Signature Quadratic Form Distance*

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

ABSTRACT

The *Signature Quadratic Form Distance* is an adaptive similarity measure for flexible content-based feature representations of multimedia data. In this paper, we present a deep survey of the mathematical foundation of this similarity measure which encompasses the classic Quadratic Form Distance defined only for the comparison between two feature histograms of the same length and structure. Moreover, we give the benefits of the Signature Quadratic Form Distance and experimental evaluation on numerous real-world databases.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval Models

General Terms

Theory, Experimentation, Performance

Keywords

Signature Quadratic Form Distance, Content-Based Similarity, Feature Signatures, Multimedia Information Retrieval

1. INTRODUCTION

Multimedia Information Retrieval is a cross-cutting field involving many different research areas which mainly focus on the retrieval of complex multimedia data in large databases. Generally, the challenging task is to find multimedia objects which are somehow related to a specified query. Whereas this relationship between query and data is typically determined via content-based similarity measures, multimedia retrieval systems are not necessarily limited to processing of one specific medium. For instance, the query can be an abstract formalization or sketch of users' thoughts put into the retrieval system which is then supposed to answer the query by retrieving the most similar multimedia objects the database contains.

One promising approach to search for similar multimedia data is to automatically extract inherent properties of the objects and compare them with each other. For this purpose, object features are mapped and aggregated in some feature space which enables the system to compare them via a distance function effectively and efficiently.

Determining the similarity between two multimedia objects via a distance computation between their feature representations is an important task of nearly all content-based retrieval systems [14]. To this end, numerous similarity models have been proposed which aim at representing object contents in an appropriate way and defining a distance function which can efficiently compute distance between these feature representations. As it turns out that the combination of flexible feature representations and adaptive distance functions achieves a high retrieval quality [12, 13, 18, 20], the *Signature Quadratic Form Distance* [1, 2] was introduced as an adaptive similarity measure for flexible feature representations. This similarity measure bridges the gap between the Quadratic Form Distance [7, 21] and feature signatures which represent more flexible structures than traditional feature histograms [19]. In this work, we make the following contributions:

- We give a deeper mathematical insight and prove that the similarity measure encompasses the classic Quadratic Form Distance.
- We thoroughly evaluate the performance of our presented similarity measure on different databases and show its effectiveness and efficiency.

We structure the rest of this paper as follows: in Section 2, we briefly review existing similarity models. In Section 3, we present the Signature Quadratic Form Distance in-depth which we evaluate in Section 4. We conclude our paper with an outlook on future work in Section 5.

2. RELATED WORK AND BACKGROUND

In this section, we first present two commonly used feature representations of multimedia data, namely feature histograms and feature signatures, which are also known as fixed-binning and adaptive-binning feature histograms, and then summarize applicable similarity measures for both representations.

*This work is partially funded by the Excellence Initiative of the German federal and state governments.

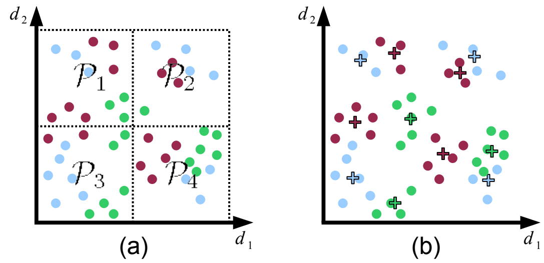


Figure 1: Fixed (a) and adaptive (b) partitioning.

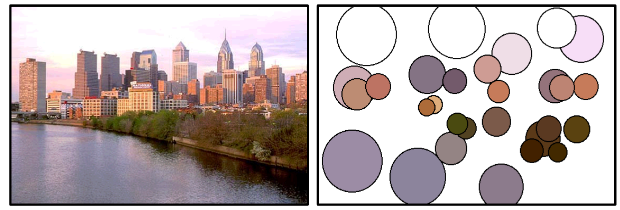


Figure 2: An example image and the visualization of its corresponding feature signature.

Feature representations are responsible for digitizing multimedia objects in a compact way by capturing their inherent properties. In this way, the object properties are represented as a set of features in an appropriate feature space $\mathcal{FS}$. For instance, for content-based image retrieval [3, 22], the feature space comprises dimensions such as position, color, or texture information [4] and each image's pixel is mapped to one feature in this feature space. The distribution of these features is then aggregated via a partitioning in the feature space. Depending on the partitioning which can be fixed for the whole database or adaptive for each data object, we can use *feature histograms* or *feature signatures*. First, we give the definition of feature histograms as below.

Definition 1. Feature Histogram
Given a feature space $\mathcal{FS}$ with a disjoint partitioning $\mathcal{P} = \mathcal{P}_1 \cup \ldots \cup \mathcal{P}_n$, the feature histogram h^o of object o with the features $f_1, \ldots, f_k \in \mathcal{FS}$ is defined as an n-dimensional vector $h^o = (h^o_1, \ldots, h^o_n)$ with $h^o_i = \frac{|\mathcal{P}_i|}{k}$ for $1 \leq i \leq n$, where $|\mathcal{P}_i|$ denotes the density of object o in partition $\mathcal{P}_i$.

