A COMPARATIVE STUDY OF SIMILARITY MEASURES FOR CONTENT-BASED MULTIMEDIA RETRIEVAL

Christian Beecks, Merih Seran Uysal, Thomas Seidl

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

ABSTRACT
Determining similarities among data objects is a core task of content-based multimedia retrieval systems. Approximating data object contents via flexible feature representations, such as feature signatures, multimedia retrieval systems frequently determine similarities among data objects by applying distance functions. In this paper, we compare major state-of-the-art similarity measures applicable to flexible feature signatures with respect to their qualities of effectiveness and efficiency. Furthermore, we study the behavior of the similarity measures by discussing their properties. Our findings can be used in guiding the development of content-based retrieval applications for numerous domains.

Keywords— similarity measures, content-based multimedia retrieval, evaluation

1. INTRODUCTION
Today, the world is interconnected by digital data whose volume has been increasing at a high rate. It is immensely desirable for multimedia systems to enable users to access, search, and explore tremendous data arising from applications generating images, videos, audio, or other non-text documents stored in large multimedia databases.

Similarity search is an increasingly important task in multimedia retrieval systems and is widely used in both commercial and scientific applications in various areas, such as copy and near-duplicate detection of images and videos [1, 2, 3, 4, 5] or content-based image, video, and audio retrieval [6, 7, 8, 9, 10, 11]. In order to meet the system requirements and user needs regarding adaptable multimedia retrieval, content-based similarity models have been developed and widely applied in the aforementioned areas. Its key challenge is that the inherent properties of data objects are gathered via feature representations which are used for the comparison of the corresponding data objects depending on their contents.

Given a query object which is either manually specified by a user or automatically generated by an application and a possibly large multimedia database, an appropriate similarity model can determine similarity between the query and each object in the database by computing the distance between their corresponding feature representations. By making use of the computed similarity values determined via distance values, it is possible to retrieve the most similar objects regarding the query. The retrieved objects can then be processed within corresponding applications or explored and evaluated by the users with respect to human perception of similarity.

As the concept of determining similarities among data objects is of crucial importance in multimedia retrieval systems and for the aforementioned applications, we provide insights into major state-of-the-art similarity measures applicable to feature signatures exhibiting flexible content-based representation of multimedia data objects. This will be carried out by extensive experiments and evaluations taking into account both effectiveness and efficiency of the corresponding similarity measures whose performance qualities are compared with each other. Our findings can be used in guiding the development of content-based retrieval applications for numerous domains.

We structure the paper as follows: we first provide preliminary information about feature extraction and common feature representations in Section 2. Then, in Section 3 we survey major state-of-the-art similarity measures applicable to feature signatures for content-based multimedia retrieval. Section 4 will be devoted to experimental evaluation of the similarity measures regarding both effectiveness and efficiency. The discussion of the experimental evaluation results will be given in Section 5. We end this work with the conclusion in Section 6.

2. FEATURE EXTRACTION AND REPRESENTATION
In this section, we describe the common feature extraction process and the frequently used feature representation form, the so-called feature signatures [12]. The extraction of data object features and their aggregation aim at digitizing and compactly storing the data objects’ inherent properties.

The extraction of features and the aggregation of these features for each individual data object are illustrated in Figure 1. In the feature extraction step, each data object is mapped into a set of features in an appropriate feature space $FS$. In the field of content-based image retrieval [6, 8, 9], the feature space frequently comprises position, color, or texture dimensions [13, 14] where each image pixel is mapped to a single feature in the corresponding feature space. In the figure, the features shown via the same color in the feature space belong to the same data object. In this way, the content of each data object is exhibited via its feature distribution in the feature space.
In order to compute similarity between two data objects efficiently, features are aggregated into a compact feature representation form which is depicted as the feature aggregation step in Figure 1. There are two common feature representation forms, feature histograms and feature signatures, which arise from global partitioning of the feature space and local clustering of features for each data object, respectively [12]. The global partitioning of the feature space is carried out regardless of feature distributions of single data objects. For each data object, a single feature histogram is generated where each entry of the histogram corresponds to the number of its features located in the corresponding global partition. In contrast, local clustering of features for a data object results in a feature signature, also named as adaptive-binning histogram [15], which comprises clusters of the object’s features with the corresponding weights and centroids. In the figure, each plus denotes a centroid of an individual cluster which can be identified via its color. For instance, the red pluses denote the centroids of the clusters which correspond to the feature distribution of the red data object in the database.

