Short communication

Aggregation of color and shape features for hybrid query generation in content based visual information retrieval

P. Androutsos, A. Kushki, K.N. Plataniotis*, A.N. Venetsanopoulos

Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Ont., Canada M5S 3G4

Received 1 August 2002; received in revised form 1 January 2004

Abstract

A fuzzy approach for the aggregation of multiple features in content-based image retrieval is outlined. Color, shape and spatial features extracted using both computational and manual segmentation techniques are used for subsequent generation of hybrid queries to a ground truth image database consisting of architectural photographs. Retrieval results for multiple-feature queries are shown in the form of precision recall graphs. The results indicate that the fuzzy approach presented herein can perform at least as well as a weighted mean approach.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Fuzzy feature aggregation; Content-based image retrieval; Shape; Color; Spatial information; Architectural image database

1. Introduction

It can be argued that the September 2001 finalization of the MPEG-7 standard [2] has signaled the dawning of a new era for content-based visual information retrieval (CBVIR). In the past, initiatives such as IBM’s QBIC [5], and MIT’s Photobook [7] pioneered differing techniques for how visual information was to be indexed and mined, and a grand history of these and other systems can be found in [9]. It is, however, unfortunate that most of the systems which currently share the CBVIR spotlight continue to exhibit a similar level of heterogeneity in their technologies as that which was expected from their predecessors. With MPEG-7, the ISO has succeeded in providing a formal framework for standardizing the way in which visual and auditory data is described and indexed, thus providing a starting point for structured and uniform content-based indexing. The scope of the standard is very clear in that it is limited to the way in which data is described, leaving other elements of the indexing puzzle such as feature extraction, query generation and similarity measurement as open research issues. For the purposes of this paper, a fuzzy...
approach for aggregating different similarity values for a varied set of features is proposed to solve the hybrid query generation problem. Although the descriptors employed within the confines of this scenario do not follow the MPEG-7 standard, their choice is not of primary importance for the purposes of this paper, which instead addresses the issue of feature aggregation rather than feature description. Furthermore, from the results presented, it is shown that good retrieval results are achieved through the use of a number of geometric shape measures much simpler to use than those suggested by the MPEG-7 standard.

The paper presented here is organized as follows. Section 2 discusses the feature descriptors and distance measures used and Section 3 provides a description of the aggregator employed. Experimental results and conclusions and comments are then provided in Sections 4 and 5.

2. Feature description & extraction and distance measures

This section provides a treatise on the features, descriptors and distance measures employed for indexing. Three features were considered in this experiment, namely color, shape and spatial layout of color regions. Section 2.1 outlines the feature extraction techniques used to acquire feature information, followed by Section 2.2 which explains descriptors and distance measures/metrics used for specific features.

2.1. Feature extraction

The dominant feature extracted in this experiment was color. In particular, colored regions as well as their spatial locations were extracted from database images using a recursive HSV segmentation scheme. As detailed in [1], perceptually similar color regions are segmented together through the thresholding of the hue histogram in order to obtain the constituent bright chromatic pixels. These are further thresholded to obtain \( m \) bright colors, where \( m \) is dependent on the image at hand. Saturation and value parameters from the HSV color space are subsequently used to extract achromatic regions. The scheme results in providing a total of \( c = \sum_{i=1}^{n_p} i + m \) colors, where \( n_p \) represents the number of peaks in the saturation histogram. Extremely small color regions with areas equivalent to less than one percent of the total image area were deemed spurious and removed in post-processing. The result of this process was a color segmented version of each image in the database from which the percentage

<table>
<thead>
<tr>
<th>Nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D )</td>
</tr>
<tr>
<td>( N )</td>
</tr>
<tr>
<td>( D_i )</td>
</tr>
<tr>
<td>( D_{wm} )</td>
</tr>
<tr>
<td>( \mu_{(D_i(Q,I))} )</td>
</tr>
<tr>
<td>( w_i )</td>
</tr>
<tr>
<td>( c )</td>
</tr>
<tr>
<td>( \mu_{ij} )</td>
</tr>
<tr>
<td>( \gamma )</td>
</tr>
<tr>
<td>( z )</td>
</tr>
<tr>
<td>( S_{O,A} )</td>
</tr>
<tr>
<td>( x_i )</td>
</tr>
<tr>
<td>( S_{c_i,c_j} )</td>
</tr>
<tr>
<td>( I )</td>
</tr>
<tr>
<td>( Q )</td>
</tr>
<tr>
<td>( m )</td>
</tr>
<tr>
<td>( p )</td>
</tr>
<tr>
<td>( S_{vad} )</td>
</tr>
<tr>
<td>(</td>
</tr>
<tr>
<td>( D_m )</td>
</tr>
<tr>
<td>( r )</td>
</tr>
</tbody>
</table>
contribution for each segmented color as well as the spatial distribution of color regions could be indexed. The spatial location of color regions was recorded according to the location of its center of mass.