According to Definition 1, histogram entries h^o_i count the proportion of features in the corresponding partition $\mathcal{P}_i$ and approximate the feature distribution in the feature space. In this way, feature histograms store the information about the object contents in a simple vectorial form based on a fixed partitioning of the feature space. We illustrate the concept of a fixed partitioning in Figure 1 (a) where we depict the features f_i of three objects as red, green, and blue points in a two-dimensional feature space $d_1 \times d_2$ and determine a grid-based partitioning with four partitions. As the partitions are fixed over all data objects, feature histograms are limited in the way they reflect the feature distribution not only within one partition, but also scattered over multiple partitions. To overcome these shortcomings, *feature signatures* [5, 20], which are also known as adaptive-binning feature histograms in the literature [13], group each object's features individually via a local partitioning. As this local partitioning is frequently obtained via clustering algorithms [6], we define feature signatures based on a local clustering as follows.

Definition 2. Feature Signature
Given a feature space $\mathcal{FS}$ and a local clustering $\mathcal{C} = \mathcal{C}_1, \ldots, \mathcal{C}_n$ 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 $\mathcal{FS} \times \mathcal{R}^+$ as follows:

$$S^o = \{\langle c^o_i, w^o_i \rangle, i = 1, \ldots, n\},$$

where $c^o_i = \frac{\sum_{f \in \mathcal{C}_i} f}{|\mathcal{C}_i|}$ and $w^o_i = \frac{|\mathcal{C}_i|}{k}$ represent the centroid and weight, respectively.

For each data object o, there exists a single feature signature S^o. Each feature signature S^o stores centroids $c^o_i \in \mathcal{FS}$ and weights $w^o_i \in \mathcal{R}^+$ of the clusters $\mathcal{C}_i$. As this is carried out individually for each data object, the resulting feature signature based on a local clustering reflects the feature distributions more meaningfully than those relying on a fixed partitioning. We illustrate this advantage in Figure 1 (b) where centroids of the clusters are depicted as pluses in the corresponding colors. It can be directly recognized that the feature distributions of data objects are more precisely reflected via feature signatures and the size can vary from object to object.

In Figure 2, we depict an example image and the visualization of its corresponding feature signature. We visualize the feature signature centroids from a five-dimensional feature space (two position and three color dimensions) as circles in a two-dimensional position space. The circles' colors and diameters reflect the colors and weights of the centroids, respectively.

Unlike feature signatures, feature histograms cannot achieve a good balance between expressiveness and efficiency [19]. Depending on specific domains, features and, thus, feature spaces, the type of partitioning or clustering, the dimensionality, and also terms used – feature vectors instead of feature histograms or adaptive-binning feature histograms instead of feature signatures – can vary.

To determine similarity between two multimedia objects based on their feature representations, multimedia retrieval systems frequently apply similarity measures. The higher the distance between two feature representations, the lower the similarity of the corresponding objects and vice versa. For feature histograms, there exists a multitude of similarity measures [10], and probably the most prominent ones are L_p norm-based ones which are defined in their weighted version as $L_p(h^q, h^o) = \left(\sum_i w_i \cdot |h^q_i - h^o_i|^p\right)^{\frac{1}{p}}$, where $w_i \in \mathcal{R}^+$ represents the weight for the dimension i. This definition includes the *Manhattan*, *Euclidean*, and *Maximum Distance* which are restricted to comparison of only the same feature histogram dimensions with each other. A more flexible and adaptive similarity measure that takes into account the cross-dimension dependencies is the *Quadratic Form Distance* [7, 21] which is defined with a similarity matrix $A \in \mathcal{R}^{n \times n}$ between two feature histograms $h^q, h^o \in \mathcal{R}^n$ as

$$QFD_A(h^q, h^o) = \sqrt{(h^q - h^o) \cdot A \cdot (h^q - h^o)^T}.$$

Each entry a_{ij} in the similarity matrix A models the similarity between dimension i and j of the feature histograms. The advantage of the Quadratic Form Distance is different dimensions of the feature histograms can be compared with each other and, as a result, it has the ability to model

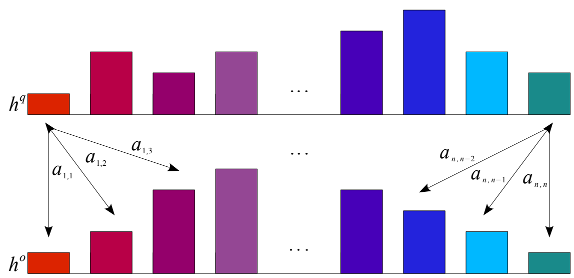


Figure 3: The illustration of the Quadratic Form Distance between feature histograms h^q and h^o.

user preferences more precisely. We illustrate the cross-dimension concept of the Quadratic Form Distance in Figure 3 where we compare two color feature histograms h^q and h^o with each other. The figure shows how the entries of both feature histograms are compared with each other in order to compute the distance value. The similarity matrix has the ability to model similarities among identical dimensions, i.e. a_{ii}, with higher similarity values than similarities among different dimensions, i.e. a_{ij} with $i \neq j$.

