Artificial Vision

Project 2 – Implementation and evaluation of visual descriptors from MPEG-7 standard

Goal:

Providing the student detailed information concerning several visual descriptors defined by the MPEG-7 standard. Furthermore, it is intended to use generic quantitative metrics to evaluate the performance of a given low-level descriptor applied to the problem of visual information retrieval on image database (CBIR - Content Based Image Retrieval).

Procedure:

The project is divided into two stages: **implementation** and **evaluation**. In the first stage, the following visual descriptors should be implemented: color (ColorStructure), texture (EdgeHistogram) and (optionally) shape (RegionShape), adopting the recommended procedures defined by the MPEG-7 standard. In the second stage, it is intended to evaluate quantitatively the performance of the implemented descriptors in the context of information retrieval in image database, using the QBE (Query by Example) paradigm. For this evaluation it will be used the ANMRR (Average Normalized Modified Retrieval Rate) criteria. Therefore, students should build a small image database, previously classified according to their content and define a small set of additional images, to be used as queries.

As a result of the first stage, it should be obtained an application that receives as input the file name of the image (for example, ***im28.jpg*** - can also, optionally, consider a second file, ***im28.pbm***, representing a binary image indicating the active pixels, otherwise all pixels are considered active) and outputs the ***im28.mp7*** file containing the corresponding metadata.

To implement the *ColorStructure* and *EdgeHistogram* visual descriptors, it should be used the steps described in the slides of the course and for the *RegionShape* descriptor (optionally), the bibliography should be consulted (to implement this descriptor is necessary to perform an image segmentation).

For the first stage (**implementation**), students should acquire prior knowledge relatively to the following questions (include those answers on the final report):

1. Why several visual descriptors are needed?
2. Why other color spaces representation are needed, different from RGB space?
3. What is the motivation for the non-uniform HMMD color space quantization?
4. What is the motivation behind the use of the "structured" color histogram, rather than a simple count of the image colors?
5. What is the motivation of dividing each sub-image in approximately 1100 blocks, in the *EdgeHistogram* descriptor?
6. What is the procedure to determine the direction of each image-block in the EHD?
7. What is the main difference between the *RegionShape* and *ContourShape* descriptors?

The evaluation stage is based on ANMRR criteria and can be computed with the following steps:

1. Select a set of *N* images, $I_1, \cdots, I_N$, used as an image database;
2. Define another additional set of *M* images, $Q_1, \cdots, Q_M$, where $M \ll N$, used as query images;
3. For each query image $Q_j, j = 1, \cdots, M$, manually define the list of images from the database that ideally should be retrieved. As a suggestion, define the variable GT (python *list* type) to store this information, so that, $GT[j]$ would be a set of integers (image id) that represents the ground-truth of query Q_j. The length of each query ground-truth could be different but should be small (for example, $len(GT[j]) \leq 5, \forall j$) and the database can include some images that not belong to any GT.
4. For each image from the database and query list, compute the visual descriptor under evaluation. The resulting information can be stored in two variables, one for the images in the database and another for the query image list.

5. Compute a distance matrix D, with size of $N \times M$, where each element d_{ij} is the distance between the descriptor of the image $I_i, i = 1, \cdots, N$ (from the database) and the descriptor of the query image Q_j.
6. Build an application that takes as input the variables D and GT and returns the ANMRR value (Average Normalized Modified Retrieval Rate).

For the previous stage, students should acquire prior knowledge relatively to the following questions (include those answers on the final report):

1. Why visual descriptors need to be evaluated quantitatively?
2. What were the main concerns underlying the definition of ANMRR criteria? Can this measure reflects the degree of user satisfaction of a given database image retrieval application?
3. Does ANMRR criteria is specific for this application or can also be used, more generally, in the evaluation of textual information search engines (google, yahoo, sapo, etc.)?
4. Can ANMRR measure be considered as a signal to noise ratio?