Investigating the Behavior of Compact Composite Descriptors in Early Fusion, Late Fusion and Distributed Image Retrieval

Savvas A. CHATZICHRISTOFIS, Avi ARAMPATZIS, Yiannis S. BOUTALIS

Department of Electrical and Computer Engineering
Democritus University of Thrace, Xanthi, Greece

schatzic@ee.duth.gr, avi@ee.duth.gr, ybout@ee.duth.gr

Abstract. *In Content-Based Image Retrieval (CBIR) systems, the visual content of the images is mapped into a new space named the feature space. The features that are chosen must be discriminative and sufficient for the description of the objects. The key to attaining a successful retrieval system is to choose the right features that represent the images as unique as possible. A feature is a set of characteristics of the image, such as color, texture, and shape. In addition, a feature can be enriched with information about the spatial distribution of the characteristic that it describes. Evaluation of the performance of low-level features is usually done on homogenous benchmarking databases with a limited number of images. In real-world image retrieval systems, databases have a much larger scale and may be heterogeneous. This paper investigates the behavior of Compact Composite Descriptors (CCDs) on heterogeneous databases of a larger scale. Early and late fusion techniques are tested and their performance in distributed image retrieval is calculated. This study demonstrates that, even if it is not possible to overcome the semantic gap in image retrieval by feature similarity, it is still possible to increase the retrieval effectiveness.*

Keywords

CBIR, Compact Composite Descriptors, Early Fusion, Late Fusion, Distributed Image Retrieval.

1. Introduction

Content-based image retrieval (CBIR) is defined as any technology that in principle helps to organize digital image archives by their visual content. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR [11]. The most common form of CBIR is an image search based on a visual example. The user inputs an image (query image), and, based on certain low level features, the system brings up similar images. This sort of features are used for describing the content of the image and must thus be appropriately selected on each occasion. These features can be global features, which describe information from the entire image, or local features, which describe segments, regions or patches of the image. In current research efforts in visual information retrieval, global features have lost part of their significance [29]. Despite this, global features are a factor often used in Content-based image retrieval (CBIR) systems as they are easy to handle and still provide basic retrieval mechanisms. Fusion in image retrieval goes hand-in-hand with practical, viable system development, which is critical for the future of image retrieval research [11]. Two main approaches to fusion have been taken: early fusion, where multiple image descriptors are composed to form a new one before index time [30], and late fusion, where result lists from individual descriptors are fused during query time [22][18], as in text meta-search.

Compact composite descriptors (CCDs) [4], [10], [5], [6] are global image features, capturing more than one type of information at the same time in a very compact representation. Their quality has so far been evaluated in retrieval from several benchmarking databases, and in the scholarly literature has been found to be better than other descriptors such as the MPEG-7 [24], [23] descriptors. The basic difference between CCDs and other descriptors in the literature lies in the fact that each of these descriptors is determined for a different type (in terms of content) of image. The structure of these descriptors is described in Section 2.

Section 3 describes the process of early fusion for the 2 descriptors of the CCD family, which combine texture and color information. These two descriptions are intended for indexing and retrieval of natural color images.

In Section 4, we consider heterogeneous databases and investigate query-time fusion techniques for CCDs. The results show that fusion is beneficial even with simple score normalization and combination methods, due to the compatibility of the score distributions produced by the CCDs considered.

In Section 5 we investigate the behavior of CCDs in a distributed image retrieval setup, where each database is described by a different CCD. We compare a variety of linear and non-linear score normalization methods in order to find the most suitable method for merging results. Experiments show that non-linear methods work better than linear, although merging without normalization works best due to the compatibility of the score distributions produced by the CCDs. Finally, the conclusions are given in Section 6.

2. Compact Composite Descriptors

The family of compact composite descriptors includes descriptors for 3 types of images. A group of 2 descriptors combines color and texture information in order to describe natural color images. One descriptor combines brightness and texture characteristics in order to describe grayscale images (primarily radiology medical images), and the other combines color and spatial distribution characteristics in order to describe artificially generated images (computer graphics, color sketches etc.)

2.1 CCDs for Natural Color Images

This category includes 2 descriptors: the Color and Edge Directivity Descriptor (CEDD) [10] and the Fuzzy Color and Texture Histogram (FCTH) [4]. The structure of these descriptors consists of n texture areas. In particular, each texture area is separated into 24 sub-regions, with each sub-region describing a color. CEDD and FCTH use the same color information, as it results from 2 fuzzy systems that map the colors of the image in a 24-color custom palette. To extract texture information, CEDD uses a fuzzy version of the five digital filters proposed by the MPEG-7 EHD [33], forming 6 texture areas. In contrast, FCTH uses the high frequency bands of the Haar Wavelet Transform in a fuzzy system, to form 8 texture areas.

An important characteristic of these 2 descriptors is the small size needed for indexing images. The CEDD length is 54 bytes per image while FCTH length is 72 bytes per image.

2.2 CCD for Grayscale Images

This category includes the Brightness and Texture Directionality Histogram (BTDH) descriptor. This descriptor was initially proposed in [5] for use with grayscale radiology images but is considered suitable for every grayscale image. This descriptor uses brightness and texture characteristics as well as the spatial distribution of these characteristics in one compact 1D vector.