So far, we have presented the similarity measures applicable to feature histograms having the same length and sharing a global, fixed partitioning of the feature space. Next, we will review similarity measures which cope with feature signatures of arbitrary structure and size. For this purpose, they apply a so-called *ground distance function* to determine distances among feature signatures' centroids in the feature space. Applicable ground distance functions can be found in [10].

The first similarity measure we present is the *Hausdorff Distance* [12] which measures the maximum nearest neighbor distance among centroids in both feature signatures. The formal definition of the Hausdorff Distance is given below.

Definition 3. 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:

$$\begin{aligned} HD_d(S^q, S^o) &= \max\{h(S^q, S^o), h(S^o, S^q)\}, \text{ where} \\ h(S^q, S^o) &= \max_{c^q \in S^q} \min_{c^o \in S^o} \{d(c^q, c^o)\}. \end{aligned}$$

As an extension for color-based image retrieval, the *Perceptually Modified Hausdorff Distance* [18] 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 4. 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:

$$\begin{aligned} PMHD_d(S^q, S^o) &= \max\{h_w(S^q, S^o), h_w(S^o, S^q)\}, \\ \text{where } h_w(S^q, S^o) &= \frac{\sum_i w_i^q \cdot \min_j \{\frac{d(c_i^q, c_j^o)}{\min\{w_i^q, w_j^o\}}\}}{\sum_i w_i^q}. \end{aligned}$$

The computation of the distance between the feature signatures S^q and S^o via the Perceptually Modified Hausdorff Distance requires the determination of the centroid c_j^o 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.

The first similarity measure we present is the well-known *Earth Mover's Distance* [20] 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 [9].

Definition 5. 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:

$$EMD_d(S^q, S^o) = \min_{f_{ij}} \left\{ \frac{\sum_i \sum_j f_{ij} \cdot d(c_i^q, c_j^o)}{\min\{\sum_i w_i^q, \sum_j w_j^o\}} \right\},$$

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

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 be solved, the run time complexity is considerably high.

The succeeding *Weighted Correlation Distance* [13] 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 value d and maximum cluster radius R which has to be specified in advance in the feature signature extraction process [13].

Definition 6. 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:

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

where

$$\begin{aligned} S^q \cdot S^o &= \sum_i \sum_j s(c_i^q, c_j^o) \cdot w_i^q \cdot w_j^o, \text{ and} \\ s(c_i, c_j) &= \begin{cases} 1 - \frac{3}{4}\frac{d}{R} + \frac{1}{16}(\frac{d}{R})^3 & \text{if } 0 \leq \frac{d}{R} \leq 2, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

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 Hausdorff Distance, Perceptually Modified Hausdorff Distance, Earth Mover's Distance, and Weighted Correlation Distance are able to compare feature signatures of different structure and size which are the preferred feature representation form of multimedia objects. As feature histograms exhibit a special case of feature signatures, the application of these similarity measures is not restricted to the latter. Despite the common thought that Quadratic Form Distances are only applicable to feature histograms, we will propose how to develop its cross-dimension concept to flexible feature signatures in the following section.

3. THE SIGNATURE QUADRATIC FORM DISTANCE

In this section, we first recall the feature signature representation in the underlying feature space $\mathcal{FS}$. Then, we give a solid mathematical definition to compute this similarity measure and show that the resulting Signature Quadratic Form Distance encompasses the classic Quadratic Form Distance which solely compares feature histograms of the same length and structure. As the Signature Quadratic Form Distance makes use of a similarity function, we finally devote the remainder of this section to appropriate choices of these functions.

3.1 Similarity Model

We generally determine similarity between two multimedia objects by computing an appropriate distance between their feature representations. For the Signature Quadratic Form Distance, we aim at computing distances among feature signatures which vary in structure and size.

Figure 4 shows an example where we depict two different structured feature signatures $S^q = \{\langle c_i^q, w_i^q\rangle, i = 1, \ldots, 3\}$ and $S^o = \{\langle c_i^o, w_i^o\rangle, i = 1, \ldots, 4\}$ with green and red pluses in a two-dimensional feature space $\mathcal{FS}$, respectively. The centroids $c_i \in \mathcal{FS}$ and corresponding weights $w_i \in \mathcal{R}^+$ of both feature signatures summarize features and, thus, reflect object contents in a compressed and flexible way. In order to compare both feature signatures with each other, we apply the cross-dimension concept of the Quadratic Form Distance to them and compare the centroids' positions and weights with each other. This is done by adopting the comparison among all dimensions of the feature histograms to the comparison among all centroids of the feature signatures. In other words, the cross-dimension concept of the Quadratic Form Distance on feature histograms is adopted to the Signature Quadratic Form Distance on feature signatures where we name this concept the cross-dependency concept.

