

# Efficient Content-Based Information Retrieval: A New Similarity Measure for Multimedia Data

Christian Beecks Thomas Seidl  
Data Management and Data Exploration Group  
RWTH Aachen University, Germany  
{beecks,seidl}@cs.rwth-aachen.de

## Abstract

**Content-based information retrieval of multimedia data is a great and attractive challenge which raises numerous research activities. As multimedia data become ubiquitous in our daily lives, information retrieval systems have to adapt their retrieval performance to different situations in order to efficiently satisfy the users' information needs anytime and anywhere. To enhance the content-based multimedia information retrieval process, we propose an efficient similarity measure based on flexible feature representations. We define the similarity model for content-based information retrieval and show its feasibility on real world multimedia data. Furthermore, we briefly discuss future research directions which we plan to investigate.**

## 1. INTRODUCTION

Information retrieval systems are ubiquitous in our daily lives. They support us to store, manage and access voluminous information based on the systems' underlying data. Thus, the primary aim of such information retrieval systems is to supply user's with relevant and useful information in order to satisfy their information needs. To this end, the system should perform the information retrieval process of a user-specific query in an effective and also efficient way.

Prominent examples of information retrieval systems are search engines in the World Wide Web. They enable us to do a search for interesting web pages, popular video clips, famous research papers, gorgeous images, and so on. Almost all multimedia information inside and outside the World Wide Web are made accessible via appropriate retrieval systems.

In order to fulfill retrieval and browsing tasks in the user's sense, information retrieval systems use different kinds of models to represent their accessible data through additional semantic and syntactic information. These information reflect the contents of the stored data and represent the core of each information retrieval process.

To perform the content-based information retrieval process in an efficient way, we propose a similarity model that extends the classic vector space model. We allow data objects to store multiple feature vectors and compare these weighted sets with our new similarity measure. Thus, the feature representation supports us to dynamically extract contents of multimedia data, whereas the new similarity measure efficiently compares these feature representations with each other.

We organize the paper's structure as follows: In Section 2, we briefly review existing works and backgrounds. We introduce our new similarity measure in Section 3 and give an impression of the first practical retrieval results in Section 4. We briefly describe future directions in Section 5 and conclude our paper in Section 6.

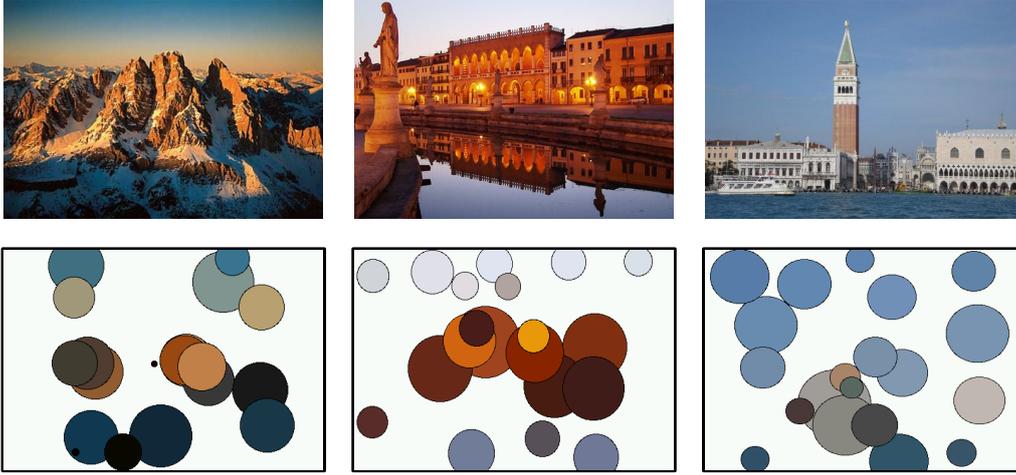


FIGURE 1: Three sample images in the top row with their signatures in the bottom row.

## 2. RELATED WORK AND BACKGROUND

The methodology of information retrieval covers a broad range of distinct interdisciplinary research areas like modeling and indexing data contents, searching information, evaluating retrieval results, designing user interfaces, and so forth. To get an impression and overview of this widespread field, we recommend the classic books of van Rijsbergen [1] and Salton et al. [2] and, for instance, the modern books of Beaza-Yates and Ribeiro-Neto [3] and Manning et al. [4].

Research in these different fields yields to a multitude of modern information retrieval systems. Each of the systems' core is the underlying information retrieval model. These models specify the way to store and access semantic and syntactic information of the data. Common classic models for information retrieval are the Boolean model, the vector space model [5], and the probabilistic model [6, 7]. These models are different in the way how they store the data and how they perform different user tasks and queries.