To extract the BTDH descriptor, a two unit fuzzy system is used. To extract the brightness information, a fuzzy unit classifies the brightness values of the image's pixels into a preset number of clusters. The cluster centers are calculated using the Gustafson Kessel Fuzzy Classifier [15]. The texture information embodied in the proposed descriptor comes from a fuzzy approach suggested by the Directionality histogram in [31]. This feature is part of the well known Tamura texture features. Finally, a Fractal Scanning method is used to capture the spatial distribution of brightness and texture information.

[1]More details about the ANMRR are given in Appendix 1.

2.3 CCD for Artificially Generated Images

The recently proposed Spatial Color Distribution Descriptor (SpCD) [6], combines color and spatial color distribution information. The descriptors of this type can be used for image retrieval by using hand-drawn sketch queries, since this descriptor captures the color layout information. In addition, the descriptors of this structure are considered to be suitable for colored graphics, since such images contain a relatively small number of color and less texture regions than the natural color images.

SpCD uses a fuzzy linking system that maps the colors of the image in a custom 8-color palette. In order to integrate the spatial information, this descriptor divides the image into a predetermined number of sub-images and, by scanning each one, captures the spatial distribution of the color. Important characteristics of SpCD are its small storage needs, which do not exceed 48 bytes an image, and its small size, which does not exceed 48 bins. During data retrieval from databases, the length of the retrieved information is of great significance.

3. CEDD and FCTH Early Fusion

The results for the 2 CCDs intended for natural color images in different benchmarking databases are notable. Tab. 1 shows the ANMRR [25] results in 3 image databases in contrast with the MPEG-7 descriptor results. The ANMRR is always in a range of 0 to 1, and the smaller the value of this measure is, the better the matching quality of the query. ANMRR is the evaluation criterion used in all of the MPEG-7 color core experiments. Evidence shows that the ANMRR measure coincides approximately linearly with the results of subjective evaluation of search engine retrieval accuracy[1].

	WANG [32]	UCID [28]	NISTER [27]
CCD			
CEDD	0.25283	0.28234	0.11297
FCTH	0.27369	0.28737	0.09463
MPEG-7			
DCD MPHSM	0.39460	-	-
DCD QHDM	0.54680	-	-
SCD	0.35520	0.46665	0.36365
CLD	0.40000	0.43216	0.2292
CSD	0.32460	-	-
EHD	0.50890	0.46061	0.3332
HTD	0.70540	-	-

Tab. 1. ANMRR Results on Several Benchmarking Databases.

The ANMRR values for the MPEG-7 descriptors in WANG's [32] database as well as the ground truths that were used are available at [34]. Since the MPEG-7 descriptor results are not available for UCID [28] and NISTER [27] database, an implementation of CLD, SCD and EHD in img(Rummager) [8] and LIRe Demo [21] retrieval systems is used. Details relating to the experimental results, the implementation of the MPEG-7 descriptors, as well as the ground truths that were used, are available in [9].

Observing the results in various queries, it is easy to ascertain that in some of the queries, better retrieval results are achieved by using CEDD, and in others by using FCTH. Could these 2 descriptors be combined to improve the retrieval results?

Based on the fact that the color information given by the 2 descriptors comes from the same fuzzy system, we can assume that joining the descriptors will rely on the combining of texture areas carried by each descriptor. Therefore, uniting CEDD and FCTH will lead to a new descriptor that will be made up of a combination of CEDD and FCTH texture areas. The types of texture areas adopted by each descriptor are illustrated in Fig. 1.

	0	1	2	3	4	5	6	7
CEDD	Linear	Non Directional	Horizontal Activation	Vertical Activation	45 Degree Diagonal	135 Degree Diagonal	-	-
FCTH	Linear Low Energy	Horizontal Low Energy	Vertical Low Energy	Both Directions Low Energy	Linear High Energy	Horizontal High Energy	Vertical High Energy	Both Directions High Energy

Fig. 1. CEDD and FCTH Texture Areas.

The combined descriptor is called Joint Composite Descriptor (JCD)[7]. It is made up of 7 texture areas, with each area made up of 24 sub-regions that correspond to color areas. The texture areas are as follows: JCD(0) Linear Area, JCD(1) Horizontal Activation, JCD(2) 45 Degrees Activation, JCD(3) Vertical Activation, JCD(4) 135 Degrees Activation, JCD(5) Horizontal and Vertical Activation and JCD(6) Non Directional Activation.

In order to make the combination process of CEDD and FCTH clear, we model the problem as follows: Let CEDD and FCTH be available for one image (j). The indicator $m \in [0,23]$ symbolises the bin of the color of each descriptor while $n \in [0,5]$ and $n' \in [0,7]$ determine the texture area for the CEDD and FCTH, respectively. Each descriptor can be described in the following way: $\mathrm{CEDD}\,(j)_n^m, \mathrm{FCTH}\,(j)_{n'}^m$.

