Pattern Recognition 42 (2009) 158 – 166

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Multiple-instance content-based image retrieval employing isometric embedded similarity measure

John Y. Chiang [a,*], Shuenn-Ren Cheng [b]

[a] Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan 80424, Taiwan
[b] Department of Business Administration, Cheng Shiu University, Kaohsiung, Taiwan

ARTICLE INFO

Article history:
Received 16 November 2007
Received in revised form 4 May 2008
Accepted 23 June 2008

Keywords:
Content-based image retrieval (CBIR)
Multiple-instance learning
Fractal orthonormal basis
Diversity density

ABSTRACT

In image-based retrieval, global or local features sufficiently discriminative to summarize the image content are commonly extracted first. Traditional features, such as color, texture, shape or corner, characterizing image content are not reliable in terms of similarity measure. A good match in the feature domain does not necessarily map to image pairs with similar relationship. Applying these features as search keys may retrieve dissimilar *false-positive* images, or leave similar *false-negative* ones behind. Moreover, images are inherently ambiguous since they contain a great amount of information that justifies many different facets of interpretation. Using a single image to query a database might employ features that do not match user's expectation and retrieve results with low precision/recall ratios. How to automatically extract reliable image features as a query key that matches user's expectation in a content-based image retrieval (CBIR) system is an important topic.

The objective of the present work is to propose a multiple-instance learning image retrieval system by incorporating an *isometric embedded* similarity measure. Multiple-instance learning is a way of modeling ambiguity in supervised learning given multiple examples. From a small collection of positive and negative example images, semantically relevant concepts can be derived automatically and employed to retrieve images from an image database. Each positive and negative example images are represented by a linear combination of fractal orthonormal basis vectors. The mapping coefficients of an image projected onto each orthonormal basis constitute a feature vector. The Euclidean-distance similarity measure is proved to remain consistent, i.e., *isometric embedded*, between any image pairs before and after the projection onto orthonormal axes. Not only similar images generate points close to each other in the feature space, but also dissimilar ones produce feature points far apart.

The utilization of an isometric-embedded fractal-based technique to extract reliable image features, combined with a multiple-instance learning paradigm to derive relevant concepts, can produce desirable retrieval results that better match user's expectation. In order to demonstrate the feasibility of the proposed approach, two sets of test for querying an image database are performed, namely, the fractal-based feature extraction algorithm vs. three other feature extractors, and single-instance vs. multiple-instance learning. Both the retrieval results, execution time and precision/recall curves show favorably for the proposed multiple-instance fractal-based approach.

1. Introduction

The retrieval of digital images from an image database is an active research area due to the inefficiency of query processing utilizing traditional textual language. Most image retrieval paradigms fall between automated pixel-oriented information models and fully human-assisted database [1–4]. These approaches differ in application domain, visual features extracted, features discrimination criteria employed and query mechanisms supported. Feature vector characterizing image properties is generally composed of color, texture, shape and location information. Distance measure, e.g., n-dimensional Euclidean distance, is utilized to compute the similarity between different feature vectors. Query specification tools are provided to allow user-constructed sketches and weight assignments among different feature components, etc. As an example, the QBIC system allows the color, texture, or shape of an image or part of an image to be compared with feature vectors from database images using Euclidean similarity measure. The retrieval of similar images

* Corresponding author. Tel.: +886 934 151515; fax: +886 7 5254301.
E-mail address: chiang@cse.nsysu.edu.tw (J.Y. Chiang).

0031-3203/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.06.020

from a database corresponds to determining neighboring points in the proximity of the feature point of a query image.

The mapping of an image to the corresponding feature vector is a process of dimensionality reduction. By finding a lower-dimensional representation of an image, an effective feature vector is expected to contain vital characteristics of the original. The pitfalls associated with traditional approaches are two-fold, namely (1) representing features extracted are not powerful enough to discriminate between similar and dissimilar images, and (2) multiple features extracted from a single image do not necessarily match user's expectation. In this paper, an *isometric embedded* feature extraction method employing fractal orthonormal bases (FOB) is introduced. The features extracted from positive and negative example images are further combined by a multiple-instance learning paradigm to induce feature commonality to query image database.

In what follows, the proposed isometric embedded feature extraction method employing fractal orthonormal basis will be introduced first. The conservation of Euclidean distance-based similarity measure before and after the mapping onto orthonormal basis will be proved. Image pairs with long feature distance in the feature domain are guaranteed to be dissimilar ones in the image domain, while feature points close to each other correspond to similar images. Next, the procedure for combining multiple features extracted from positive and negative example images to forge a unified query key will be outlined. This multiple-instance learning procedure identifies common positive features and excludes negative ones to further clarify the user's searching criteria. The last section shows the effectiveness of this novel approach by comparing the retrieval results of (1) single-instance *vs.* multiple-instance learning by applying the same fractal-based feature extraction paradigm, and (2) multiple-instance retrieval by employing the proposed fractal orthonormal basis approach and other feature extraction techniques. Due to the preservation of distance relationship in both the image and feature domains for the fractal-based feature extraction method, and the derivation of commonality from multiple example images, consistent search results matching user's expectation are obtained in both tests.