Despite the fact that color segmentation provides a somewhat decent indication of object shape, using color segmentation for shape extraction is usually less than optimal. When the images at hand are simple, there is sometimes a one-to-one correspondence between the extracted color masks and the optimal shape masks. The images used here, however, did not exhibit such behavior, and as a result, separate extraction of shape masks was necessary. Since the focus of this paper is geared towards the fusion of information rather than shape segmentation, the development of an effective shape extraction algorithm was not pursued and shape masks were extracted manually using a commercial image editing program and tablet interface. To maintain perceptual uniformity, this task was performed by a single person, and extracted shape masks were limited to simple geometric figures as often as possible (i.e. rectangles, circles, ellipses, and triangles). Fig. 1 illustrates the overall process used for feature extraction as well as for feature description where the image shown is processed for color separately from shape.

2.2. Feature description and similarity measures

For each element in the image database, the descriptions of the features outlined in Section 2.1 were written to an index. For color, both the set of constituent colors as well as their coverage as a percentage of the overall image area were recorded. The dominant color in the image was also recorded as a special element in the index to eliminate the need to determine color dominance during querying. Similarity between colors was calculated using the vector angular distance measure shown in Eq. (1) and outlined in [1].

\[
S_{\text{vad}}(c_i, c_j) = 1 - \left( 1 - \frac{|c_i - c_j|}{\sqrt{3 \cdot 255^2}} \right) \cdot \frac{1 - \frac{2}{\pi} \cos^{-1} \left( \frac{c_i \cdot c_j}{|c_i||c_j|} \right)}{1 - \frac{2}{\pi} \cos^{-1} \left( \frac{c_i \cdot c_j}{|c_i||c_j|} \right)}.
\]  

In the expression for \(S_{\text{vad}}\), the similarity between the two colors \(c_i\) and \(c_j\) has two components. The first half of the measure determines the magnitude of the difference between the vectors, while the latter is for determining the difference in angle between the two colors in the color space.
Since $S_{vad} \in [0, 1]$, this similarity value is easily converted to a distance value by subtracting it from unity (if required).

As mentioned in the Introduction, this paper does not directly deal with MPEG-7 descriptors, but rather addresses the issue of combining features for hybrid query generation. As a result, the actual choice of descriptors is made according to the requirements of the scenario at hand. Due to the complexity associated with the ISO suggested shape descriptors, much simpler measures were chosen for this application. Specifically, a variety of geometric measures were employed. Besides basic features such as bounding rectangles, perimeter and area, values for compactness, major axis direction, and eccentricity were determined for shape description in addition to the use of the first four invariant moments proposed by Hu in [4]. The end result of this geometric shape analysis is a descriptor vector with 12 elements.

Similarity between geometric descriptor values were determined with the generalized Minkowski distance [3] shown in Eq. (2) where the use of $r = 1$ and $r = 2$, respectively provide the Manhattan and Euclidean distances.

$$S_m(x_i, x_j) = \left( \sum_{k=1}^{|x_i|} |x_i^k - x_j^k|^r \right)^{1/r}, \quad |x_i| = |x_j|. \quad (2)$$

The Euclidean distance ($r = 2$) was employed for all geometric shape measures. The Euclidean distance was also used to measure the distance between the locations of centroids of colored regions and as explained previously, the vector angular distance was employed for measuring color distance $(1 - S_{vad})$.

Although it is possible to use a single uniform technique for measuring the distances for all descriptors, in practice this is not a good approach (e.g. for non-Euclidean spaces, it does not make sense to apply a Minkowski metric using $r = 2$). In addition, when issues of subjectivity are of interest, it is beneficial to use a perceptually tuned distance measure in order to ensure that descriptor pairs with low distance values are indeed perceived as similar from a subjective standpoint.

3. Fuzzy aggregator

Typically, in order to establish an overall measure of the distance between two images, the distance between the descriptions of the constituent features of the images must first be determined. These descriptions are different in nature and therefore different distance measures need to be employed for each descriptor. As a result, an image with $N$ descriptions has a distance set $D = \{D_1, D_2, \ldots, D_N\}$ where $D_i$ is the distance between two images with respect to descriptor $i$.

