A FRACTALS-INSPIRED APPROACH TO CONTENT-BASED IMAGE INDEXING

Mathieu Vissac[†], *Jean–Luc Dugelay*[†] *and Kenneth Rose*[††]

[†]Institut Eurécom, Dept. of Multimedia Communication
06904 Sophia-Antipolis Cedex, France
Tel : +33 (0)4.93.00.26.{41.71}, Fax : +33 (0)4.93.00.26.27
Email : {vissac, dugelay}@eurecom.fr
[††]Dept. of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106, USA
Email : rose@ece.ucsb.edu

ABSTRACT

This paper applies ideas from fractal compression and optimization theory to attack the problem of efficient content-based image indexing and retrieval. Similarity of images is measured by block matching after optimal (geometric, photometric, etc.) transformation. Such block matching which, by definition, consists of localized optimization, is further governed by a global dynamic programming technique (Viterbi algorithm) that ensures continuity and coherence of the localized block matching results. Thus, the overall optimal transformation relating two images is determined by a combination of local block-transformation operations subject to a regularization constraint. Experimental results on a sample of seventy five binary images from the MPEG-7 database demonstrate the power and potential of the proposed approach.

1. INTRODUCTION

Recent years have seen a dramatic increase in the size and variety of digital image databases. Huge volumes of image data accumulate and are stored for diverse applications. These include medical imagery, satellite imagery, entertainment image data, etc. As manual search and retrieval of images in large databases is impractical, efficient automatic access techniques are needed. The problem is exacerbated by the need for content-based retrieval rather than the alphanumeric search typical of traditional database management. We assume the basic problem of query by example. The user presents a "query image", also called the "example image", and the algorithm searches the database (of so-called "test images") for images that are most similar to the query image. The query by example is a major requirement of users as it circumvents the need to specify the query in words, which is often hard or even impossible. Moreover, it eliminates the need for a costly preliminary stage of manual alphanumeric indexing to enable future queries. Such manual indexing must also forsee and account for all types of future queries – an impractical requirement in many applications.

Content-based retrieval is a major challenge that has been recognized by many researchers and developers [1, 2]. Several methods have been used : color histograms [3, 4], shape retrieval [5], texture indexing [6, 7] and others [8, 9].
Several products are currently available on the internet, and offer good performance when queries are well represented by color indexing via histograms, etc. Nevertheless, the main difficulty is due to the wide diversity of users and queries whose needs are not satisfied by such simple search techniques. The human notion of similarity is subjective and hard to define precisely. It depends on the type of database, the context, the application and the user.

In this paper we propose to define a new distance between pictures, by applying local similarity principles from the fractal coding theory [10]. This approach to formulating similarity offers much flexibility, to represent the subjectivity of queries, and the premise of this work is that it is a useful tool for approximating the subjective aspects of an image database query. In most fractal-based image indexing algorithms, such as [11, 12], images are divided into blocks, and decisions are made independently for each block. In this work, however, the fractal local similarity measure is complemented by a regularization constraint that enforces coherence of consecutive matching decisions. The regularization constraint is globally and efficiently imposed by a dynamic programming procedure which is commonly known as the Viterbi algorithm [13, 14].

In section 2, we introduce the concept of local similarity based on fractal coding theory. To maintain global coherence we employ the Viterbi algorithm. For this purpose we first define states and transition probabilities as described in section 3. Preliminary results for an MPEG-7 image database, are presented in section 4.

2. LOCAL SIMILARITY

As is commonly done in fractal image compression, the example image is partitioned into blocks that are referred to as range blocks. Block matching is then performed to associate (range) blocks of the query image with (domain) blocks of a candidate image in the database, by adjusting the transformation parameters. In our experiments, the domain blocks have the same size than the range blocks, but the basic framework allows for accomodating zoom effects as well. We consider the geometric parameters $(a_i, b_i, c_i, d_i, e_i, f_i)$:

$$\tau_i . \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} a_i & b_i & 0 \\ c_i & d_i & 0 \\ 0 & 0 & 1 \end{pmatrix} . \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} e_i \\ f_i \\ 0 \end{pmatrix} . \tag{1}$$

One may notice that the luminance parameters (denoted by o_i and s_i in the fractal compression literature) do not appear. This is because our main concern here is with binary pictures. Work is currently under way to extend and optimize the approach to natural gray-level and color images (or multi-spectral images).