For example, the symbol $\mathrm{CEDD}(j)_2^5$ corresponds to the $\mathrm{bin}(2 \times 24 + 5 = 53)$ of the CEDD descriptor of image (j). The algorithm for the Joint Composite Descriptor can be analyzed as follows:

$$\mathrm{JCD}\,(j)_0^i = \frac{\mathrm{FCTH}\,(j)_0^i + \mathrm{FCTH}\,(j)_4^i + \mathrm{CEDD}\,(j)_0^i}{2}, \quad (1)$$

$$\mathrm{JCD}\,(j)_1^i = \frac{\mathrm{FCTH}\,(j)_1^i + \mathrm{FCTH}\,(j)_5^i + \mathrm{CEDD}\,(j)_2^i}{2}, \quad (2)$$

$$\mathrm{JCD}\,(j)_2^i = \mathrm{CEDD}\,(j)_4^i, \quad (3)$$

$$\mathrm{JCD}\,(j)_3^i = \frac{\mathrm{FCTH}\,(j)_2^i + \mathrm{FCTH}\,(j)_6^i + \mathrm{CEDD}\,(j)_3^i}{2}, \quad (4)$$

$$\mathrm{JCD}\,(j)_4^i = \mathrm{CEDD}\,(j)_5^i, \quad (5)$$

$$\mathrm{JCD}\,(j)_5^i = \mathrm{FCTH}\,(j)_3^i + \mathrm{FCTH}\,(j)_7^i, \quad (6)$$

$$\mathrm{JCD}\,(j)_6^i = \mathrm{CEDD}\,(j)_1^i \quad (7)$$

with $i \in [0,23]$.

In order to measure the similarity of the images on the basis of JCD, the Tanimoto coefficient is used, just as in the CEDD and FCTH. The distance T of images (a) and (b) is determined as T_{ab} and is calculated as follows:

$$T_{ab} = T(\mathrm{JCD}(a)_n^m, \mathrm{JCD}(b)_n^m) = \quad (8)$$

$$= \frac{\mathrm{JCD}(a)_n^{mT}\mathrm{JCD}(b)_n^m}{\mathrm{JCD}(a)_n^{mT}\mathrm{JCD}(a)_n^m + \mathrm{JCD}(b)_n^{mT}\mathrm{JCD}(b)_n^m - \mathrm{JCD}(a)_n^{mT}\mathrm{JCD}(b)_n^m}$$

where $\mathrm{JCD}(a)_n^{mT}$ is the transposed vector of the $\mathrm{JCD}(a)_n^m$. In the absolute congruence of the vectors the Tanimoto coefficient takes the value 1, while in the maximum deviation the coefficient tends to zero.

Joint Composite Descriptor has been integrated in the retrieval software system Img(Rummager).

Experiments were carried out on 3 benchmarking image databases. Initially, experiments were carried out in Wang's database of 1000 images [32]. In particular, queries and ground truths proposed by the MIRROR [34] image retrieval system are used. MIRROR separates the WANG database into 20 queries. Experiments were also carried out in the UCID image database [28]. This database currently consists of 1338 uncompressed TIFF images. All the UCID images were subjected to manual relevance assessments against 262 selected images, creating 262 ground truth image sets for performance evaluation. UCID and WANG databases are part of the benchmark databases proposed by [12]. These databases are used for a meaningful comparison of feature performances.

Fig. 2. Retrieval Results on query "'270"' (Wang Database) using (a) CEDD; NMRR = 0.387, (b) FCTH; NMRR = 0.343 and (c) JCD; NMRR = 0.335.

Finally, experiments were carried out in the Nister [27] image database. The Nister image database is made up of 10200 images divided into 2550 groups of 4 images. Each group includes images of a single object. The experiments took place in the first 1000 images of the database with 250 queries. The objective Averaged Normalized Modified Retrieval Rank (ANMRR) is employed to evaluate the performance of the image retrieval system that uses the proposed method in the retrieval procedure.

	WANG	UCID	NISTER
CEDD	0.25283	0.28234	0.11297
FCTH	0.27369	0.28737	0.09463
JCD	0.25606	0.26832	0.085486

Tab. 2. Early Fusion Results (ANMRR).

As Tab. 2 demonstrates, the JCD succeeds in approaching the values of CEDD and FCTH in the Wang database, whereas in the remaining databases, it presents better results than in the other two CCD. These experimental results show that the early fusion method is able to present better results than CEDD and FCTH.

4. CCDs Late Fusion

The quality of the CCDs has so far been evaluated through retrieval from homogeneous benchmarking databases, containing images of only the type that each CCD is intended for. For example, the JCD is tested on Wang, UCID and NISTER databases which contain natural color images, the BTDH on the IRMA database consisting of grayscale medical radiology images, and the SpCD on two benchmarking databases with artificially generated images.

In this study, we evaluate the retrieval effectiveness of late fusion techniques which enable the combined use of the JCD, BTDH, and SpCD, on heterogeneous databases.

We created a heterogeneous database with 20230 images by joining the following: 9000 grayscale images from the IRMA 2005 database; 10200 natural color images from the NISTER database and 1030 artificially generated images from the Flags database [6]. We used 40 queries: The first 20 natural color image queries from the NISTER database and the first 20 grayscale queries of the IRMA 2005 database.

