Approximating the Signature Quadratic Form Distance Using Scalable Feature Signatures*

Jakub Lokoč

SIRET research group, Department of Software Engineering,
Faculty of Mathematics and Physics, Charles University in Prague
lokoc@ksi.mff.cuni.cz

Abstract. The feature signatures in connection with the signature quadratic form distance have become a respected similarity model for effective multimedia retrieval. However, the efficiency of the model is still a challenging task because the signature quadratic form distance has quadratic time complexity according to the number of tuples in feature signatures. In order to reduce the number of tuples in feature signatures, we introduce the scalable feature signatures, a new formal framework based on hierarchical clustering enabling definition of various feature signature reduction techniques. As an example, we use the framework to define a new feature signature reduction technique based on joining of the tuples. We experimentally demonstrate our new feature signature reduction technique can be used to implement more efficient yet effective filter distances approximating the original signature quadratic form distance. We also show the filter distances using our new feature signature reduction technique significantly outperform the filter distances based on the related maximal component feature signatures.

Keywords: Similarity Search, Approximate Search, Content-based Retrieval, Signature Quadratic Form Distance, Scalable Descriptor.

1 Introduction and Related Work

The content-based multimedia retrieval [6] has become an integral part of various information systems managing multimedia data (e.g., e-shops, image banks, industry and medical systems), providing users an alternative to the keyword-based retrieval approaches. In order to search the multimedia data in the content-based way, the systems often employ a similarity model enabling ranking of the database objects according to a query object, where the similarity model comprises multimedia data descriptors and a suitable similarity measure defined for the utilized descriptors. The selection of a proper similarity model then belongs among key tasks when designing an effective and efficient content-based multimedia retrieval system. During the last decades, many types of similarity models

* This research has been supported by Czech Science Foundation project GAČR P202/12/P297.

C. Gurrin et al. (Eds.): MMM 2014, Part I, LNCS 8325, pp. 86–97, 2014.

have been designed and even standardized for a particular multimedia retrieval tasks (e.g., the MPEG-7 standard [15]). One of the most popular similarity models investigated during the last decade is the bag of visual words (BoVW) model [18], utilizing a statically-created vocabulary of codewords, so called codebook. In the BoVW model, each object is represented as a frequency histogram of codewords present in the object, where all the objects in a database share one codebook. Such representation enables efficient retrieval using inverted files, a well established technique for the text-based retrieval area. Whereas the efficiency of the BoVW model is sufficient for large scale multimedia retrieval, the practical effectiveness of the model is still an open problem. While recent works have tried to improve the effectiveness of the BoVW model using semantic preserving models [19], Hamming embedding [10], compressed Fisher vectors [16] or vectors of locally aggregated features [11], several new approaches have relaxed from a common static vocabulary and investigated more general similarity models based on the feature signatures [17] and the adaptive distance measures (e.g., Signature Quadratic Form Distance [3] or Signature Matching Distance [1]). The signature-based models utilize an object specific vocabulary and thus can flexibly represent the contents of an object. Hence, the feature signatures can capture more disparities in the data, which can be beneficial in dynamic databases rapidly changing in content (e.g., multimedia streams). As recently shown, several signature-based models can outperform the BoVW approaches in the terms of effectiveness [1], however, the efficiency of the signature-based models is still a challenging task, especially for feature signatures comprising a high number of tuples. In [9], the authors employ metric/ptolemaic indexing to improve the efficiency of the retrieval, however, the approach is restricted only to distances satisfying metric/ptolemaic postulates. In [13], the authors show signature-based models can be utilized for effective re-ranking when obtaining a candidate result set using an efficient model based on a subset of the MPEG-7 descriptors. In this paper, we focus on new feature signature reduction techniques enabling more efficient yet still effective retrieval. Furthermore, we consider also scalability of the reduction techniques enabling adjusting the size of the feature signatures according to the actual system load. Let us now recall several basic concepts and definitions referred in this paper.

1.1 Feature Signatures and Signature Quadratic Form Distance

Feature signatures [17] have been introduced to flexibly aggregate and represent the contents of a multimedia object mapped into a feature space $\mathbb{F}$. Whether the requested feature space $\mathbb{F}$ comprises color, position, texture information, SIFT gradient vectors or other complex features [7,14], the feature signatures are often obtained by an adaptive variant of the k-means clustering selecting the most significant centroids. In Figure 1, we depict an example of image feature signatures according to a CPT feature space[1]. The feature signatures were extracted using the GPU extractor [12] employing an adaptive k-means clustering

[1] Color $\langle L, a, b\rangle$, position $\langle x, y\rangle$ and texture information $\langle contrast, entropy\rangle$, $\mathbb{F} \subseteq \mathbb{R}^7$.