In this work, we focus on similarity measures applicable to feature signatures which achieve a better balance between expressiveness and efficiency than feature histograms [16]. Furthermore, feature signatures aggregate the data objects’ inherent properties which are expressed via features more flexible than feature histograms and preserve the local appearance of features better than feature histograms. The definition of feature signatures is given below.

**Definition 1 Feature Signature**

Given a feature space $\mathcal{FS}$ and a local clustering $\mathcal{C} = C_1, \ldots, C_n$ of the features $f_1, \ldots, f_k \in \mathcal{FS}$ of object $o$, the feature signature $S^o$ is defined as a set of tuples from $\mathcal{FS} \times \mathbb{R}^+$ as $S^o = \{ (c_i^o, w_i^o), i = 1, \ldots, n \}$, where $c_i^o = \frac{\sum_{f \in C_i} f}{|C_i|}$ and $w_i^o = \frac{|C_i|}{k}$ represent the centroid and weight, respectively.

Intuitively, a feature signature $S^o$ of object $o$ is a set of centroids $c_i^o \in \mathcal{FS}$ with the corresponding weights $w_i^o \in \mathbb{R}^+$ of the clusters $C_i$. Carrying out the feature clustering individually for each data object, we recognize that each feature signature reflects the distribution more meaningfully than any feature histogram. In addition, feature histograms can be regarded as a special case of feature signatures where the clustering is replaced with a global partitioning of the feature space.

We depict three example images with their corresponding feature signatures in Figure 2. We visualize the feature signatures’ centroids from a seven-dimensional feature space (two position, three color, and two texture dimensions) as circles in a two-dimensional position space. The circles’ colors and diameters reflect the colors and weights of the centroids, respectively.

Throughout the present work, we apply an adaptive variant of the $k$-means clustering algorithm to generate adjustable feature signatures, similar to the one proposed in [15]. This clustering algorithm performs like a $k$-means one without specifying the number $k$ of clusters in advance. Instead, the following parameters have to be defined: the maximum radius $R$ of a cluster, the nominal cluster separation $S$, and the minimum number of points in a cluster $M$. For the choice of these parameters, we refer to Section 4.

In the following section, we include a short survey of state-of-the-art similarity measures applicable to feature signatures.

### 3. A SHORT SURVEY OF ADAPTIVE SIMILARITY MEASURES

In this section, we summarize major state-of-the-art similarity measures for feature signatures exhibiting content-based feature representations of multimedia objects. Furthermore, we discuss theoretical fundamentals and ideas behind each similarity measure in order to contribute to our understanding.

Adaptive similarity measures apply a so-called ground distance function to determine distances among feature signatures’ centroids in the feature space. Applicable ground distance functions are for instance those evaluated in [17].

The first similarity measure we present is the Hausdorff Distance [18] which measures the maximum nearest neighbor distance among centroids in both feature signatures. The formal definition of the Hausdorff Distance is given below.

**Definition 2 Hausdorff Distance**

Given two feature signatures $S^q$ and $S^o$ and a ground distance function $d$, the Hausdorff Distance $HD_d$ between $S^q$ and $S^o$ is defined as:

$$HD_d(S^q, S^o) = \max \{ h(S^q, S^o), h(S^o, S^q) \},$$

where

$$h(S^q, S^o) = \max_{c^q \in S^q} \min_{c^o \in S^o} \{ d(c^q, c^o) \}.$$
The Hausdorff Distance is only based on the centroid structure of both feature signatures and does not consider their weights. On top of this, it does not take into account the whole structures of feature signatures which causes a limitation while computing the distance between any two feature signatures.

As an extension for color-based image retrieval, the Perceptually Modified Hausdorff Distance [19] was proposed which uses the information of both weights and centroids of the feature signatures. Below, the formal definition of the Perceptually Modified Hausdorff Distance is given.

**Definition 3 Perceptually Modified Hausdorff Distance**

Given two feature signatures $S^q$ and $S^o$ and a ground distance function $d$, the Perceptually Modified Hausdorff Distance $PMHD_d$ between $S^q$ and $S^o$ is defined as:

$$PMHD_d(S^q, S^o) = \max \left\{ h_w(S^q, S^o), h_w(S^o, S^q) \right\},$$

where $h_w(S^q, S^o) = \frac{\sum_i w_i^q \cdot \min\left\{ \frac{d(c_i^q, c_o^q)}{\min(w_i^q, w_o^q)} \right\}}{\sum_i w_i^q}$.