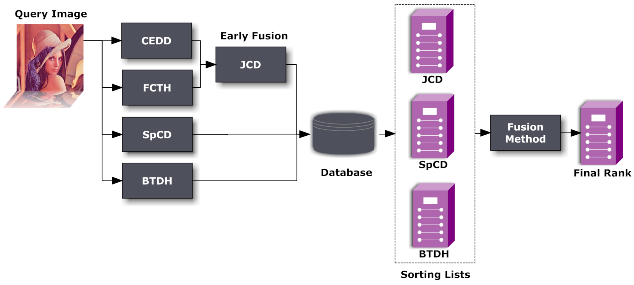

Fig. 3. Late Fusion Implementation Process.

Fusion methods consist of a score normalization and a score combination component. We mainly focus on the normalization, using each time the combination method more natural to the normalization at hand, although we investigated other possibilities in initial experiments not reported here. More details about the combination methods are available at [13].

Six fusion methods were tested:

- **CombSUM**: This is a classic method for integrating results from different ranking lists. The sum of the deviation presented in each ranking list is defined as the deviation of each image. Finally, the images are ranked on the basis of this sum. In general, this method is described as the addition of all scores per image, without normalization.

- **Borda Count + CombSUM**: The Borda Count originates from social theory in voting. The image with the highest rank on each ranked list gets n votes, where n is the collection size. Each subsequent rank gets one vote fewer than the previous. Votes across ranked-lists are naturally combined with CombSUM.

  Let the query be A. The search is performed on a database according to the JCD. The results are sorted according to the distance D, which each image presents from the query A. Each image l, depending on the position shown in the results, is scored as follows:

$$Rank'(l) = N - R(l) \tag{9}$$

  where N is the total number of the images in the database and $R(l)$ is the Rank of the image l after classification.

  The same procedure is followed for the BTDH and the SpCD descriptors. The results are classified and each image is scored with $Rank(l)''$ and $Rank(l)'''$, respectively. Finally, for each l image the $Rank(l) =$

$Rank(l)' + Rank(l)'' + Rank(l)'''$ is calculated and a final classification of the results according to the *Rank* of each image is made.

- **IRP**: The Inverse Rank Position merges ranked lists in the decreasing order of the inverse of the sum of inverses of individual ranks.

  The operating principle of the method is as follows: A search is performed using each of the 3 descriptors. The IRP Rank of each image is determined as:

$$\mathrm{IRP}(l) = \frac{1}{\sum_{D=0}^{2} \frac{1}{Rank_D(l)}} \quad (10)$$

  where $Rank_D(l)$ is the position where the image was found when the search was performed with descriptor d, $d \in \{0,2\}$. At the end of the process the images are ranked on the basis of their IRP distance.

- **Z-SCORE**: A standard method for score normalization that takes the SD (score distribution) into account is the Z-SCORE. Scores are normalized, per topic and engine, to the number of standard deviations by which they are higher (or lower) than the mean score:

  Assume that for query A there were 3 ranking lists, one for each descriptor. For each image l, its Z-Score is calculated for each of the ranking lists according to the function:

$$s'(l) = \frac{s(l) - \mu}{\delta} \quad (11)$$

  where $s(l)$ is the distance which image l presents with query image A, μ the average value of the distances, and δ the typical deviation.

  With the completion of the process, the $s'(l)$ values of each image for each ranking list are summed and the final ranking is done on the basis of these values.

  The mean and standard deviation depend on the length of the ranking. Z-SCORE seems to assume a normal distribution of scores, where the mean would be a meaningful “neutral” score. As it is well-known, actual SDs are highly skewed and clearly violate the assumption underlying the Z-SCORE. Although not very popular in information retrieval, Z-SCORE was used with reasonable success in [19]. In the field of image retrieval, it was used by [14].

- **Z-SCORE with Median**: In order to compensate for any skews in the score distributions of the descriptors we replaced the mean value with the median.

- **Aggregate Historical CDF (HIS)**: HIS is a non-linear normalization which maps each score to the probability of a historical query scoring a collection image below that score. This enables normalization of scores to probabilities - albeit not of relevance - comparable across different engines. Nevertheless, it is not clear - if it is even possible - how using a single distribution can be applied to thresholding, where for optimizing most common measures a reference to (or probabilities of) relevance are needed. Per engine, the proposed normalization is:

$$s'(l) = F^{-1}(P(S \leq s(l))) \quad (12)$$

  where $P(S \leq s(l))$ is the cumulative density function (CDF) of the probability distribution of all scores aggregated by submitting a number of queries to the engine, and F is the CDF of the score distribution of an ideal scoring function that matches the ranking by actual relevance. The F^{-1} transformation is called *standardization step*, it is common across all engines participating in a fusion or distributed setup, and is considered critical to the method for compensating for potential individual system biases.

  Further study of this method in [1] demonstrated that this is a very promising method, albeit unnecessarily complicated.

  The same study proposed the Aggregate Historical CDF Simplified according to which:

$$\mathrm{HIS} : s'(l) - P(S_{\mathrm{HIS}} \leq s(l)) \quad (13)$$

  where HIS refers to the fact that historical queries are used for aggregating the SD that the random variable SHIS follows. HIS normalizes input scores $s(l)$ to the probability of a historical query scoring at or below $s(l)$. The aggregate historical SD is an average which can be seen as produced by an ’average’ historical query. In this respect, HIS normalizes the SD of the ’average’ query to uniform in $[0,1]$. This is equivalent to the Cormack model [16], assuming such an ’average’ query is sensible and exists. It should be noted that this approach is used for the first time in image retrieval.