To perform this cross-dependency concept which compares all centroids of the feature signatures with each other without knowing the similarities among the centroids in advance, the distance computation is distinguished in three different parts which we highlight in Figure 4. On the left-hand side, we visualize intra-dependencies, that is, similarity computations within each feature signature via green and red lines, whereas the inter-dependencies, similarity computations between both feature signatures, are shown with gray lines on the right-hand side. Intuitively, the distance computation involves the measurement of feature signatures' similarity (green and red lines) which is then compared with the measurement of similarity between them (gray lines). In this way, we consider the inherent structure of both feature signatures which are then compared with each other.

Mathematically, the similarity measure is calculated by multiplying the similarity of two centroids with their according weights and finally summarizing over all possibilities. We show that the described similarity measure indeed matches the cross-dependency concept and encompasses the classic Quadratic Form Distance in the following.

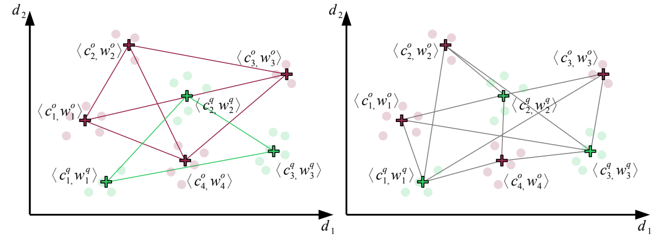


Figure 4: Intra-dependencies and inter-dependencies between two feature signatures S^q and S^o, depicted in a two dimensional feature Space $\mathcal{FS}$.

3.2 Mathematical Definition

Instead of taking the distances between the feature signatures' centroids into the computation, the Signature Quadratic Form Distance applies similarity functions among the centroids. The values of these similarity functions vary inversely with those of ground distance functions. We refer to Section 3.3 for the concept of similarity functions and continue with the definition of the Signature Quadratic Form Distance.

Definition 7. Signature Quadratic Form Distance
Given two feature signatures $S^q = \{\langle c_i^q, w_i^q\rangle | \; i = 1, \ldots, n\}$ and $S^o = \{\langle c_i^o, w_i^o\rangle | \; i = 1, \ldots, m\}$, and a similarity function $f_s(c_i, c_j) \mapsto \mathcal{R}$, the Signature Quadratic Form Distance $SQFD_{f_s}$ between S^q and S^o is defined as:

$$SQFD_{f_s}(S^q, S^o) = \sqrt{(w_q| - w_o) \cdot A_{f_s} \cdot (w_q| - w_o)^T},$$

where $A_{f_s} \in \mathcal{R}^{(n+m)\times(n+m)}$ is the similarity matrix arising from applying the similarity function f_s to the corresponding centroids, i.e. $a_{ij} = f_s(c_i, c_j)$. Furthermore, $w_q = (w_1^q, \ldots, w_n^q)$ and $w_o = (w_1^o, \ldots, w_m^o)$ 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 $-w_o$.

In order to compute the Signature Quadratic Form Distance between two feature signatures, we concatenate both weight vectors w_q and w_o which comprise the weights of both feature signatures, and multiply them with the similarity matrix A_{f_s} which reflects similarities among the feature signatures' centroids. The similarity matrix A_{f_s} is dynamically determined for each comparison of two feature signatures and models the intra-dependencies and inter-dependencies as described above. Intuitively, the concatenation $(w_q|-w_o)$ of the weight vectors reflects the weight information of both feature signatures' centroids, while the positional information regarding feature signatures' centroids is exhibited via the similarity matrix A_{f_s}. Furthermore, the entries of the similarity matrix depend on the order of the centroids in which they appear in the feature signatures. To determine these entries, we present the concept of similarity functions in Section 3.3 where we study the special case of mapping distances between centroids to similarity values.

We elucidate the idea and the computation of the Signature Quadratic Form Distance in the example below where we compare two feature signatures with each other.

Example 1. Distance Computation

Suppose we are given a two-dimensional feature space $d_1 \times d_2$ and two feature signatures $S^q = \{\langle\binom{3}{3}, 0.5\rangle, \langle\binom{8}{7}, 0.5\rangle\}$ and $S^o = \{\langle\binom{4}{7}, 0.5\rangle, \langle\binom{9}{5}, 0.25\rangle, \langle\binom{8}{1}, 0.25\rangle\}$ which are depicted in Figure 5 (a). According to Definition 7, we obtain the corresponding weight vectors $w_q = (0.5, 0.5)$ and $w_o = (0.5, 0.25, 0.25)$ of the feature signature S^q and S^o, respectively. The concatenation of both weight vectors is given as below:

$$\begin{array}{cccccccc} & & \binom{3}{3} & \binom{8}{7} & \binom{4}{7} & \binom{9}{5} & \binom{8}{1} & \\ (w_q| - w_o) = (& & 0.5, & 0.5, & -0.5, & -0.25, & -0.25 &). \end{array}$$

We depict the feature signatures' centroids above the concatenated weight vector, as their order implies the similarity matrix $A_{f_s} \in \mathcal{R}^{5\times 5}$ which is given as follows:

$$A_{f_s} = \begin{pmatrix} 1 & 0.135 & 0.195 & 0.137 & 0.157 \\ 0.135 & 1 & 0.2 & 0.309 & 0.143 \\ 0.195 & 0.2 & 1 & 0.157 & 0.122 \\ 0.137 & 0.309 & 0.157 & 1 & 0.195 \\ 0.157 & 0.143 & 0.122 & 0.195 & 1 \end{pmatrix}$$