Thus, for each block of the example image, and for each possible geometric transformation, we search for the K most similar blocks in the test image (more on the choice of K in subsectionc 3.1). "Most similar" blocks are defined as the ones that minimize the matching cost:

$$\Delta(n,i) = \min_j \left(\sqrt{\frac{\sum_{pixels} (S_n - \tau_i . D_j)^2}{\# \text{ pixels}}} \right) \tag{2}$$

where S_n is the nth block of the example image, D_j is the jth block of the test image and τ_i is the ith transformation.

These matching costs are converted into a probabilistic representation by a function f which, for simplicity, we select to be the linear function : $f(x) = 1 - \frac{x}{255}$). (A future implementation will define f as a half Gaussian.)

Specifically, block n is in state i with probability:

$$O_n(i) = f(\Delta(n,i)) \tag{3}$$

It should be emphasized that while the above formulation of similarity offers substantial flexibility which may capture non-trivial notions of similarity, it may entail too many degrees of freedom. It is conceivable that non-similar images might "match" by mixing and transforming unrelated blocks (a phenomenon we refer to as the "jigsaw puzzle" effect). In order to eliminate this shortcoming we introduce a requirement of "global coherence". In other words, we wish to impose a degree of continuity on the block matching results (geometric transform continuity, spatial continuity, etc). We propose to optimize the trade off between local matching scores and the global coherence constraint by application of a dynamic programming technique known as the Viterbi algorithm [13, 14].

3. GLOBAL COHERENCE

The Viterbi algorithm (VA) finds the optimal path in the trellis of a Markov chain (i.e., a random state machine). In our case, VA is used to find the optimal sequence of block matching pairs for the given pair of example and test images. Since the optimality and simplicity of VA requires a one-dimensional Markov chain, we use a zigzag scan of the image (from top to bottom). This path ensures that consecutively scanned blocks are neighbors on the image plane.

The jth state value of the current block n is given by,

$$E_n(j) = \max_i(E_{n-1}(i) \times P(j|i) \times O_n(j))$$

Where,
$E_{n-1}(i)$ is the ith state value of the block n-1.
$O_n(j)$ is the simple probability for state i and the block n.
$P(j|i)$ is the i-to-j probability.

3.1. The States

The states enumerated below were selected for the case of binary images on which we focus in the results section, but can be extended to the case of gray-level, color or multi-spectral images. The set of states is the product of two sets of sub-states as follows.

First, we define eight sub-states corresponding to the eight geometric transforms : Identity, horizontal and vertical reflections, first and second diagonal reflections and rotations of $\frac{\pi}{2}$, π and $\frac{3\pi}{2}$.

To combat the jigsaw puzzle effect, we add K sub-states, to each of the above states. These sub-states correspond to K possible locations of the test block. The Viterbi algorithm will eventually select the optimal state for each block so as to achieve global geometric, photometric and spatial coherence. Thus, for each block and each geometric transform, we keep the K best (most probable) test-block locations. For example, in Figure 1, we would keep the K highest peaks. It should be noted that the choice of K location sub-states represents a complexity compromise. For optimality, one should consider all possible test-block locations, but that would result in an impractical number of states. Restricting to the best K locations is a form of pruning that is minimally suboptimal but maintains manageable complexity.

3.2. The transition probability

Transition between states is penalized (by assigning an appropriate level of probability) to underline the importance of continuity (geometric and spatial). In the current experiments, we used a three-level cost (no transition, soft transition and abrupt transition) for geometric continuity, and a gaussian probability density function (see figure 2) for implementing spatial continuity constraints. However, optimized selection of the coherence cost is under current investigation.