  As historical queries we used 50 images drawn randomly from the database. Since HIS returns probabilities, the natural combination would be multiplication; addition gave inferior results in initial experiments.

We evaluated with two measures: the Average Normalized Modified Retrieval Rank (ANMRR), and the Mean Average Precision (MAP). Since the goal of fusion is to achieve better results than those achieved by any of the CCDs in isolation, we used the performance of SpCD as a baseline, as shown in Tab. 3.

All fusion methods beat the baseline. Best effectiveness overall is achieved by Z-score which beats the baselines of the individual CCDs by wide margins. Both versions (with the mean or median) perform similarly. While HIS performs better than IRP and close to BC, it lags behind Z-score and the bare CombSUM. We tried using more than 50 historical queries (up to 1000) in order to deduce smoother normalization functions, but the effectiveness of HIS did not improve. This is in line with [1], where 50 queries were deemed sufficient to achieve a performance plateau.

Descriptor	ANMRR	MAP
JCD	0.3554	0.5899
BTDH	0.4015	0.5555
SpCD	**0.3081**	**0.6311**
CombSUM	0.2491	0.7121
BC + CombSUM	0.2678	0.6848
Z-score with Mean + CombSUM	**0.2400**	**0.7194**
Z-score with Median + CombSUM	0.2420	0.7193
IRP	0.2729	0.6674
HIS + multiplication	0.2664	0.6846

Tab. 3. Late Fusion Results.

The performance of the bare CombSUM is remarkable. Although it is considered a naive method [2], it is found to be effective and robust. On further investigation it transpired that the reason for this is the similarity of the SDs, in both shape and range, across the CCDs (Fig. 4).

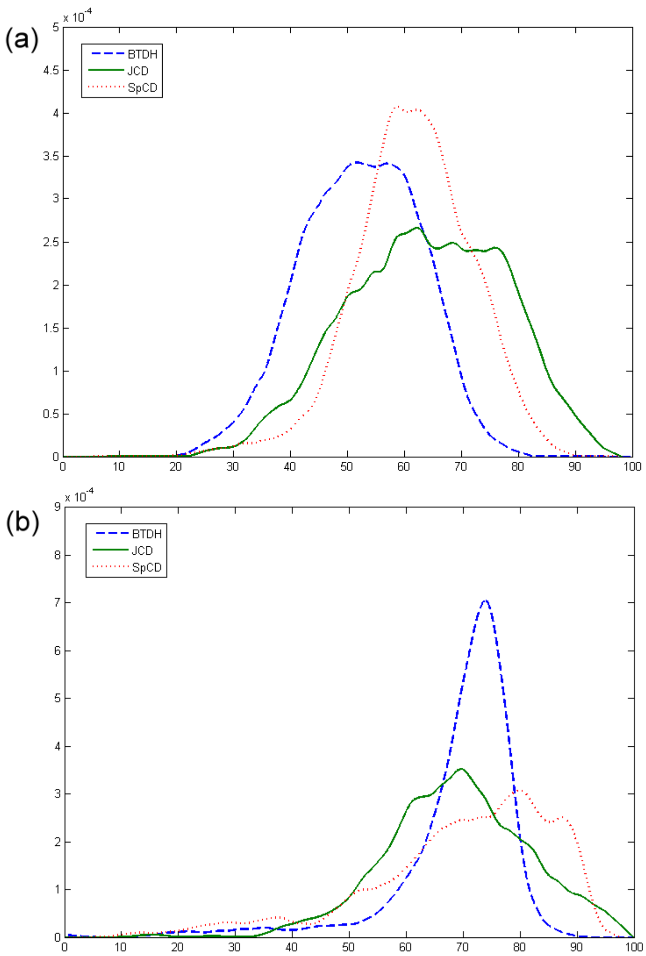

Fig. 4. SDs of the three CCDs for 2 queries.

5. Distributed Image Retrieval

With the growing popularity of digital databases, the focus of research in the area of CBIR has shifted toward content query from distributed databases. A peer-clustering model for the query is proposed in [26], with the assumption that the image collection at each peer node falls under one category. In [20], a novel approach is introduced studying practical scenarios where multiple image categories exist in each individual database in a distributed storage network. In [17], the behavior of a CBIR engine in an interactive distributed environment is examined. In the latter approach, the query image is sent to all registered databases in the network. The response of each database is then collected and transferred to a local server where a supervised relevance identification approach is applied to identify the final outcome of the search. The issue of fusing results returned by different image repositories is also examined in [3].

In this section, we assume databases with disjointed content, and simply merge results by first normalizing the scores assigned to retrieved images by the individual libraries. The libraries in the current setup are described by a variety of single compact composite descriptors. The query is forwarded to all libraries, without any resource selection. Each library extracts from the query its own descriptor and executes the search. In [2] is postulated that effective normalization methods should be non-linear taking into account the shape of score distributions (SDs) especially for non-text descriptors where a wider variety of SDs is assumed. In this respect, we adapt and employ two non-linear methods. We compare against linear functions, as well as the simple ranked-based merging with round-robin.

In the previous section, for fusing result-lists obtained for individual CCDs from a heterogeneous database, we found that simple score normalization and combination methods, such as combining by adding bare or linearly normalized scores with Z-score, work best. This was due to the similarity of the SDs, in both shape and range, across the CCDs, resulting in compatible scores. This implies that merging without normalization may be a good baseline.