The computation of a single entry $a_{ij} \in A_{f_s}$ is performed by measuring the similarity between the centroids c_i and c_j with the similarity function $f_s(c_i, c_j) = \frac{1}{1+L_2(c_i,c_j)}$ which is based on the Euclidean ground distance function L_2. The entry $a_{24} = 0.309$ models the similarity between centroid $c_2^q = \binom{8}{7}$ of feature signature S^q and centroid $c_2^o = \binom{9}{5}$ of feature signature S^o. As both centroids are the closest ones to each other, this entry reflects the highest similarity value in matrix A_{f_s} except the values on its diagonal. Finally, we compute the Signature Quadratic Form Distance as in the following:

$$\begin{aligned} SQFD_{f_s}(S^q, S^o) &= \sqrt{(w_q| - w_o) \cdot A_{f_s} \cdot (w_q| - w_o)^T} \\ &= 0.808\ . \end{aligned}$$

Figure 5 (a) shows the corresponding feature signatures with the individual centroids and weights taken from the example above. Each plus denotes a centroid whose weight is given by the proportion of features, shown as light-green and light-red points, which belong to that plus. The green centroids belong to the feature signature S^q, and the red ones to the feature signature S^o.

When we examine the structure of the similarity matrix depicted in Figure 5 (b), we recognize the similarity matrix block structure comprising four blocks. Each of these blocks reflects the cross-dependencies among the corresponding centroids of the feature signatures as shown in Figure 4: the green and red blocks model the intra-dependency among the centroids within feature signature S^q and S^o, respectively. The remaining gray blocks are generated by comparing the centroids of both S^q and S^o with each other and model the inter-dependency among their centroids.

Before we show how the Signature Quadratic Form Distance encompasses the classic Quadratic Form Distance in Theorem 1, we briefly recall the Quadratic Form Distance on feature histograms.

In the feature histogram case, each entry of the Quadratic Form Distance's similarity matrix models similarity between different dimensions of the feature histograms. Whether this similarity matrix is statistically specified by considering only the feature histograms entries, for instance by using the inverse covariance matrix, or manually designed on

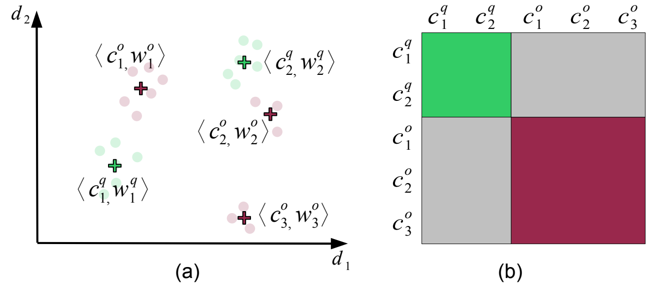


Figure 5: (a) Two feature signatures with their centroids and weights. (b) The structure of the similarity matrix.

the basis of the feature histograms' underlying global partitions, as defined in Section 2, the similarity matrix can be expressed as a similarity function among the feature histograms' dimensions. By turning the similarity matrix A of the Quadratic Form Distance QFD_A into the corresponding similarity function f_s of the Signature Quadratic Form Distance $SQFD_{f_s}$, we can show their equivalence in case of comparing feature histograms with each other as shown in the following theorem.

Theorem 1. The Feature Histogram Case

Given the Quadratic Form Distance QFD_A and the Signature Quadratic Form Distance $SQFD_{f_s}$ with the similarity function f_s which corresponds to the specified similarity entries in the matrix $A \in \mathcal{R}^{n\times n}$ modeling the similarities among the dimensions of two feature histograms $h^q = (h_1^q, \ldots, h_n^q)$ and $h^o = (h_1^o, \ldots, h_n^o)$, the following holds:

$$QFD_A(h^q, h^o) = SQFD_{f_s}(h^q, h^o).$$

PROOF.

$$\begin{aligned} & QFD_A^2(h^q, h^o) \\ = \ & (h^q - h^o) \cdot A \cdot (h^q - h^o)^T \\ = \ & h^q \cdot A \cdot h^{qT} + h^o \cdot A \cdot h^{oT} - h^q \cdot A \cdot h^{oT} - h^o \cdot A \cdot h^{qT} \\ = \ & (h^q | - h^o) \cdot \begin{pmatrix} A & A \\ A & A \end{pmatrix} \cdot (h^q| - h^o)^T \\ = \ & (h^q| - h^o) \cdot A_{f_s} \cdot (h^q| - h^o)^T \\ = \ & SQFD_{f_s}^2(h^q, h^o). \end{aligned}$$

Consequently, we obtain that $QFD_A(h^q, h^o)$ is the same as $SQFD_{f_s}(h^q, h^o)$ for any feature histograms h^q and h^o and the corresponding similarity function f_s derived from the matrix A. $\square$

Theorem 1 shows that the Signature Quadratic Form Distance encompasses the classic Quadratic Form Distance for feature histograms by applying the corresponding similarity function among the dimensions of the feature histograms.

