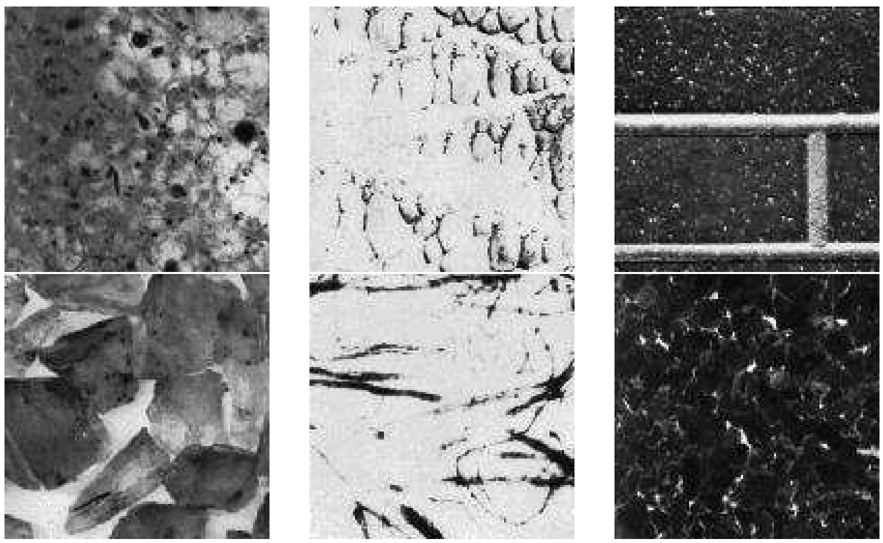

Fig. 2. Couples of images in which small patches are similar. (Brodatz images D73/D98, D10/D108, D26/D33 cropped to 128×128 pixels.)

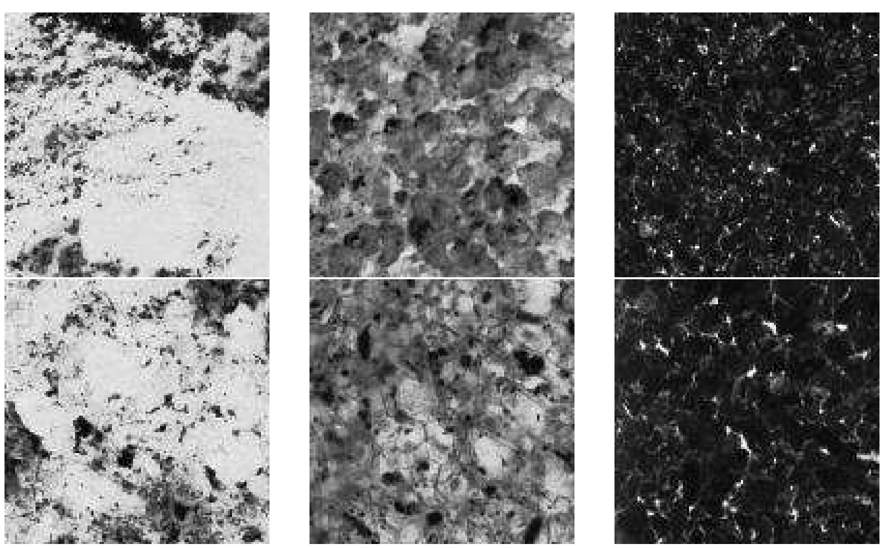

Fig. 3. Couples of similar 128×128 textured patches. (Brodatz images D7/D60, D28/D73, D32/D33.)