In the present work, however, we focus on an extended vector space model that is suitable for content-based multimedia information retrieval. Therefore, we store the contents of each data object in a weighted set of vectors and call this set a signature (cf. the work of Rubner et al. [8]).

### Definition 1. (Signature)

Given a data object  $o$ , the corresponding signature  $S_o$  with length  $n^s$  is defined as

$$S_o = \{(v_i^s, w_i^s) \mid i = 1, \dots, n^s\},$$

where  $n^s$  is the number of vectors  $v_i^s \in \mathcal{R}^n$  with the corresponding weights  $w_i^s \in \mathcal{R}^+$ . The benefit of this feature representation, in contrast to the traditional vector approach, is the flexibility and adaptability of signatures to cover dynamic properties of vivid multimedia data. By extracting signatures from multimedia data objects, information retrieval systems are able to capture global as well as local object's properties to ensure effective content-based information retrieval processes.

Figure 1 shows an example of signatures from three different images. In this example, we extract signatures comprising 20 vectors with the underlying feature dimension equal to seven. These seven dimensions include two position components, three color components, one contrast component, and one coarseness component. In the figure, we visualize the vectors according to their position information with colored circles. The diameters of the circles reflect the weights of the vectors. As we see in this example, signatures represent a flexible and also compressed way to store data contents automatically without any additional semantic structure information.

To judge contents of multimedia objects based on their automatically extracted signatures, information retrieval systems compute similarities among signatures. Common similarity measures for signatures are the Earth Mover's Distance [8] or the Hausdorff Distance [9]. Both are applicable to multimedia data, because they use an adaptable ground distance to determine the distance between the signatures' vectors. Nevertheless, both distances limit the performance of modern information retrieval systems: Earth Mover's Distances have high computation times, whereas Hausdorff Distances suffer from the missing possibility to consider the weights of the signatures.

### 3. A NEW SIMILARITY MEASURE

In this section, we present our new similarity measure for multimedia information retrieval based on signatures. Therefore, we adapt the well-known concept of quadratic form distance measures [10, 11] from simple feature vectors to the more flexible feature representation of signatures. In order to compare two signatures, the challenging task consists in matching all of the vectors of both signatures among each other. To achieve this in the distance computation, we propose the following definition.

#### Definition 2. (Signature Distance)

Given two signatures  $Q = \{(v_i^q, w_i^q) \mid i = 1, \dots, m\}$  and  $P = \{(v_i^p, w_i^p) \mid i = 1, \dots, n\}$  of any length  $m, n$ , respectively, we define the Signature Distance measure  $SD$  as

$$SD(Q, P) = \sqrt{(\tilde{w}_q - \tilde{w}_p) \cdot A \cdot (\tilde{w}_q - \tilde{w}_p)^T},$$

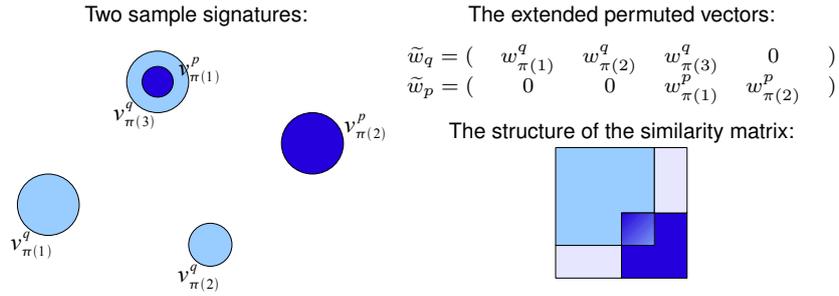
where  $\tilde{w}_q \in \mathcal{R}^{n+m-k}$  and  $\tilde{w}_p \in \mathcal{R}^{n+m-k}$  are the extended permuted vectors with the same length which result from the signatures' weights  $w_i^q \in \mathcal{R}^+$  and  $w_i^p \in \mathcal{R}^+$  as follows:

$$\begin{aligned} \tilde{w}_q &= \overbrace{(w_{\pi(1)}^q, \dots, w_{\pi(n-k)}^q)}^{(n-k) \text{ weights of } Q}, \underbrace{(w_{\pi(n-k-1)}^q, \dots, w_{\pi(n)}^q)}_{k \text{ common weights}}, (0, \dots, 0) \\ \tilde{w}_p &= (0, \dots, 0, \underbrace{(w_{\pi(1)}^p, \dots, w_{\pi(k)}^p)}_{k \text{ common weights}}, \underbrace{(w_{\pi(k+1)}^p, \dots, w_{\pi(m)}^p)}_{(m-k) \text{ weights of } P}). \end{aligned}$$

The extended permuted vectors  $\tilde{w}_q$  and  $\tilde{w}_p$  in the preceding definition consist of three blocks, according to the vectors  $v_i^q$  and  $v_i^p$  of the signatures. The first and the last block exclusively comprise weights from signatures  $Q$  and  $P$ , respectively, sharing no common vectors, whereas the block in the middle only consists of weights from  $Q$  and  $P$  sharing the same vectors. As a result, the permutation  $\pi$  aligns the weights of the signatures to each other and enables the multiplication with the similarity matrix  $A \in \mathcal{R}^{(n+m-k) \times (n+m-k)}$  which is determined dynamically per distance computation. The dynamically generated similarity matrix  $A$  models the similarity among all vectors and depends on the compared signatures.