Since similarity functions play a major role in computing the Signature Quadratic Form Distance, we devote the next subsection to them where we demonstrate similarity functions by means of examples.

3.3 Similarity Functions

When we compute the Signature Quadratic Form Distance between two feature signatures, we apply a similarity function to generate the similarity matrix and gather similarities among all centroids of the objects' feature signatures. Comparing similarities among centroids, we assume that the same centroids exhibit the highest possible similarity value. The definition of similarity function which uses the preceding property is given in the following.

Definition 8. Similarity Function
A similarity function is a function $f_s(c_i, c_j) \mapsto \mathcal{R}$ for which the following holds:

$$f_s(c_i, c_i) \geq f_s(c_j, c_k) \text{ for all } c_i, c_j, c_k, \text{ and } c_j \neq c_k.$$

In Definition 8, we require similarity functions to assign the highest similarity values among identical centroids which corresponds to our intuition of similarity that nothing is more similar than the same.

One promising way to determine such similarity functions is based on *ground distance functions* – as introduced in Section 2, the term ground distances frequently refers to distances defined in the feature space and which are included in the computation of distances among feature signatures. The higher the value of the ground distance function, the lower the value of the chosen similarity function. We define three similarity functions in the following.

Definition 9. Three Similarity Functions
Given a ground distance function $d(c_i, c_j) \mapsto \mathcal{R}^+$ and a parameter $\alpha \in \mathcal{R}$, we define the three typical similarity functions:

- Minus function: $f_-(c_i, c_j) = -d(c_i, c_j)$
- Gaussian function: $f_g(c_i, c_j) = e^{-\alpha \cdot d^2(c_i, c_j)}$
- Heuristic function: $f_h(c_i, c_j) = \frac{1}{\alpha + d(c_i, c_j)}$

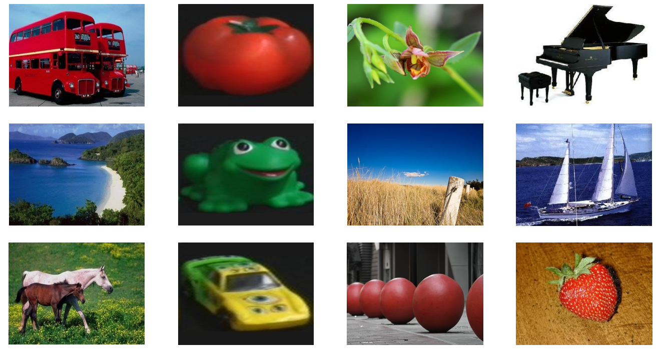
Obviously, the three defined functions in Definition 9 are similarity functions according to Definition 8, since they always reach the maximum value when both input parameters are identical. Whereas the minus function f_- is free of any additional parameter, both gaussian f_g and heuristic f_h function have an additional parameter $\alpha \in \mathcal{R}$ which has to be specified to the underlying database accordingly.

In the forthcoming section, we experimentally evaluate the Signature Quadratic Form Distance. For this purpose, we will determine the constant α which strongly depends on type of database and the used features based on the Euclidean ground distance function.

4. EXPERIMENTAL EVALUATION

In this section, we evaluate the retrieval performance of the Signature Quadratic Form Distance and the other presented state-of-the-art similarity measures in terms of effectiveness and efficiency on the following databases: the *Wang* database [23], the *Coil100* database [17], the *MIR Flickr* database [11], and the *101objects* database [8]. We depict example images from these databases in Figure 6.

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 after each returned object [15, 16]. 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.



(a) (b) (c) (d)

Figure 6: Example images of the (a) *Wang*, (b) *Coil100*, (c) *MIR Flickr*, and (d) *101objects* database.

We extracted feature signatures with respect to images in the aforementioned databases by applying an adaptive variant of the k-means clustering algorithm [13], in different feature spaces comprising up to seven dimensions: three color, two position, and two texture dimensions [4]. As color of an image is its fundamental perceptual property, we evaluated the performance of the similarity measures on the following feature spaces: only color (3d), color + texture (5d), color + position (5d), and color + position + texture (7d). We fixed the internal parameters of the adaptive k-means clustering algorithm to $R = 30$, $S = 50$, and $M = 40$, more details regarding the choice of these parameters can be found in [13]. Furthermore, we excluded the black surroundings of objects in images of the *Coil100* database by filtering out features which are almost black.

We ran all experiments on a 2.33GHz Intel machine with 16GB memory and measured the performance of the similarity measures based on a JAVA implementation.

Table 1 to Table 4 show the *mean average precision* values [15, 16] gathered by $1,000$ 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. The tables differ from features which are used, as described above. We highlight the highest mean average precision values of each row.