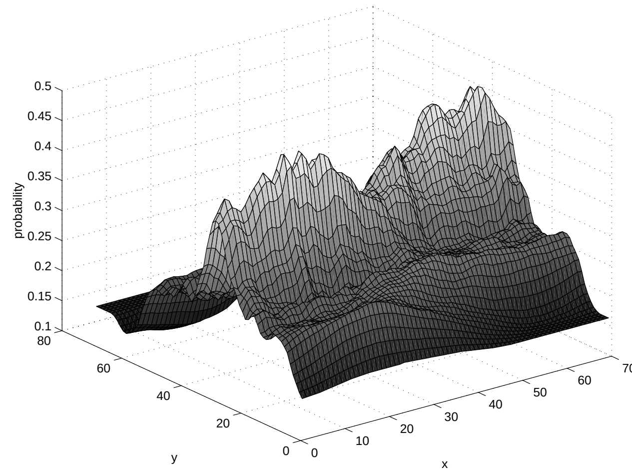


Figure 1: Simple probability

4. EXPERIMENTAL RESULTS

Preliminary experimentation has been performed on a subset of the MPEG-7 database: (*Trademark images captured by a scanner*, CD-ROM no. 10). The set of images consists of four logos, each appearing in nine variations with various degrees of degradation and modification. In addition, there are thirty nine different logos.

As input to our algorithm we provided a logo from the database. For each original logo, the algorithm performed the search and listed the database images in decreasing order of similarity to the query image. As shown on the logo “BAYER” example in Figure 3, our approach retrieves first in the list the query image itself followed by its eight modified logos, and that result is

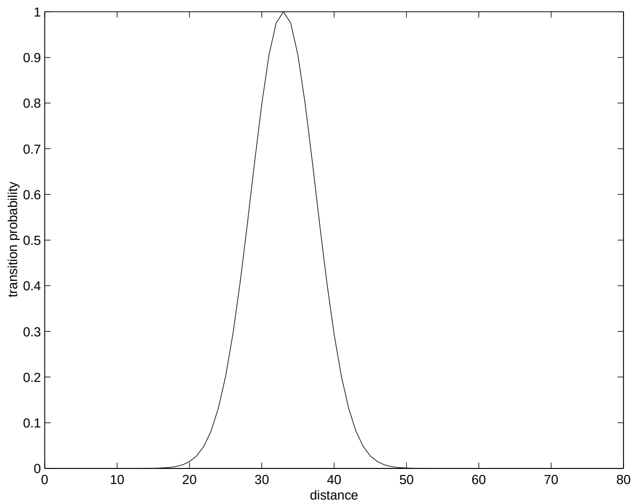


Figure 2: Spatial transition probability

verified in each of the queries carried out on the four logos considered in any of the 9 variations.

5. CONCLUDING REMARKS

A novel image indexing and retrieval method was presented. This method is based on calculation of local similarity while allowing for a variety of geometric and photometric transformations, which is complemented by a global coherence constraint that ensures geometric, photometric, and spatial continuity. Preliminary results demonstrate the method's robustness to various local and global transformations and modifications of the image. The proposed approach shows much promise as a means to capture the subjective notion of similarity.

Our algorithm can be adapted to handle other types of image modification, queries and applications. The use of Viterbi algorithm allows easy extension via a richer set of states including new transformations. Future works will include: (i) implementation of additional states for natural grey-level and color images,

