A New Framework to Combine Descriptors for Content-based Image Retrieval *

Ricardo da S. Torres, Alexandre X. Falcão
University of Campinas, Campinas, SP, Brazil
{rtorres, afalcao}@ic.unicamp.br

Marcos André Gonçalves
Federal University of Minas Gerais
Belo Horizonte, MG, Brazil
mgoncalv@dcc.ufmg.br

Baoping Zhang, Weiguo Fan, and Edward A. Fox
Virginia Tech, Blacksburg, VA 24061 USA
{bzhang, wfan, fox}@vt.edu

Pavel Calado
INESC - ID
Portugal
pavel@algos.inesc-id.pt

ABSTRACT

In this paper, we propose a novel framework using *Genetic Programming* to combine image database descriptors for content-based image retrieval (CBIR). Our framework is validated through several experiments involving two image databases and specific domains, where the images are retrieved based on the shape of their objects.

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval[Query formulation, Search process]

General Terms: Measurement, Experimentation

Keywords: Content-based Image Retrieval, Genetic Programming

1. INTRODUCTION

Advances in data storage and image acquisition technologies have enabled the creation of large image databases. In order to deal with these data, it is necessary to develop appropriate information systems which can support different services. The focus of this paper is on CBIR systems. Basically, CBIR systems try to retrieve images similar to a user-defined specification (e.g., image example), according to image content properties (such as color and shape).

Different descriptors encoding different or even the same image properties have been proposed to support image retrieval by content [4]. These descriptors are commonly chosen in a domain-dependent fashion, and, generally, are combined in order to meet users' perception. This paper proposes a novel framework to combine image database descriptors, improving effectiveness in retrieval tasks. This framework is based on an artificial intelligence (AI) optimization technique, called *Genetic Programming (GP)* [5]. We validate the proposed framework in using shape-based image retrieval, through various experiments. Our approach to search for optimal functions to combine descriptors appears to be flexible and powerful [3].

*This work was supported in part by AOL, CNPq, FAPESP (Proc.03/14096-8), Microsoft Escience project, FCT project POSI/EIA/58194/2004, and NSF grants DUE-0121679, DUE-0333531, IIS-0307867, DUE-0136690, IIS-0086227, and IIS-0325579.

CIKM'05, October 31–November 5, 2005, Bremen, Germany.
ACM 1-59593-140-6/05/0010.

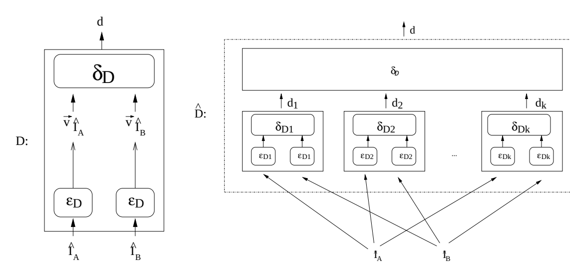
2. CBIR MODEL

In this section, we formalize our CBIR system model.

Definition 1. A **simple descriptor** D is defined as a pair (ϵ_D, δ_D), where: $\epsilon_D : \hat{I} \rightarrow \mathbb{R}^n$ is a function which extracts a *feature vector* $\vec{v}_{\hat{I}}$ from an *image* $\hat{I}$, and $\delta_D : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is a *similarity function* that computes the similarity between two images as the inverse of the distance between their corresponding *feature vectors*.

Definition 2. A **feature vector** $\vec{v}_{\hat{I}}$ of an image $\hat{I}$ is a point in $\mathbb{R}^n$ space: $\vec{v}_{\hat{I}} = (v_1, v_2, ..., v_n)$, where n is the dimension of the vector.

Figure 1(a) illustrates the use of a simple descriptor D to compute the similarity d between two images $\hat{I}_A$ and $\hat{I}_B$.



(a) (b)

Figure 1: (a) Simple and (b) Composite descriptors.

Definition 3. A **composite descriptor** $\hat{D}$ is a pair $(\mathcal{D}, \delta_{\mathcal{D}})$ (see Figure 1(b)), where: $\mathcal{D} = \{D_1, D_2, \ldots, D_k\}$ is a set of k pre-defined simple descriptors, and $\delta_{\mathcal{D}}$ is a similarity function which combines the similarity values obtained from each descriptor $D_i \in \mathcal{D}$, $i = 1, 2, \ldots, k$.

Our solution relies on the creation of a *composite descriptor*, $\hat{\mathcal{D}}_{GP}$, where $\delta_{\mathcal{D}_{GP}}$ is a mathematical expression uniquely represented as a GP expression tree, whose non-leaf nodes are numerical operators and the leaf node set is composed of the similarity values d_i, $i = 1, 2, \ldots, k$.

The entire combination discovery framework can be seen as an iterative process. Starting with a set of training images with known relevance judgments, GP first operates on a population of random combination functions. These combination functions are then evaluated based on the relevance information for the training images. If a stopping criterion is not met, our GP technique will go through the genetic transformation steps to create and evaluate the population of the next generation.

3. EXPERIMENTS

GP System

Terminals: This list includes widely used and recently proposed shape descriptors for comparison purposes — Moment Invariants (MI) [4], Fourier Descriptors (FD) [4], Beam Angle Statistics (BAS) [1] (with 40 and 60 samples), and Contour Multiscale Fractal Dimension (CMFD) [2].