Regarding the experimental behavior, we can make the following observations: first, the Signature Quadratic Form Distance ($SQFD_{f_g}$) with the Gaussian similarity function f_g exhibits the highest effectiveness for the used feature spaces by setting the constant α accordingly. We depict the optimal values of α_{f_h} and α_{f_g} in the last two columns in the tables. Second, the well-known Earth Mover's Distance (EMD) is outperformed by the Signature Quadratic Form Distance regarding effectiveness except for the *101objects* database where the Earth Mover's Distance always achieves

Table 1: Mean average precision values by making use of position, color, and texture features.

database	$SQFD_{f_-}$	$SQFD_{f_h}$	$SQFD_{f_g}$	HD	PMHD	WCD	EMD	α_{f_h}	α_{f_g}
Wang	0.592	0.598	**0.613**	0.308	0.476	0.591	0.598	2.7	0.9
Coil100	0.726	0.721	**0.776**	0.425	0.606	0.726	0.710	2.7	0.6
MIR Flickr	0.336	0.338	**0.343**	0.307	0.322	0.335	0.333	2.5	0.6
101objects	0.117	0.128	0.139	0.072	0.105	0.117	**0.141**	1.4	1.6
average:	0.443	0.446	**0.468**	0.278	0.377	0.442	0.446		

Table 2: Mean average precision values by making use of position and color features.

database	$SQFD_{f_-}$	$SQFD_{f_h}$	$SQFD_{f_g}$	HD	PMHD	WCD	EMD	α_{f_h}	α_{f_g}
Wang	0.536	0.531	**0.568**	0.391	0.461	0.536	0.563	2.9	1.2
Coil100	0.743	0.720	**0.802**	0.498	0.664	0.743	0.706	2.9	0.7
MIR Flickr	0.317	0.315	**0.321**	0.316	0.319	0.317	0.314	2.9	0.2
101objects	0.097	0.115	0.113	0.081	0.120	0.097	**0.130**	<0.1	1.6
average:	0.423	0.420	**0.451**	0.322	0.391	0.423	0.428		

Table 3: Mean average precision values by making use of color and texture features.

database	$SQFD_{f_-}$	$SQFD_{f_h}$	$SQFD_{f_g}$	HD	PMHD	WCD	EMD	α_{f_h}	α_{f_g}
Wang	0.604	0.615	**0.620**	0.345	0.476	0.604	0.599	1.6	1.5
Coil100	0.728	0.721	**0.770**	0.477	0.696	0.728	0.750	2.8	1.5
MIR Flickr	0.338	0.341	**0.343**	0.317	0.331	0.338	0.337	1.6	1.0
101objects	0.108	0.115	0.113	0.064	0.110	0.109	**0.118**	<0.1	0.7
average:	0.445	0.448	**0.462**	0.301	0.403	0.445	0.451		

Table 4: Mean average precision values by making use of color features.

database	$SQFD_{f_-}$	$SQFD_{f_h}$	$SQFD_{f_g}$	HD	PMHD	WCD	EMD	α_{f_h}	α_{f_g}
Wang	0.567	0.580	**0.586**	0.317	0.439	0.567	0.580	0.9	2.9
Coil100	0.772	0.771	0.811	0.731	**0.828**	0.772	0.809	2.9	2.9
MIR Flickr	0.321	0.321	**0.322**	0.318	**0.322**	0.321	**0.322**	2.9	0.7
101objects	0.097	**0.115**	0.106	0.060	0.076	0.097	0.109	<0.1	0.4
average:	0.439	0.447	**0.456**	0.357	0.416	0.439	0.455		

the highest mean average precision values regardless of the chosen features. Third, the Hausdorff Distance (HD) exhibits the lowest mean average precision values regardless of the combination of database and features. Fourth, the mean average precision values of the Weighted Correlation Distance (WCD) and those of the Perceptually Modified Hausdorff Distance (PMHD) lie in between the values of the similarity measures mentioned before. The latter, however, achieves the highest mean average precision values for the *Coil100* database using color features only.

Concluding the experimental evaluation regarding the effectiveness of the similarity measures, we state that the Signature Quadratic Form Distance outperforms the presented state-of-the-art similarity measures. In the remainder, we will evaluate the second important aspect of retrieval performance, the efficiency of the similarity measures.

Table 5 depicts the computation time values in milliseconds needed to generate a complete ranking of the corresponding database. As the usage of position, color, and texture features exhibits the highest computation times values, we only show the results for these features. Obviously, the computation time values correlate with the database sizes depicted in the second column of Table 5. The Hausdorff Distance and Perceptually Modified Hausdorff Distance exhibit the lowest computation time values followed by the Signature Quadratic Form Distance on the basis of the minus similarity function f_-. The efficiency of the Signature Quadratic Form Distance computation depends on its similarity function. We conclude that the overall efficiency of the Signature Quadratic Form Distance competes with that of the Weighted Correlation Distance and outperforms the efficiency of the Earth Mover's Distance.

To sum up, we state the Signature Quadratic Form Distance is an effective and efficient similarity measure for today's content-based multimedia retrieval systems.