We give the following example to illustrate the distance computation of the proposed Signature Distance measure. For this purpose, we depict two signatures on the left-hand side in Figure 2. Signature  $Q$  consists of three vectors  $v_{\pi(i)}^q$  which are depicted with the color light-blue, whereas signature  $P$  consists of two vectors  $v_{\pi(i)}^p$  which are depicted with the color dark-blue. Both signatures share one common vector, namely  $v_{\pi(3)}^q$  in signature  $Q$  and  $v_{\pi(1)}^p$  in signature  $P$ . Thus, the extended permuted vectors have the total length of five and they share exactly one common component: the weights  $w_{\pi(3)}^q$  and  $w_{\pi(1)}^p$ .

Based on the extended permuted vectors of the signatures which have to be compared, the similarity matrix is generated. In this example, the similarity matrix comprises three major blocks. The light-blue block and the dark-blue block determines the similarity among the vectors of signature  $Q$  and those of signature  $P$ , respectively. The overlapping block between the light-



**FIGURE 2:** Two sample signatures on the left-hand side, the extended permuted vectors and the structure of the similarity matrix on the right-hand side.

and dark-blue blocks models the similarity of the common vectors in both signatures. As in this example exists only one vector which appears in both signatures, this block consists of exactly one value. The other blocks of the similarity matrix model the similarities among the vectors of signature  $Q$  and signature  $P$ . As a result, the distance computation is finalized by multiplying the difference of the extended permuted vectors  $\tilde{w}_q$  and  $\tilde{w}_p$  with the generated similarity matrix according to Definition 2.

In order to generate the similarity matrix, we have to determine the similarities between two vectors. We recommend to choose a function which is inverse proportional to a distance function. Such a function guarantees the highest similarity value for the same vectors and lower similarity values for different vectors.

#### 4. PRACTICAL RETRIEVAL RESULTS

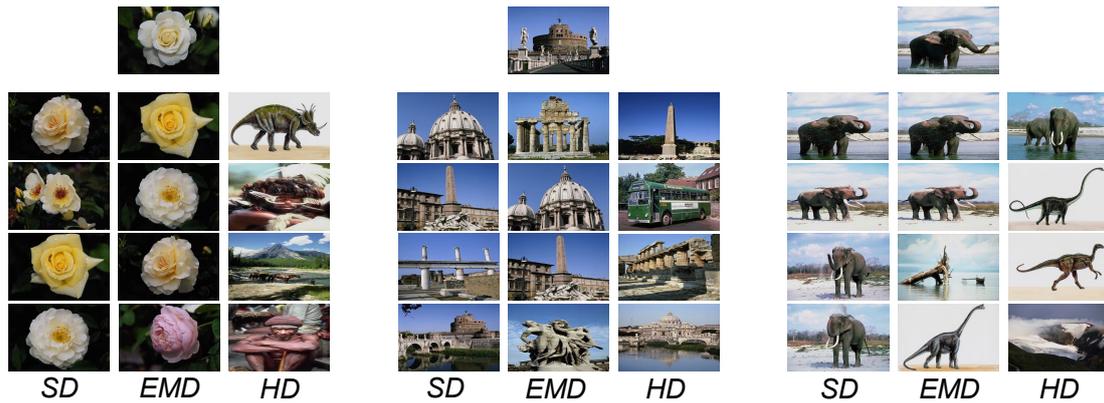
In this section, we present the first practical retrieval results of our new similarity measure. For this purpose, we conducted several similarity queries on the *Wang* image collection [12] which consists of 1.000 images from 10 different themes. Based on a  $k$ -means clustering in the feature space, we extracted 20-dimensional signatures of the images including position, color, contrast, and coarseness information.