Functions: $+, \times, /, sqrt$.

Fitness Functions: A fitness function measures how effective a combination function represented by an individual tree is for ranking images. A formal definition of the chosen fitness can be found in [3].

Reproduction: We used 0.05 of the population size as the reproduction rate.

Crossover: For crossover, a method called *tournament selection* [5] is used.

Mutation: Our experiments considered 0.25 as the percentage of individuals selected for mutation.

Stopping Criterion: We stop the GP discovery process after 50 generations.

Image Databases: Two different databases have been used to compare the proposed GP-based shape descriptors: (a) Fish Shapes Database: This database contains 100 classes with 10 fish shapes each, obtained by rotating and scaling one hundred fish contours available online[1]; and (b) MPEG-7: shape database of the MPEG-7 project with 70 classes and 20 images each, for a total of 1400 images.

We randomly split the data into training and test parts. The training set used a random 50% sample for each class. We also considered two different samples (S1 and S2) for each data set.

4. RESULTS

Our experiments compare the GP-based approach with a composite descriptor derived from a *Genetic Algorithm (GA)* and the best simple descriptor, taken as baselines. The GA-based descriptor uses a fixed-length sequence of real numbers (weights) to indicate the importance of each descriptor.

We used precision after 10 images are returned as our comparison criterion. Table 1 shows the average precision for each simple descriptor. Note that the BAS60 descriptor yields the best result with both collections.

Table 2 presents the average precision of the GP-based shape descriptors, using different fitness functions. With regard to the MPEG-7 collection, GP-based descriptors outperform the BAS60 baseline. Note also that GP presents a better result when compared to the GA-based descriptor, except for the CHK fitness function, using S2. For the Fish Shapes collection, despite the high effectiveness of the baseline (BAS60 descriptor), the results based on the GP approach are better.

[1] www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html

Descriptor	MPEG-7		Fish Shapes	
	S1	S2	S1	S2
BAS40	65.35	64.84	83.35	81.10
BAS60	**66.27**	**65.37**	**93.25**	**92.30**
CMFD	40.71	40.05	71.35	68.85
FD	20.25	20.44	24.20	23.75
MI	34.68	35.02	63.20	61.45

Table 1: Average precision after 10 returned images, considering the evidence in isolation.

Descriptor	MPEG-7	
	S1	S2
BAS60	66.27	64.84
GP with PAVG@10	70.56 (6.47%)	69.21 (6.74%)
GP with FFP1	**70.92 (7.02%)**	69.59 (7.33%)
GP with FFP2	70.79 (6.82%)	69.76 (7.59%)
GP with FFP3	70.75 (6.76%)	69.44 (7.09%)
GP with FFP4	70.40 (6.23%)	68.97 (6.37%)
GP with CHK	70.73 (6.73%)	66.78 (2.99%)
GP with LGM	70.86 (6.93%)	**70.90 (9.35%)**
GA	69.37 (4.68%)	68.30 (5.38%)
Descriptor	Fish Shapes	
	S1	S2
BAS60	93.25	92.30
GP with PAVG@10	93.75 (0.54%)	92.75 (0.49%)
GP with FFP1	94.20 (1.02%)	93.30 (1.08%)
GP with FFP2	**94.30 (1.13%)**	**93.35 (1.14%)**
GP with FFP3	94.05 (0.86%)	93.30 (1.08%)
GP with FFP4	94.05 (0.86%)	93.30 (1.08%)
GP with CHK	94.20 (1.02%)	93.30 (1.08%)
GP with LGM	94.15 (0.97%)	93.20 (0.98%)
GA	93.40 (0.16%)	92.55 (0.27%)

Table 2: Average precision after 10 returned images, considering the GP-based descriptors.

5. CONCLUSIONS

We considered the problem of combining simple descriptors for CBIR. Our solution uses GP to search for an optimal combination function. The proposed framework was validated for shape-based image retrieval, through several experiments involving two image databases, and many simple descriptors and fitness functions. The effectiveness results demonstrate that the GP framework can find better similarity functions than the ones obtained from the individual descriptors. Our experiments also show better results with GP than using a GA approach. Future work will focus on using automatic mechanisms to incorporate the GP-based descriptors in search engines.

6. REFERENCES

[1] N. Arica and F. T. Y. Vural. BAS: A Perceptual Shape Descriptor Based on the Beam Angle Statistics. *Pattern Recognition Letters*, 24(9-10):1627–1639, June 2003.

[2] R. da S. Torres, A. X. Falcão, and L. da F. Costa. A Graph-based Approach for Multiscale Shape Analysis. *Pattern Recognition*, 37(6):1163–1174, June 2004.

[3] R. da S. Torres, A. X. Falcão, M. A. Gonçalves, B. Zhang, W. Fan, and E. A. Fox. A New Framework to Combine Descriptors for Content-based Image Retrieval. Technical Report IC-021, Institute of Computing, University of Campinas, Campinas, SP, Brazil, September 2005.

[4] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, Reading, MA, USA, 1992.

[5] J. R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA, 1992.