5. CONCLUSIONS AND FUTURE WORK

We presented the Signature Quadratic Form Distance as an adaptive similarity measure for flexible content-based feature representations of multimedia data. We gave mathematical inside into this similarity measure and proved that the Signature Quadratic Form Distance encompasses the classic Quadratic Form Distance which is limited to the comparison of feature histograms of only the same length and structure. The experiments on different real-world multimedia databases showed that the effectiveness of the Signature Quadratic Form Distance outperforms that of state-of-the-art similarity measures. Moreover, the efficiency of the Signature Quadratic Form Distance outperforms that of the well-known Earth Mover's Distance. Thus, the Signa-

Table 5: Computation time values in milliseconds by making use of position, color, and texture features.

database	size	$SQFD_{f_-}$	$SQFD_{f_h}$	$SQFD_{f_g}$	HD	PMHD	WCD	EMD
Wang	1000	458.9	565.1	1127.0	136.0	234.4	615.3	4880.9
Coil100	7200	1981.5	2422.9	4753.8	631.3	1084.6	2591.2	19393.2
MIR Flickr	25000	9254.7	11399.0	22821.1	2701.5	4635.6	12572.2	73182.9
101objects	9196	3443.2	4223.6	8434.9	1019.5	1749.7	4608.2	33226.7
average:		3784.6	4652.6	9284.2	1122.0	1926.1	5096.7	32670.9

ture Quadratic Form Distance is well-suited for efficient and effective multimedia information retrieval based on flexible content-based feature representations of complex multimedia objects.

As future work, we plan to investigate the Signature Quadratic Form Distance via different similarity functions and parameters to improve the effectiveness of the retrieval results. In addition, we plan to study approximation and indexing techniques which enable us to further improve the efficiency of the retrieval process.

6. REFERENCES

[1] C. Beecks, M. S. Uysal, and T. Seidl. Signature Quadratic Form Distances for Content-Based Similarity. In *Proc. 17th ACM Int. Conf. on Multimedia*, pages 697–700, 2009.

[2] C. Beecks, M. S. Uysal, and T. Seidl. Efficient k-Nearest Neighbor Queries with the Signature Quadratic Form Distance. In *Proc. 4th Int. Workshop on Ranking in Databases*, pages 10–15, 2010.

[3] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys*, 40(2):1–60, 2008.

[4] T. Deselaers, D. Keysers, and H. Ney. Features for Image Retrieval: An Experimental Comparison. *Information Retrieval*, 11(2):77–107, 2008.

[5] A. Dorado and E. Izquierdo. Fuzzy color signatures. In *Proc. IEEE Int. Conf. on Image Processing*, volume 1, pages 433–436, 2002.

[6] R. Duda, P. Hart, and D. Stork. *Pattern classification.* Wiley New York, 2001.

[7] 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*, 3(3/4):231–262, 1994.

[8] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples an incremental bayesian approach tested on 101 object categories. In *Proc. Workshop on Generative-Model Based Vision*, 2004.

[9] F. Hillier and G. Lieberman. *Introduction to Linear Programming.* McGraw-Hill, 1990.

[10] R. Hu, S. M. Rüger, D. Song, H. Liu, and Z. Huang. Dissimilarity measures for content-based image retrieval. In *Proc. IEEE Int. Conf. on Multimedia and Expo*, pages 1365–1368, 2008.

[11] M. J. Huiskes and M. S. Lew. The MIR Flickr retrieval evaluation. In *Proc. 1st ACM Int. Conf. on Multimedia Information Retrieval*, pages 39–43, 2008.

[12] D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the Hausdorff Distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

[13] W. K. Leow and R. Li. The analysis and applications of adaptive-binning color histograms. *Computer Vision and Image Understanding*, 94(1-3):67–91, 2004.

[14] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-Based Multimedia Information Retrieval: State of the Art and Challenges. *ACM Trans. on Multimedia Computing, Communications and Applications*, 2(1):1–19, 2006.

[15] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval.* Cambridge University Press, 2008.

[16] S. Marchand-Maillet and M. Worring. Benchmarking Image and Video Retrieval: an Overview. In *Proc. 8th ACM Int. Workshop on Multimedia Information Retrieval*, pages 297–300, 2006.

[17] S. Nene, S. K. Nayar, and H. Murase. Columbia Object Image Library (COIL-100). Technical report, Department of Computer Science, Columbia University, 1996.

[18] B. G. Park, K. M. Lee, and S. U. Lee. Color-based image retrieval using perceptually modified Hausdorff distance. *Journal on Image and Video Processing*, 2008:1–10, 2008.

[19] Y. Rubner. *Perceptual metrics for image database navigation.* PhD thesis, 1999. Adviser: Tomasi, Carlo.

[20] Y. Rubner, C. Tomasi, and L. J. Guibas. The Earth Mover's Distance as a Metric for Image Retrieval. *Int. Journal of Computer Vision*, V40(2):99–121, 2000.

[21] T. Seidl and H.-P. Kriegel. Efficient User-Adaptable Similarity Search in Large Multimedia Databases. In *Proc. 23rd Int. Conf. on Very Large Data Bases*, pages 506–515, 1997.

[22] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[23] J. Wang, J. Li, and G. Wiederhold. SIMPLIcity: semantics-sensitive integrated matching for picture libraries. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(9):947–963, 2001.