In order to overcome the baseline, we initially attempted the classic round-robin method. Round-robin merges results by simply sorting ranks produced by individual descriptors, ignoring scores.

We also tried the Min/Max method. Min/Max normalization maps the resulting score range linearly to the $[0,1]$ interval. Its two obvious drawbacks are: outlier scores have a large impact, and it creates a round-robin effect at the merged ranking due to the top normalized score in all engines being 1.

We then attempted to merge the results with the Z-SCORE and Z-SCORE with Median methods.

For non-linear normalization we attempted the HIS method, analyzed in Section 4. Having no historical queries available, we used 50 images drawn randomly from the union of databases (HIS-rfu). We also tried the alternative method of using 50 random images drawn for each database from its content (HIS-rfe).

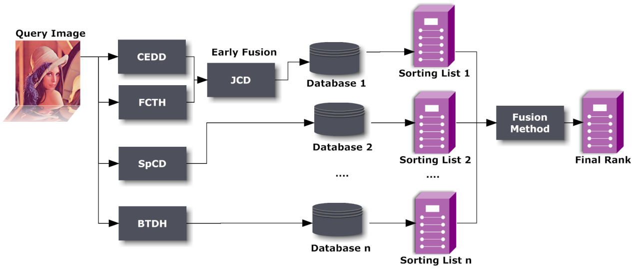

Fig. 5. Distributed Image Retrieval Implementation Process.

We created a distributed image retrieval testbed consisting of 6 databases with disjointed content of the same type: 9000 grayscale images from the IRMA 2005 database indexed with the BTDH are randomly partitioned to 2 libraries; 10200 natural color images from the Nister indexed with the JCD partitioned to 2 libraries; 1030 artificially generated images from the Flags database [24] indexed with the SpCD partitioned to 2 libraries. We used 40 queries: the first 20 natural color image queries from the NISTER database and the first 20 grayscale queries of the IRMA 2005 database.

Normal. Method	40 Queries			
	ANMRR	MAP	P@1	P@5
No-Normaliz.	**0.3812**	**0.5520**	**0.6750**	0.5150
Round Robin	0.6464	0.2126	0.1750	0.2350
MinMax	0.6450	0.2794	0.1750	0.2800
Z-Score Mean	0.4200	0.4408	0.3500	0.2800
Z-Score Median	0.4230	0.4382	0.3500	0.2800
HISrfu	0.4475	0.5068	0.6250	**0.5400**
HISrpe	0.4161	0.5170	**0.6750**	0.4900

Tab. 4. Distributed Image Retrieval Results.

We evaluated using four measures: the Average Normalized Modified Retrieval Rank (ANMRR), the Mean Average Precision (MAP), and the Precision at 1 and 5.

Tab. 4 summarizes the merging results. The effectiveness of round-robin depends on the order of the engines. To deal with this arbitrariness, we repeated the experiment 6 times with random engine orderings, and presented the average effectiveness; round-robin is clearly the worst method.

We removed the engine ordering arbitrariness similarly for the top result per engine also for Min/Max, which performs slightly better than round-robin. Z-score is the best performer of both linear methods and much better than round-robin.

There is a clear advantage of the non-linear HIS methods over the linear ones, although they lag slightly behind in the bare score merging. Using compatible descriptors from the family of CCDs and merging with bare non-normalized scores works best.

6. Conclusions

This paper presented 3 experiments. We initially attempted a technique combining CEDD and FCTH into a fused descriptor, the Joint Composite Descriptor (JCD). This descriptor was tested on various image benchmarking databases and had better results than the descriptors which formed it. We then tried techniques for fusing retrieval results obtained from a heterogeneous image database using multiple CCDs individually. This type of fusion, known as late fusion, is found to be a viable method for retrieving from heterogeneous databases, which improves effectiveness over single descriptor baselines even with simple score normalization and combination methods.

While [29] postulates that effective normalization methods should be non-linear taking into account the shape of SDs especially for non-text descriptors where a wilder variety of SDs is assumed we found these claims to be not necessarily true. Using compatible descriptors from the family of CCDs, combined with adding bare or linearly normalized scores with Z-score works best.

Finally, the use of CCDs in distributed image retrieval was attempted. In observing the results, we found that non-linear normalizations are better than linear. By using compatible descriptors from the family of CCDs, merging by bare non-normalized scores works best.

References

[1] ARAMPATZIS, A., KAMPS, J. A signal-to-noise approach to score normalization. In *Proceeding of the 18th ACM Conference on Information and Knowledge Management*. Hong Kong (China), 2009, p. 797 – 806.

[2] ARAMPATZIS, A., ROBERTSON, S., KAMPS, J. Score distributions in information retrieval. In *Proceedings of the 2nd International Conference on Theory of Information Retrieval: Advances in Information Retrieval Theory*. Cambridge (UK), 2009, p. 139 – 151.

[3] BERRETTI, S., DEL BIMBO, A., PALA, P. Merging results for distributed content based image retrieval. *Multimedia Tools and Applications*, 2004, vol. 24, no. 3, p. 215 – 232.