These descriptor distances must somehow be combined in order to obtain an overall measure of distance between the two images. A relatively straightforward technique for performing this combination is using the weighted mean shown in Eq. (3), where $w_i$ indicates the weight of the $i$th feature and $\sum_{i=1}^{N} w_i = 1$.

$$D_{wm} = \sum_{i=1}^{N} w_i D_i. \quad (3)$$

The proposed fuzzy aggregation operator can be employed to the same ends as the weighted mean to combine descriptor-level distance values and generate a single, overall similarity value. Unlike the weighted mean approach, however, the fuzzy aggregator combines decisions on distances as opposed to distances/similarities directly. This approach provides added flexibility in modelling various conceptual user queries by allowing the use of logical operators such as AND, OR, and NOT. It must be noted that fuzzy decisions represent the grade of membership to the set of images similar to the query and are generated by passing the distance values through fuzzy membership functions in order to generate a similarity value as illustrated in Fig. 2. An exponential was used as a membership function in this scenario in order to comply with human perceptual characteristics.

$$\mu_{D_{Q,I}} = \exp \frac{-D_{Q,I}}{\alpha}. \quad (4)$$

Above, $\alpha$ is a normalization factor which was set to the median value of all descriptor distances and $D_{Q,I}$ represents the distance between images $Q$ and $I$ for descriptor $i$. 

---

This text seems to be a continuation from a previous page discussing geometric shape analysis and distance measurement in images. The focus is on the methods used to measure geometrical properties and how to combine these measurements to obtain an overall similarity value between images. The text explains the use of the Minkowski distance and the Euclidean distance for different geometrical features and the application of these distances in a weighted mean approach for combining descriptor-level distances. It also introduces a fuzzy aggregation operator that allows for the combination of decisions on distances, using logical operators such as AND, OR, and NOT, to achieve a single, overall similarity value.
In devising a suitable fuzzy operator for combining descriptor decisions, various logical operators were considered. Since content-based image retrieval systems typically seek to find images which are similar to a query rather than exactly the same, using a logical AND operation that requires high similarity in all descriptors may be too restrictive. Furthermore, it is often enough if the query and the candidate images are similar only in some of the given descriptors and this leads to the use of a logical OR operator. Yet, matching according to only one query requirement is overly optimistic. With this in mind, the compensative operator outlined in [8] was chosen to provide a compromise between the logical AND and OR operators. This is shown in Eq. (5) where \( m_1 \) and \( m_2 \) are fuzzy decisions made on two descriptor distance values \( D_1 \) and \( D_2 \), respectively.

\[
\mu_1 \otimes \mu_2 = [\max(\mu_1, \mu_2)]^\gamma \min(\mu_1, \mu_2)]^{1-\gamma}.
\] (5)

The compensative operator’s behavior ranges between the logical AND and OR depending on the parameter \( \gamma \). The AND operator can be obtained by setting \( \gamma = 0 \), \( \gamma = 1 \) results in the OR operator, and for all other values of the \( \gamma \) parameter, the aggregation result will be a compromise among the arguments of the operator.

In addition to aggregating the results of the various descriptors, it may be necessary to combine the results based on various instances of the same descriptor. For example, if the user is interested in images that contain the colors red and yellow, the overall similarity of each of the candidate images to this query will be a combination of their similarity to the individual colors red and yellow. In other words, the decision for each descriptor is in fact an aggregation of the various required instances of that descriptor. Therefore, if the decision based on the \( j \)th required instance of descriptor \( i \) is denoted as \( \mu_{ij} \), the similarity decision for that descriptor based on \( n \) instances can be obtained as

\[
\mu_i = \mu_{i1} \otimes \mu_{i2} \cdots \otimes \mu_{in}.
\] (6)

Finally, Eq. (7) defines the calculation for overall image similarity as a fuzzy aggregation of all descriptor decisions, where each \( \mu_i \) represents the aggregation of a particular descriptor’s similarity decision (for shape, color, etc.), and \( \gamma \) is set to a value of 0.3.

\[
S_{i,n,\gamma} = \max_{\forall i} (\mu_i)^\gamma \min_{\forall i} (\mu_i)^{1-\gamma}.
\] (7)

4. Experimental results

This section deals with the experiments done to test the proposed aggregation method as well as provide results showing performance in comparison to the weighted mean approach using precision recall graphs. The database used consisted of 220 architectural images (i.e. buildings, homes, castles, etc.) taken from a Corel image library, containing ground truth. The database chosen was
selected as a specific application scenario, but the aggregator can be applied to any image database containing ground truth without loss of generality. The ground truth was established using ten subjects who determined similarity for all database