The computation of the distance between $S^q$ and $S^o$ via the Perceptually Modified Hausdorff Distance requires the determination of the centroid $c_o^q$ located as near as possible with the highest possible weight for each centroid $c_i^q$, and vice versa. In other words, in spite of the consideration of the weight and position information the whole structures of the feature signatures are not taken into consideration.

Another similarity measure is the well-known Earth Mover’s Distance [12, 16] originated in the computer vision domain. Its successful utilization gave raise to numerous applications in different domains. This similarity measure describes the cost for transforming one feature signature into another one. Similarity is considered to be a transportation problem and, thus, is the solution to a minimization problem which can be solved via a specialized simplex algorithm.

**Definition 4 Earth Mover’s Distance**

Given two feature signatures $S^q$ and $S^o$ and a ground distance function $d$, the Earth Mover’s Distance $EMD_d$ between $S^q$ and $S^o$ is defined as a minimum cost flow over all possible flows $f_{ij} \in \mathcal{R}$ as:

$$EMD_d(S^q, S^o) = \min \left\{ \sum_i \sum_j f_{ij} \cdot d(c_i^q, c_o^j) \right\} \left\{ \min\left\{ \sum_i w_i^q, \sum_j w_o^j \right\} \right\},$$

under the constraints: $\forall i : \sum_j f_{ij} \leq w_i^q$, $\forall j : \sum_i f_{ij} \leq w_o^j$, $\forall i, j : f_{ij} \geq 0$, and $\sum_i \sum_j f_{ij} = \min\left\{ \sum_i w_i^q, \sum_j w_o^j \right\}$.

The constraints guarantee a feasible solution, i.e. all costs are positive and do not exceed the limitations given by the weights in both feature signatures. However, as there is a minimization problem to solve, the run time complexity is considerably high.

The succeeding Weighted Correlation Distance [15] follows a slightly different approach. Instead of taking only distances between the feature signatures’ centroids into the computation, it measures the intersection among the clusters represented by the corresponding centroids via their distance $d$ and maximum cluster radius $R$ which has to be specified in the feature extraction process, as can be seen in Section 2.

**Definition 5 Weighted Correlation Distance**

Given two feature signatures $S^q$ and $S^o$, a ground distance function $d$, and the maximum cluster radius $R$, the Weighted Correlation Distance $WCD_{d,R}$ between $S^q$ and $S^o$ is defined as:

$$WCD_{d,R}(S^q, S^o) = 1 - \sum_i \sum_j s(c_i^q, c_o^j) \cdot \frac{w_i^q}{\sqrt{w_i^q + w_o^j}} - \frac{w_o^j}{\sqrt{w_i^q + w_o^j}},$$

where

$$S^q \cdot S^o = \sum_i \sum_j s(c_i^q, c_o^j) \cdot w_i^q \cdot w_o^j$$

and

$$s(c_i, c_j) = \begin{cases} \frac{1}{1 + \frac{d}{R}} & \text{if } 0 \leq \frac{d}{R} \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

Based on the intersection $s(c_i, c_j)$ between two centroids $c_i$ and $c_j$, the weighted correlation $S^q \cdot S^o$ between both feature signatures is normalized and used to determine the corresponding distance value.

The last similarity measure we present is the Signature Quadratic Form Distance [20, 21, 22] which bridges the gap between the traditional Quadratic Form Distance [23] and feature signatures. Its formal definition is given below.

**Definition 6 Signature Quadratic Form Distance**

Given two feature signatures $S^q$ and $S^o$, the Signature Quadratic Form Distance $SQFD_A$ between $S^q$ and $S^o$ is defined as:

$$SQFD_A(S^q, S^o) = \sqrt{(w_q - w_o) \cdot A \cdot (w_q - w_o)^T},$$

where $A \in \mathcal{R}^{(n+m) \times (n+m)}$ is the similarity matrix, $w_q = (w_1^q, \ldots, w_n^q)$ and $w_o = (w_1^q, \ldots, w_m^o)$ form weight vectors, and $(w_q - w_o) = (w_1^q, \ldots, w_n^q, -w_1^o, \ldots, -w_m^o)$ denotes the concatenation of $w_q$ and $-w_o$.