[4] CHATZICHRISTOFIS, S., BOUTALIS, Y. FCTH: Fuzzy color and texture histogram-a low level feature for accurate image retrieval. In *Proceedings of the 2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS'08*. Klagenfurt (Austria), 2008, p. 191 – 196.

[5] CHATZICHRISTOFIS, S., BOUTALIS, Y. Content based radiology image retrieval using a fuzzy rule based scalable composite descriptor. *Multimedia Tools and Applications*, 2010, vol.46, no. 2-3, p. 493 – 519.

[6] CHATZICHRISTOFIS, S., BOUTALIS, Y., LUX, M. SPCD - spatial color distribution descriptor. a fuzzy rule based compact composite descriptor appropriate for hand drawn color sketches retrieval. In *2nd International Conference on Agents and Artificial Intelligence ICAART*. Valencia (Spain), 2010, p. 58 – 63.

[7] CHATZICHRISTOFIS, S. A., BOUTALIS, Y. S., LUX, M. Selection of the proper compact composite descriptor for improving content based image retrieval. In *Signal Processing, Pattern Recognition*

and Applications SPPRA 2009. Innsbruck (Austria), 2009, p. 134 – 140.

[8] CHATZICHRISTOFIS, S. A., BOUTALIS, Y. S., LUX, M. Img(rummager): An interactive content based image retrieval system. In *Second International Workshop on Similarity Search and Applications SISAP 2009.* Prague (Czech Republic), 2009, p. 151 – 153.

[9] CHATZICHRISTOFIS, S. A., ZAGORIS, K., BOUTALIS, Y. S., PAPAMARKOS, N. Accurate image retrieval based on compact composite descriptors and relevance feedback information. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, 2010, vol. 24, no. 2, p. 207 – 244.

[10] CHATZICHRISTOFIS, S. A., BOUTALIS, Y. S. CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. In6th *International Conference on Computer Vision Systems ICVS 2008.* Santorini (Greece), 2008.

[11] DATTA, R., JOSHI, D., LI, J., WANG, J. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys,* 2008, vol. 40, no. 2, p. 1 – 60.

[12] DESELAERS, T., KEYSERS, D., NEY, H. Features for image retrieval: An experimental comparison. *Information Retrieval*, 2008, vol. 11, no. 2, p. 77 – 107.

[13] FRANK HSU, D., TAKSA, I. Comparing rank and score combination methods for data fusion in information retrieval. *Information Retrieval*, 2005, vol. 8, no. 3, p. 449 – 480.

[14] FU, X., LI, Y., HARRISON, R., BELKASIM, S. Content-based image retrieval using gabor-zernike features. In *18th International Conference on Pattern Recognition.* Hong Kong (China), 2006, vol. 2.

[15] GUSTAFSON, D., KESSEL, W. Fuzzy clustering with a fuzzy covariance matrix. In *1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes.* 1978, vol. 17., p. 761 – 766.

[16] HAWKING, D., ROBERTSON, S. On collection size and retrieval effectiveness. *Information Retrieval*, 2003, vol. 6, no. 1, p. 99 – 105.

[17] JARRAH, K., GUAN, L. Content-based image retrieval via distributed databases. In*International conference on Content-based image and video retrieval.* Niagara Falls (Canada), 2008, p. 389 – 394.

[18] JOVIC, M., HATAKEYAMA, Y., DONG, F., HIROTA, K. Image retrieval based on similarity score fusion from feature similarity ranking lists. *Lecture Notes in Computer Science*, 2006, vol. 4223, p. 461 – 470.

[19] KAMPS, J., DE RIJKE, M., SIGURBJORNSSON, B. Combination methods for crosslingual web retrieval. *Lecture Notes in Computer Science,* 2006, vol. 4022, p. 856 – 864.

[20] LEE, I., GUAN, L. Semi-automated relevance feedback for distributed content based image retrieval. In *International Conference on Multimedia and Expo ICME 2004.* Taipei (Taiwan), 2004, vol. 3., p. 1871 – 1874.

[21] LUX, M., CHATZICHRISTOFIS, S. Lire: lucene image retrieval: an extensible java cbir library. In *Proceedings of the 16th International Conference on Multimedia*. Vancouver (Canada), 2008.

[22] M. BLESCHKE , R. MADONSKI, R. RUDNICKI. Image retrieval system based on combined mpeg-7 texture and colour descriptors. In *16th International Conference Mixed Design of Integrated Circuits and Systems MIXDES 2009.* Lodz (Poland), 2009, p. 635 – 639.

[23] MANJUNATH, B., OHM, J., VASUDEVAN, V., YAMADA, A. Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 2001, vol. 11, no. 6, p. 703 – 715.

[24] MANJUNATH, B., SALEMBIER, P., SIKORA, T. *Introduction to MPEG-7: Multimedia Content Description Interface*. New York: Wiley, 2002.

[25] MPEG-7. *Subjective Evaluation of the MPEG-7 Retrieval Accuracy Measure (ANMRR).* ISO/WG11, Doc. M6029, 2000.

[26] NG, C., SIA, K., CHANG, C. Advanced peer clustering and firework query model in the peer-to-peer network. In *12th International World Wide Web Conference (WWW).* Budapest (Hungary), 2003.