2. Isometric embedded (FOB)

Even though similar images generally derive feature points close to each other, there is no guarantee that dissimilar images will map to distant feature points. For example, the comparison of color feature usually employs certain measure of histogram. Images with similar histogram distributions will be regarded as similar under this scheme. However, color histogram represents an image's global feature. With an analogous histogram distribution, the color within a dissimilar image might be locally distributed in a totally different manner. Using color feature as a measure of similarity between images is not powerful enough to exclude false-positive cases. Moreover, a query image might be rotational, scaling, shifted, or noise-corrupted variations of database images. A traditional retrieval algorithm might not be sufficiently robust to include similar database images of these variations, causing the occurrence of false dismissal.

The corresponding feature vectors f_1,f_2,f_3, and f_q of images i_1, i_2, i_3, and i_q, respectively, are shown in Fig. 1. The derived feature points in the feature domain might not preserve the same spatial distance relationship as their counterparts in the image domain. When an image i_q is used for querying a database, i_1, i_3 will be included in the search result due to the proximity of points f_1,f_3 with f_q in the feature space. However, image i_2 will be excluded since point f_2 is considered as too distant from f_q. A dissimilar image, e.g., image i_3, mistakenly classified as similar is called a false-positive, while a similar image, e.g., image i_2, incorrectly excluded from the final search result is referred to as a false-negative. Being unable to provide stable distance measure, most systems try to minimize false-negative results at the expense of an increased number of false-positives. A compact, perceptually relevant representation of an image content that preserves the distance relationship in terms of similarity metric in both image and feature spaces is highly desirable.

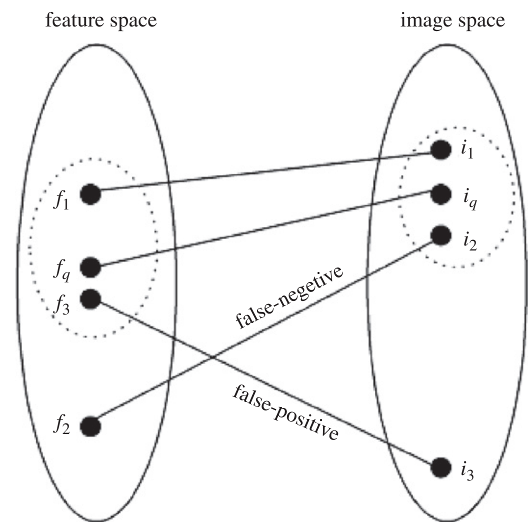

Fig. 1. The distance relationship between image points and corresponding feature points is not preserved through most feature extraction processes. Image-feature pairs $\{i_2,f_2\}$ and $\{i_3,f_3\}$ illustrate the false-negative and false-positive cases, respectively.

Barnsley suggested that storing images as collections of transformations could lead to image compression [5]. Jacquin was among the first to publish a fractal image compression scheme by regular partitioning the image [6]. Jacquin's method is based on partitioned iterated function systems and affine transform acting locally rather than globally. The accurate coding of a range block is dependent upon there being a self-similar domain block in the codebook. Because this piecewise self-similar model is an approximation of real-world data, there is no guarantee that a perfect mapping can be found. Observing that the iterative function system (IFS) coding technique seems to have a limit in the accuracy that an image can be coded, Vines proposed a scheme by finding a set of basis vectors to best represent the image in the sense of achieving higher fidelity with good compression [7]. Vines' method was intended to improve the decoded signal-to-noise ratio of fractal compression. However no application of Vines' approach to image database retrieval was ever suggested.

According to Vines' approach, a set of orthonormal basis vectors is created by the Gram–Schmidt procedure and the range blocks are coded by projecting the block elements onto this basis. The principle in determining the orthonormal set is to create a basis that allows each range block to be accurately represented with a minimum number of the basis vectors. These FOB are derived from the domain vectors. With these vectors, the range blocks can be encoded with a simple projection operation, and the map parameters will be the corresponding weights for this orthonormal basis.

For a range block of size $L_R \times L_R$, let $M_r = L_R^2$ be the length of the range vectors, and $\boldsymbol{R}_I = \{\tilde{r}_i^I\}_{i=1}^{N_R}$ be the set of all range vectors $\tilde{r}_i^I$, $i = 1 \sim N_R$, in an image I. Three basis vectors v_1, v_2, and v_3, determined a priori according to Vines' approach, are orthonormalized later to form the first three vectors of the required M_r orthonormal basis vectors, where $\tilde{v}_1 = \{1, 1, \ldots, 1\}^T$, the DC value, $\tilde{v}_2 = \{0, 1, 2, 3, 4, 5, 6, 7, 0, 1, 2, 3, 4, 5, 6, 7, \ldots, 0, 1, 2, 3, 4, 5, 6, 7\}^T$, the tilt along x-axis, and $\tilde{v}_3 = \{0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, \ldots, 7, 7, 7, 7, 7, 7, 7, 7\}^T$, the tilt along y-axis.

The remaining basis vectors will be chosen to span the $(M_r - 3)$-dimensional subspace S^0 perpendicular to the subspace spanned by