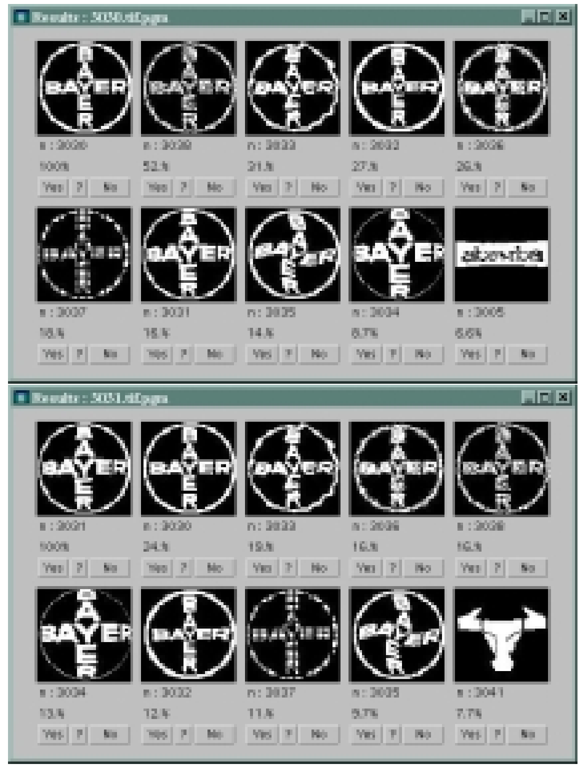
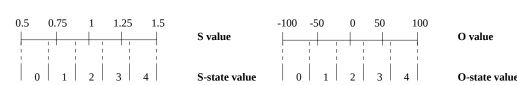


Figure 3: Example results

(ii) optimization of the test image sweep code to further reduce the computational complexity, and (iii) implementation of a training system to optimize the choice of system parameters.

For gray-level images, in particular, several substates will be added to represent photometric parameters and to quantify the tolerance to luminance variation. We expect that transitions between these substates should not be as heavily penalized as in the geometric case, as they are not believed to represent a necessary condition of similarity.



6. ACKNOWLEDGMENTS

This work was supported in part by a regional grant, and by Eurécom's industrial partners: Ascom, Cegetel, France Telecom, Hitachi, IBM France, Motorola, Swisscom, Texas Instruments, and Thomson CSF. K. Rose was supported in part by the University of California MICRO program, Cisco Systems, Inc., Conexant Systems, Inc., Dialogic Corp., Fujitsu Laboratories of America, Inc., General Electric Co., Hughes Network Systems, Intel Corp., Lernout & Hauspie Speech Products, Lockheed Martin, Lucent Technologies, Inc., Panasonic Technologies, Inc., Qualcomm, Inc., and Texas Instruments, Inc.

7. REFERENCES

[1] T. Randen and J.H. Husoy. Image content search by color and texture properties. In *ICIP-97*, volume 1, pages 580–583, 1997.

[2] H. Yu and W. Wolf. A hierarchical, multi-resolution method for dictionary-driven content-based image retrieval. In *ICIP-97*, volume 2, pages 823–826, 1997.

[3] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[4] J. R. Smith and S.-F. Chang. Single color extraction and image query. In *IEEE International Conference on Image Processing*, pages 528–531, 1995.

[5] C. Nastar. The image shape spectrum for image retrieval. Technical Report 3206, INRIA, July 1997.

[6] J. R. Smith and S.-F. Chang. Automated binary texture feature sets for image retrieval. In *IEEE ICASSP'96*, pages 2239–2242, May 1996.

[7] C. Li and Castelli V. Deriving texture feature set for content-based retrieval of satellite image database. In *ICIP-97*, volume 1, pages 576–579, 1997.

[8] J.M. Marie-Julie and H. Essafi. Image indexing by using rotation and scale invariant partition. *ECMAST'97*, pages 163–175, 1997.

[9] M. Beatty and Manjunath B.S. Dimensionality reduction using multi-dimensional scaling for content-based retrieval. In *ICIP-97*, volume 2, pages 835–837, 1997.

[10] A. E. Jacquin. Image coding based on a fractal theory of iterated contrative image transformation. *IEEE Transactions on Image Processing*, 1(1):18–30, January 1991.

[11] A. Zhang, B. Cheng, and R. Acharya. A fractal-based clustering approach in large visual database systems. *Multimedia Tools and Applications*, (3):225–244, 1996.

[12] J.M. Marie-Julie and H. Essafi. Image database indexing and retrieval using the fractal transform. In *ECMAST'97*, pages 169–182, 1997.

[13] R. Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[14] A. J. Viterbi. Errors bounds for convolutional codes and asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, IT-13:260–269, 1967.