[27] NISTER, D., STEWENIUS, H. Scalable recognition with a vocabulary tree. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2006, vol. 5, p. 2161 – 2168.

[28] SCHAEFER, G., STICH, M. Ucid-an uncompressed colour image database. *Storage and Retrieval Methods and Applications for Multimedia*, 2004, vol. 5307, p. 472 – 480.

[29] SMEULDERS, A., WORRING, M., SANTINI, S., GUPTA, A., JAIN, R. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, vol. 22, no. 12, p. 1349 – 1380.

[30] SPYROU, E., LE BORGNE, H., MAILIS, T., COOKE, E., AVRITHIS, Y., O CONNOR, N. Fusing mpeg-7 visual descriptors for image classification. *Lecture Notes in Computer Science*, 2005, vol. 3697, p. 847 – 852.

[31] TAMURA, H., MORI, S., YAMAWAKI, T. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 1978, vol. 8, no. 6, p. 460 – 473.

[32] WANG, J., LI, J., WIEDERHOLD, G. Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, vol. 23, no. 9, p. 947–963.

[33] WON, C., PARK, D., PARK, S. Efficient use of mpeg-7 edge histogram descriptor. *Etri Journal*, 2002, vol. 24, no. 1, p. 23 – 30.

[34] WONG, K., CHEUNG, K., PO, L. Mirror: an interactive content based image retrieval system. In *IEEE International Symposium on Circuits and Systems ISCAS 2005.* Kobe (Japan), 2005, vol. 2, p. 1541 – 1544.

1. ANMRR Calculation

The average rank AVR(q) for query q is:

$$\mathrm{AVR}(q) = \sum_{k=1}^{NG(q)} \frac{Rank(k)}{NG(q)} \tag{14}$$

where

- NG(q) is the number of ground truth images for query q. A ground truth is defined as a set of visually similar images.
- $K = min(X_{\mathrm{NG}} \times \mathrm{NG}(q), 2 \times \mathrm{GTM})$.
- $\mathrm{GTM} = max(\mathrm{NG})$.
- If $\mathrm{NG}(q) > 50$ then $X_{\mathrm{NG}} = 2$ else $X_{\mathrm{NG}} = 4$.
- $Rank(k)$ is the retrieval rank of the ground truth image. Consider a query. Assume that as a result of the retrieval, the kth ground truth image for this query q is found at a position R. If this image is in the first K retrievals then $Rank(k) = R$ else $Rank(k) = (K+1)$.

The modified retrieval rank is:

$$\mathrm{MRR}(q) = \mathrm{AVR}(q) - 0.5 \times [1 + \mathrm{NG}(q)]. \quad (15)$$

Note that MRR is 0 in case of perfect retrieval. The normalized modified retrieval rank is computed as follows:

$$\mathrm{NMRR}(q) = \frac{\mathrm{MRR}(q)}{1.25 \times K - 0.5 \times [1 + \mathrm{NG}(q)]}. \quad (16)$$

Finally the average of NMRR over all queries is defined as:

$$\mathrm{ANMRR} = \frac{1}{Q} \sum_{q=1}^{Q} \mathrm{NMRR}(q) \quad (17)$$

where

- Q is the total number of queries.

About Authors...

Savvas A. CHATZICHRISTOFIS received the diploma of Electrical Engineer in 2005 from the Democritus University of Thrace (DUTH), Greece. He is currently a research and teaching assistant and studies towards the Ph.D. degree at the Department of Electrical and Computer Engineering, Democritus University of Thrace. His research interests include image retrieval, image processing and analysis, document analysis, machine intelligence, neural networks, Fuzzy Logic and pattern recognition. He is an establishing member of the ≪Greek Open Source Adherents' Club≫. He is also Member in Cyprus Scientific and Technical Chamber since 2005, licentiate in the fields of Electronics, Information Science and Electrical Mechanics.

Avi ARAMPATZIS graduated from the Computer Engineering and Informatics department, University of Patras, Greece, in 1996. He then moved to the Netherlands where he completed his PhD in Adaptive Document Filtering at the University of Nijmegen in 2001. Since then, he has been a researcher in various academic institutions, such as the Computing Science Department of University of Utrecht and the Department of Media Studies of University of Amsterdam, as well as an independent consultant. Currently, he is a lecturer in databases at the Department of Electrical and Computer Engineering, Democritus University of Thrace, Greece. His current interests are in Information Retrieval and related areas.

Yiannis S. BOUTALIS received the diploma of Electrical Engineer in 1983 from the Democritus University of Thrace (DUTH), Greece and the PhD degree in Electrical and Computer Engineering (topic Image Processing) in 1988 from the Computer Science Division of National Technical University of Athens, Greece. Since 1996, he serves as a faculty member, at the Department of Electrical and Computer Engineering, DUTH, Greece, where he is currently an associate professor and director of the automatic control systems lab. He served as an assistant visiting professor at University of Thessaly, Greece, and as a visiting professor in Air Defense Academy of General Star of air forces of Greece. He also served as a researcher in the Institute of Language and Speech Processing (ILSP), Greece, and as a managing director of the R&D SME Ideatech S.A, Greece, specializing in pattern recognition and signal processing applications. His current research interests are focused in the development of Computational Intelligence techniques with applications in Control, Pattern Recognition, Signal and Image Processing Problems.