The similarity matrix $A$ which is dynamically determined for each comparison of two feature signatures reflects cross-similarities among the feature signatures’ centroids. Entries of $A$ depend on the order of the centroids in which they appear in the feature signatures and can be obtained via similarity functions [20, 21], for instance

$$a_{ij} = e^{-\alpha L^2(c_i,c_j)}.$$
complete weight and position information given by the feature signatures which decreases their efficiency.

We evaluate the retrieval performance of the similarity measures in the forthcoming section.

4. EXPERIMENTAL EVALUATION

In this section, we conduct experiments on different multimedia databases. First, we describe the experimental setup. Second, we evaluate the retrieval performance of the similarity measures in terms of effectiveness and efficiency.

4.1. Experimental Setup

In order to thoroughly evaluate the presented similarity measures, we measure their retrieval performance on the following databases: the Wang database [24], the Coil100 database [25], the MIR Flickr database [26], and the 101objects database [27]. We depict example images from these databases in Figure 3.

The Wang database comprises 1,000 images which are classified into ten themes. The themes cover a multitude of topics, such as beaches, flowers, buses, food, etc. The Coil100 database consists of 7,200 images classified into 100 different classes. Each class depicts one object photographed from 72 different directions. The MIR Flickr database contains 25,000 images downloaded from http://flickr.com/ including textual annotations. The 101objects database contains 9,196 images which are classified into 101 categories.

The themes, classes, textual annotations, and categories are used as ground truth to measure precision-recall values [28] after each returned object. In the MIR Flickr database we define virtual classes which contain all images sharing at least two common textual annotations and are used as ground truth.

We extracted feature signatures with respect to images in the aforementioned databases by applying an adaptive variant of the $k$-means clustering algorithm, described in Section 2, in different feature spaces comprising up to seven dimensions: three color ($L, a, b$), two position ($x, y$), and two texture dimensions ($\chi, \eta$) consisting of contrast $\chi$ and coarseness $\eta$. In this way, image pixels are first mapped to features in the corresponding feature space and then clustered accordingly. As color of an image is its fundamental perceptual property, we evaluated the performance of the similarity measures on the following feature spaces: color only ($L, a, b$), color + texture ($L, a, b, \chi, \eta$), color + position ($L, a, b, x, y$), and color + position + texture ($L, a, b, x, y, \chi, \eta$).

We fixed the parameters of the clustering algorithm to $R = 30$, $S = 50$, and $M = 40$. Furthermore, we excluded the black surroundings of objects in images of the Coil100 database by filtering out features which are almost black.

All similarity measures were evaluated by using the Euclidean ground distance function as it turned out that different ground distance functions did not influence the performance ratios of the similarity measures. For the Signature Quadratic Form Distance, we figured out that an exponential similarity function $a_{ij} = e^{-\alpha(L_{i,j}^2(a_e))}$ exhibits the best results and, therefore, we used this similarity function throughout our experimental evaluation. We dynamically determine the best value of $\alpha$ with respect to the corresponding database.

We ran all experiments on a 2.33GHz Intel machine and measured the performance of the presented similarity measures based on a JAVA implementation.

4.2. Retrieval Results

Tables 2 - 5 show the detailed mean average precision values for each combination of database, feature space, and similarity measure. Each value is gathered by receiving the mean of average precision values of 2,000 randomized queries which vary in their size and structure, i.e. the databases are queried with different images and each query image is also represented by feature signatures of different sizes. We highlight the highest mean average precision values of each row.

Comparing the retrieval results of the similarity measures, we observe that the mean average precision values of each similarity measure depend on the underlying feature space and database. For the Wang, the Coil100, and the MIR Flickr databases the Signature Quadratic Form Distance (SQFD) exhibits the highest mean average precision values. When only color features are used, the Perceptually Modified Hausdorff Distance (PMHD) and the Earth Mover’s Distance (EMD) show the best retrieval performance in terms of mean average precision values. The latter exhibits the highest mean average precision values for the 101objects database, regardless of the selected features. While the Hausdorff Distance (HD) achieves

