Late Fusion of Compact Composite Descriptors for Retrieval from Heterogeneous Image Databases

Savvas A. Chatzichristofis Avi Arampatzis
Department of Electrical and Computer Engineering
Democritus University of Thrace
Xanthi, Greece
schatzic@ee.duth.gr, avi@ee.duth.gr

ABSTRACT

Compact composite descriptors (CCDs) are global image features, capturing more than one types of information at the same time in a very compact representation. Their quality has so far been evaluated in retrieval from several homogeneous databases containing images of only the type that each CCD is intended for, and has been found better than other descriptors in the literature such as the MPEG-7 descriptors. In this study, 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.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*retrieval models, search process*; H.2.5 [**Database Management**]: Heterogeneous Databases
General Terms: Measurement, Experimentation, Theory
Keywords: Image Retrieval, CCD, Fusion, Normalization, Combination

1. INTRODUCTION

Fusion in image retrieval goes hand-in-hand with practical, viable system development, which is critical for the future of image retrieval research [3]. Two main approaches to fusion have been taken: *early fusion*, where multiple image descriptors are composed to form a new one before index time [8], and *late fusion*, where result-lists from individual descriptors are fused during query time [4, 5], as in text meta-search. While early fusion has been common, late fusion still remains an under-explored possibility.

Compact Composite Descriptors (CCDs) are global image features for content-based image retrieval. CCDs capture more than one types of information at the same time in a very compact representation. An example of early fusion is the Joint Composite Descriptor (JCD), created by combining two CCDs [8]: the Color and Edge Directivity Descriptor (CEDD), and the Fuzzy Color and Texture Histogram (FCTH). The CEDD and FCTH can be seen as products of early fusion themselves, since they both combine color and texture information. Both descriptors are developed for color natural images, and experiments have shown that they perform better than the MPEG-7 descriptors and other descriptors in the related literature.

The Brightness and Texture Directionality Histogram, or BTDH, is proposed for grayscale and medical radiology images [7], and has been found to perform better than other descriptors in several benchmarking databases. The recently proposed Spatial Color Distribution (SpCD) combines color information and its spatial distribution in a quantized histogram [9]. The SpCD is considered suitable for colored graphics, since they contain a relatively small number of colors and less texture regions than natural color images.

The quality of the aforementioned CCDs has so far been evaluated in retrieval from homogeneous benchmarking databases, containing images of only the type that each CCD is intended for. For example, the JCD is tested on NISTER [6] and Wang 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.

2. A CCD LATE FUSION EXPERIMENT

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 [9].

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.

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. Five fusion methods were tested:

- CombSUM: It is the addition of all scores per image, without normalization.

- BC+CombSUM: 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 less than the previous. Votes across ranked-lists are naturally combined with CombSUM.

SIGIR'10, July 19–23, 2010, Geneva, Switzerland.
ACM 978-1-60558-896-4/10/07.

- Z-score+CombSUM: Z-score is a linear normalization per query which maps each score to its number of standard deviations above or below the mean score. It is more suitable for non-skewed score distributions (SDs) where the mean would make more sense. In this respect, we also tried it with the median score instead of the mean. We present results with CombSUM; we also tried multiplication, but it gave inferior performance.

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

- HIS+multiplication: HIS is a non-linear normalization which maps each score to the probability of a historical query scoring a collection image below that score. It is recently proposed by [1] and found to be robust and effective in a distributed retrieval testbed. 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 evaluate with two measures: the Average Normalized Modified Retrieval Rank (ANMRR), and the Mean Average Precision (MAP). ANMRR is the evaluation measure used in all the MPEG-7 color core experiments; it ranges between 0 and 1, with 0 being the maximum effectiveness.

Since the goal of fusion is to achieve better results than those achieved by any of the CCDs in isolation, we use the performance of SpCD as baseline, as shown in Table 1.

Descriptor / Fused	40 Queries	
	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

Table 1: Experimental Results.

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 lacks 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.

The performance of the bare CombSUM is remarkable. Although it is considered a naive method [2], it is found effective and robust. On a further investigation it turned out that the reason for this is the similarity of the SDs, in both shape and range, across the CCDs (Fig. 1).

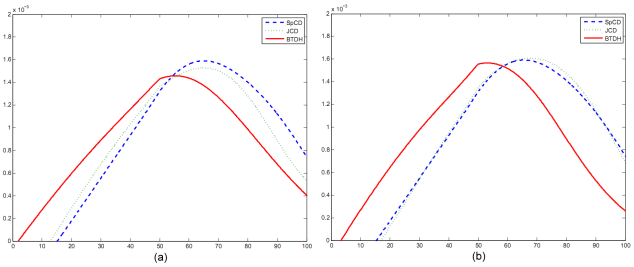

Figure 1: SDs of the three CCDs for 2 queries.

3. CONCLUSIONS

We investigated methods for fusing retrieval results obtained from an heterogeneous image database using multiple descriptors 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 [2] 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. By using compatible descriptors from the family of CCDs, combining by adding bare or linearly normalized scores with Z-score works best.

4. REFERENCES

[1] A. Arampatzis and J. Kamps. A signal-to-noise approach to score normalization. In *CIKM*, pages 797–806. ACM, 2009.

[2] A. Arampatzis, S. Robertson, and J. Kamps. Score distributions in information retrieval. In *ICTIR*, volume 5766 of *LNCS*, pages 139–151. Springer, 2009.

[3] R. Datta, D. Joshi, J. Li, and J. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60, 2008.

[4] M. Jovic, Y. Hatakeyama, F. Dong, and K. Hirota. Image retrieval based on similarity score fusion from feature similarity ranking lists. In *FSKD*, volume 4223 of *LNCS*, pages 461–470. Springer, 2006.

[5] M. Bleschke , R. Madonski and R. Rudnicki. Image retrieval system based on combined mpeg-7 texture and colour descriptors. In *MIXDES*, pages 635 – 639, 2009.

[6] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *Proc. CVPR*, volume 5, pages 2161–2168. Citeseer, 2006.

[7] S. A. Chatzichristofis and Y. S. Boutalis. Content based radiology image retrieval using a fuzzy rule based scalable composite descriptor. *Multimedia Tools and Applications*, 46:493–519, 2010.

[8] S. A. Chatzichristofis, Y. S. Boutalis, and M. Lux. Selection of the proper compact composite descriptor for improving content based image retrieval. In *SPPRA*, pages 134–140. ACTA Press, 2009.

[9] S. A. Chatzichristofis, Y. S. Boutalis and M. Lux. SpCD—spatial color distribution descriptor. A fuzzy rule based compact composite descriptor appropriate for hand drawn color sketches retrieval. In *ICAART*, pages 58–63, 2010.