<table>
<thead>
<tr>
<th>Query type</th>
<th>Query data</th>
<th>Image source</th>
</tr>
</thead>
</table>
| Color only   | - white (255, 255, 255)  
- blue (29, 68, 131)  
- black (0, 0, 0) | Extracted from source image in Fig. 3(a) |
| Shape only   | Triangular |              |
| Layout only  | Layout of colors from image in Fig. 3(a) |

<table>
<thead>
<tr>
<th>Query type</th>
<th>Query data</th>
<th>Image source</th>
</tr>
</thead>
</table>
| Color only   | - white (255, 255, 255)  
- green (84, 140, 229)  
- brown (29, 68, 131)  
- black (0, 0, 0) | Extracted from source example image in Fig. 3(b) |
| Shape only   | Rectangular |              |
| Layout only  | Layout of colors from image in Fig. 3(b) |

Table 1
Data for query set 1

Table 2
Data for query set 2

Each subject determined overall image similarity according to color and shape content, and no time restrictions were imposed. The two query sets (for which data are shown in Tables 1 and 2) examined here each consisted of a color-only query, a shape-only query, and a color-layout-only query all generated from a common source image (see Figs. 3(a) and (b)). For each query set, two hybrid queries were performed; a weighted mean query and a fuzzy aggregated query each combining all elements from the query set.

Retrieval results for the two combination schemes (fuzzy aggregator and the weighted mean) are shown for query sets 1 and 2 in Figs. 4 and 5, respectively. From a purely visual standpoint, it can be argued that the retrieval results shown in Fig. 4(a) using the fuzzy aggregator are better than

Fig. 3. Source images used in creation of data for (a) query set 1 and (b) query set 2.
the ones obtained via the weighted mean in (b). The retrieved images in for the results in Fig. 5 seem to exhibit a greater degree of uniformity to the example image which in both cases holds the top rank in each result set. Of interest is the fact that the weighted mean approach in Fig. 5(b) seems to retrieve a greater number of images which contain unspecified colors (i.e. green, and red). Thus, from a simple visual standpoint, it can be said that both retrieval sets seem to have relatively equal performances.

Although the visual inspection of results ultimately decides whether or not a particular retrieval is good or not, it is impractical from a numerical standpoint. Employing the ground truth for this database, the precision-recall graphs in Figs. 6 and 7 were generated. From these plots, it is evident that hybrid queries consisting of combinations of features are more effective than individual queries for retrieving similar images. In addition, it can be seen from the trends that for the two query sets shown, the retrieval performance of the proposed fuzzy aggregator is superior to the weighted mean approach. For the specific case of query set 1, both hybrid query generation schemes perform identically until they suddenly and drastically diverge at two retrievals. For query set 2, the fuzzy aggregator consistently performs better until after around seventy retrievals where the weighted mean surpasses it.

5. Concluding remarks

In this paper, a fuzzy aggregator used for the generation of hybrid queries in content-based visual information retrieval is presented. This operator is applied to aggregate color and shape features with the respective descriptions being color coverage and color layout as well as a number of geometric shape measures. The similarities for these features are the vector angular measure for dominant color, and Euclidean distance for the remainders. It must be noted that this fuzzy operator can be used to aggregate any number of features in order to arrive at a single
decision on overall image similarity. Unlike vector-based approaches which can stack vectors to generate supervectors, the incorporation of new data into a hybrid query does not force this fuzzy technique to use a new similarity operator which is typically fixed to the size of the supervector used. This property makes it highly scalable and easily extended to any number of features or descriptors. The fact that it can easily adapt to a large number of features means that it has the capacity to incorporate many different feature descriptors in order to achieve improved retrieval results. More importantly, this aggregator is able to act as a general framework for many of the hybridization schemes presented herein since their operations are merely special cases of the fuzzy approach. Systems employing the geometric mean in order to perform feature combination can be implemented using the fuzzy aggregator by setting the fuzzy parameter $\gamma$ equal to 0.5. Furthermore, the aggregator proposed here has the operate in the same capacity as the Boolean algebra based MARS system [6]. Once again, this combination technique is a special case of the proposed aggregator which behaves like a purely boolean-based system when $\gamma$ is set to zero or unity in order to simulate AND and OR operators (NOT can also be implemented by using a $1 - \mu$ function). Based on the precision recall results shown, it is evident that the proposed fuzzy aggregator is superior to the weighted mean approach for the data sets investigated. The power of this
aggregation technique, however, does not necessarily lie in the fact that it can provide better results, but rather in its natural ability to behave like the arithmetic, geometric, and boolean based combination methods, as well as its ability to easily scale up or down as the number of descriptors used for retrieval changes.

References