---

**Table 1.** The similarity measures’ properties.

<table>
<thead>
<tr>
<th>Similarity Measure</th>
<th>Efficiency</th>
<th>Weight Information</th>
<th>Position Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD</td>
<td>++++</td>
<td>-</td>
<td>partially</td>
</tr>
<tr>
<td>PMHD</td>
<td>+++</td>
<td>partially</td>
<td>completely</td>
</tr>
<tr>
<td>EMD</td>
<td>+</td>
<td>completely</td>
<td>completely</td>
</tr>
<tr>
<td>WCD</td>
<td>++</td>
<td>completely</td>
<td>completely</td>
</tr>
<tr>
<td>SQFD</td>
<td>++</td>
<td>completely</td>
<td>completely</td>
</tr>
</tbody>
</table>

---

**Fig. 3.** Example images of the (a) Wang, (b) Coil100, (c) MIR Flickr, and (d) 101objects database.
Table 2. Mean average precision values for the Wang database.

<table>
<thead>
<tr>
<th>features</th>
<th>SQFD</th>
<th>HD</th>
<th>PMHD</th>
<th>WCD</th>
<th>EMFD</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>color</td>
<td>0.586</td>
<td>0.317</td>
<td>0.439</td>
<td>0.567</td>
<td>0.580</td>
<td>2.9</td>
</tr>
<tr>
<td>color + texture</td>
<td>0.620</td>
<td>0.345</td>
<td>0.476</td>
<td>0.604</td>
<td>0.599</td>
<td>1.5</td>
</tr>
<tr>
<td>color + position</td>
<td>0.568</td>
<td>0.391</td>
<td>0.461</td>
<td>0.536</td>
<td>0.563</td>
<td>1.2</td>
</tr>
<tr>
<td>color + pos. + text.</td>
<td>0.613</td>
<td>0.308</td>
<td>0.476</td>
<td>0.591</td>
<td>0.598</td>
<td>0.9</td>
</tr>
<tr>
<td>mean:</td>
<td>0.597</td>
<td>0.340</td>
<td>0.463</td>
<td>0.575</td>
<td>0.585</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Mean average precision values for the Coil100 database.

<table>
<thead>
<tr>
<th>features</th>
<th>SQFD</th>
<th>HD</th>
<th>PMHD</th>
<th>WCD</th>
<th>EMFD</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>color</td>
<td>0.811</td>
<td>0.731</td>
<td>0.828</td>
<td>0.772</td>
<td>0.809</td>
<td>2.9</td>
</tr>
<tr>
<td>color + texture</td>
<td>0.770</td>
<td>0.477</td>
<td>0.696</td>
<td>0.728</td>
<td>0.750</td>
<td>1.5</td>
</tr>
<tr>
<td>color + position</td>
<td>0.802</td>
<td>0.498</td>
<td>0.664</td>
<td>0.743</td>
<td>0.706</td>
<td>0.7</td>
</tr>
<tr>
<td>color + pos. + text.</td>
<td>0.776</td>
<td>0.425</td>
<td>0.606</td>
<td>0.726</td>
<td>0.710</td>
<td>0.6</td>
</tr>
<tr>
<td>mean:</td>
<td>0.790</td>
<td>0.533</td>
<td>0.699</td>
<td>0.742</td>
<td>0.744</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Mean average precision values for the MIR Flickr database.

<table>
<thead>
<tr>
<th>features</th>
<th>SQFD</th>
<th>HD</th>
<th>PMHD</th>
<th>WCD</th>
<th>EMFD</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>color</td>
<td>0.322</td>
<td>0.318</td>
<td>0.322</td>
<td>0.321</td>
<td>0.322</td>
<td>0.7</td>
</tr>
<tr>
<td>color + texture</td>
<td>0.343</td>
<td>0.317</td>
<td>0.331</td>
<td>0.338</td>
<td>0.337</td>
<td>1.0</td>
</tr>
<tr>
<td>color + position</td>
<td>0.321</td>
<td>0.316</td>
<td>0.319</td>
<td>0.317</td>
<td>0.314</td>
<td>0.2</td>
</tr>
<tr>
<td>color + pos. + text.</td>
<td>0.343</td>
<td>0.307</td>
<td>0.322</td>
<td>0.335</td>
<td>0.333</td>
<td>0.6</td>
</tr>
<tr>
<td>mean:</td>
<td>0.332</td>
<td>0.315</td>
<td>0.324</td>
<td>0.328</td>
<td>0.327</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Mean average precision values for the 101objects database.