Figure 3 visualizes some of the retrieval results which are given below the three different query images with the most similar images on top of each column. The columns of each query show the ranked results of the proposed Signature Distance ( $SD$ ), the Earth Mover's Distance ( $EMD$ ), and the Hausdorff Distance ( $HD$ ), respectively. The figure reveals that the results of the Signature Distance and Earth Mover's Distance continually have a higher quality compared to that of the Hausdorff Distance. Comparing the Earth Mover's Distance with the Signature Distance, we see that the latter has a slightly higher perceptual retrieval quality. Thus, we conclude that the computed similarities of the Signature Distance are well comparable to the ones of the Earth Mover's Distance. In order to verify the perceptual results, we list the aggregated mean average precision [4] values measured over 100 randomly chosen queries, based on the *Wang* image collection, in the following table.

|                         | mean average precision |
|-------------------------|------------------------|
| Quadratic Form Distance | 0.51                   |
| Earth Mover's Distance  | 0.50                   |
| Hausdorff Distance      | 0.33                   |

**TABLE 1:** Aggregated mean average precision.

In addition to the quality of the retrieval results, we also measured the computation times needed to generate the resulting rankings. As a result, the Hausdorff Distance has the lowest run-time of 23 milliseconds to generate the ranking. The Signature Distance has a run-time of 76 milliseconds, whereas the Earth Mover's Distance requires 261 milliseconds to finish the retrieval process. The



**FIGURE 3:** Three different query examples and their ranked results. For each query, the columns depict the results of the Signature Distance (*SD*), Earth Mover's Distance (*EMD*), and Hausdorff Distance (*HD*).

experiments were implemented in Java and the run-times were measured on Intel XEON E5345 CPU-based machine with 2.33GHz.

Combining the perceptual qualities and the run-times of the retrieval experiments, we summarize that our new Signature Distance advantages high retrieval qualities and comparatively low computation times. Both properties are fundamental for efficient content-based multimedia information retrieval.

## 5. FUTURE DIRECTIONS

For our future work, we plan to apply the proposed similarity measure on multimedia data to retrieve content-based information. Therefore, we identify the following future research directions on which we plan to contribute with our similarity measure:

In the field of region-based similarity search, we are interested in multimedia objects that are similar to a region of a given query object. Thus, instead of querying multimedia databases with complete signatures, we only use relevant components of query signatures to find those regions in the signatures containing relevant information of the objects. The arising question is, if this relevant components of the query signatures and of the signatures stored in multimedia databases can be determined automatically within the proposed similarity measure.

In addition to region-based similarity search, we plan to focus our research in the field of adaptable similarity search which considers the adaptation of the proposed similarity measure to different user preferences. In order to improve the retrieval quality of content-based similarity search, we plan to examine the properties of the underlying similarity matrix to capture those user preferences.

The third aspect that we want to consider is the content-based retrieval of heavily sized databases. As information retrieval is generally not restricted to a fixed size of the databases, we investigate on techniques to query voluminous data in an efficient way. To support the retrieval process, we plan to study approximation and indexing techniques of the proposed similarity measure.

## 6. CONCLUSION

We presented a new similarity measure based on flexible feature representations for efficient content-based information retrieval of multimedia data. We define and illustrate this similarity measure and show its feasibility and efficiency on real world multimedia data. As a result, we conclude that our new similarity measure combines a low run-time of distance computation with a high retrieval performance. Furthermore, we identify future research directions on which we plan to contribute with our new similarity measure.

## REFERENCES

- [1] C. Rijsbergen, *Information Retrieval*. Butterworth, 1979.
- [2] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, Inc., 1986.
- [3] R. A. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.
- [4] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.
- [5] G. Salton, *The SMART Retrieval System—Experiments in Automatic Document Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1971.
- [6] S. E. Robertson and S. K. Jones, “Relevance weighting of search terms,” *Journal of the American Society for Information Science*, vol. 27, no. 3, pp. 129–146, 1976.
- [7] N. Fuhr, “Probabilistic models in information retrieval,” *Computer Journal*, vol. 35, no. 3, pp. 243–255, 1992.
- [8] Y. Rubner, C. Tomasi, and L. J. Guibas, “The Earth Mover’s Distance as a Metric for Image Retrieval,” *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.
- [9] D. Huttenlocher, G. Klanderman, and W. Rucklidge, “Comparing images using the Hausdorff Distance,” *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol. 15, no. 9, pp. 850–863, 1993.
- [10] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, “Efficient and Effective Querying by Image Content,” *Journal of Intelligent Information Systems*, vol. 3, no. 3/4, pp. 231–262, 1994.
- [11] T. Seidl and H.-P. Kriegel, “Efficient User-Adaptable Similarity Search in Large Multimedia Databases,” in *VLDB*, 1997, pp. 506–515.
- [12] J. Wang, J. Li, and G. Wiederhold, “Simplicity: semantics-sensitive integrated matching for picture libraries,” *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol. 23, no. 9, pp. 947–963, 2001.