<table>
<thead>
<tr>
<th>features</th>
<th>SQFD</th>
<th>HD</th>
<th>PMHD</th>
<th>WCD</th>
<th>EMFD</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>color</td>
<td>0.106</td>
<td>0.060</td>
<td>0.076</td>
<td>0.097</td>
<td>0.109</td>
<td>0.4</td>
</tr>
<tr>
<td>color + texture</td>
<td>0.113</td>
<td>0.064</td>
<td>0.110</td>
<td>0.109</td>
<td>0.118</td>
<td>0.7</td>
</tr>
<tr>
<td>color + position</td>
<td>0.113</td>
<td>0.081</td>
<td>0.120</td>
<td>0.097</td>
<td>0.130</td>
<td>1.6</td>
</tr>
<tr>
<td>color + pos. + text.</td>
<td>0.139</td>
<td>0.072</td>
<td>0.105</td>
<td>0.117</td>
<td>0.141</td>
<td>1.6</td>
</tr>
<tr>
<td>mean:</td>
<td>0.118</td>
<td>0.069</td>
<td>0.103</td>
<td>0.105</td>
<td>0.125</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Computation time values in milliseconds for the MIR Flickr database.

<table>
<thead>
<tr>
<th>features</th>
<th>SQFD</th>
<th>HD</th>
<th>PMHD</th>
<th>WCD</th>
<th>EMFD</th>
</tr>
</thead>
<tbody>
<tr>
<td>color</td>
<td>2.374</td>
<td>346</td>
<td>502</td>
<td>1.253</td>
<td>24.281</td>
</tr>
<tr>
<td>color + texture</td>
<td>10.409</td>
<td>1.429</td>
<td>2.201</td>
<td>5.333</td>
<td>58.706</td>
</tr>
<tr>
<td>color + position</td>
<td>6.561</td>
<td>882</td>
<td>1.349</td>
<td>3.283</td>
<td>35.369</td>
</tr>
<tr>
<td>color + pos. + text.</td>
<td>20.666</td>
<td>2.657</td>
<td>4.168</td>
<td>10.526</td>
<td>110.527</td>
</tr>
<tr>
<td>mean:</td>
<td>10.003</td>
<td>1.329</td>
<td>2.055</td>
<td>5.099</td>
<td>57.221</td>
</tr>
</tbody>
</table>

5. DISCUSSION

According to Table 1 in Section 3, we can classify the similarity measures into two groups. The first group examined includes the Hausdorff Distance and Perceptually Modified Hausdorff Distance. While the Hausdorff Distance completely ignores weight information of the feature signatures, the latter includes this information. Their computations gather position information of the feature signatures only partially, thus they are very fast to compute. In our detailed experimental evaluation, we figured out that the Hausdorff Distances achieve the best results when the size of the query feature signatures is approximately the same as that of the database. As mentioned in the previous section, we vary the query feature signature size to obtain realistic real-world experimental evaluation.

The second group which handles this discrepancy between query and database more successfully includes the Weighted Correlation Distance, Earth Mover’s Distance, and Signature Quadratic Form Distance. They completely take into account weight and position information of the feature signatures to be compared. This comes at the cost of higher computation time values, which particularly hold for the Earth Mover’s Distance solving a costly optimization problem. We figured out that the similarity measures of the second group achieve good results even if the size of the query feature signature is much smaller than that of any feature signature corresponding a data object in the database. The computation of the Weighted Correlation Distance is limited by the parameter \( R \) which needs to be specified in advance according to individual clusters. The Signature Quadratic Form Distance ensures to use the complete position and weight information in the computation. It achieves the highest retrieval performance in terms of effectiveness.

6. CONCLUSIONS

In this paper, we presented new insights into the state-of-the-art similarity measures applicable to feature signatures for content-based multimedia retrieval. We compared the Hausdorff Distance, Perceptually Modified Hausdorff Distance, Weighted Correlation Distance, Earth Mover’s Distance, and Signature Quadratic Form Distance, and evaluated experimental results with respect to their qualities of effectiveness, as well as efficiency. We observed that the Signature Quadratic Form Distance comes up with the highest retrieval performance in terms of effectiveness and outperforms the aforementioned similarity measures.
7. REFERENCES