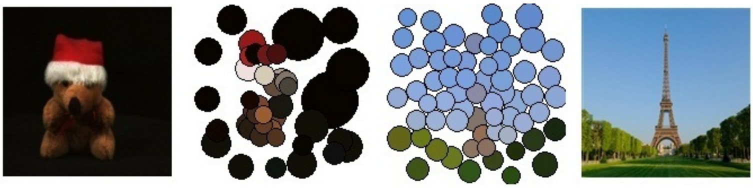

Fig. 1. Example of feature signatures

algorithm, where the extraction of the first image in Figure 1 has put stress on the color, while the extraction of the second image in the figure has put stress on the position. The representatives $r_i \in \mathbb{F}$ corresponding to the selected centroids are depicted by circles in the corresponding position and color, while the weights $w_i \in \mathbb{R}^+$ corresponding to the size of the cluster determine the diameter of the circles (texture information is not depicted). Formally, the feature signatures are defined as:

Definition 1 (Feature Signature). *Given a feature space $\mathbb{F}$, the feature signature S^o of a multimedia object o is defined as a set of tuples $\{\langle r_i^o, w_i^o \rangle\}_{i=1}^n$ from $\mathbb{F} \times \mathbb{R}^+$, consisting of representatives $r_i^o \in \mathbb{F}$ and weights $w_i^o \in \mathbb{R}^+$*

The number of tuples in a feature signature can vary depending on a complexity of a corresponding multimedia object and the parameters used for the extraction. As a consequence, a feature signature can comprise tens or hundreds of tuples, which significantly affects the time for similarity computations. In [2], the authors propose a simple feature signature reduction technique based on maximal components of a feature signature O, where the maximal component feature signature O_{MC} with c components is defined as: $O_{MC} \subseteq O, |O_{MC}| = c$, such that $\forall \langle r_i^o, w_i^o \rangle \in O_{MC}, \forall \langle r_j^o, w_j^o \rangle \in O - O_{MC} : w_i^o \geq w_j^o$. In other words, the maximal component feature signature contains c tuples with the highest weights. The authors also define a signature quadratic form filter distance, applicable for approximate filter and refine retrieval, where the filter distance just evaluates the signature quadratic form distance using maximal component feature signatures. In Figure 2, we depict an example of maximal component feature signatures with 10 and 20 components. We may observe the maximal component feature signatures can omit representative tuples from the original feature signatures (the rightmost signatures in Figure 2) when few maximal components are utilized.

Let us now shortly recall the Signature Quadratic Form Distance [3], an effective adaptive distance measure generalizing the quadratic form distance.

Definition 2 (SQFD). *Given two feature signatures $S^o = \{\langle r_i^o, w_i^o \rangle\}_{i=1}^n$ and $S^p = \{\langle r_i^p, w_i^p \rangle\}_{i=1}^m$ and a similarity function $f_s : \mathbb{F} \times \mathbb{F} \to \mathbb{R}$ over a feature space $\mathbb{F}$, the signature quadratic form distance SQFD_{f_s} between S^o and S^p is defined as:*

$$\mathrm{SQFD}_{f_s}(S^o, S^p) = \sqrt{(w_o \mid -w_p) \cdot A_{f_s} \cdot (w_o \mid -w_p)^T} \,,$$

where $A_{f_s} \in \mathbb{R}^{(n+m)\times(n+m)}$ *is the similarity matrix arising from applying the similarity function* f_s *to the corresponding feature representatives, i.e.,* $a_{ij} = f_s(r_i, r_j)$*. Furthermore,* $w_o = (w_1^o, \ldots, w_n^o)$ *and* $w_p = (w_1^p, \ldots, w_m^p)$ *form weight vectors, and* $(w_o \mid -w_p) = (w_1^o, \ldots, w_n^o, -w_1^p, \ldots, -w_m^p)$ *denotes the concatenation of weight vectors* w_o *and* $-w_p$*.*

To determine similarity values between all pairs of representatives from the feature signatures, the Gaussian similarity function $f_{gauss}(r_i, r_j) = e^{-\alpha L_2^2(r_i, r_j)}$ or the Heuristic similarity function $f_{heuristic}(r_i, r_j) = 1/(\alpha + L_2(r_i, r_j))$ can be utilized, where α is a parameter for controlling the precision, and L_2 denotes the Euclidean distance. If we utilize similarity function $f_{L_2}(r_i, r_j) = -L_2^2(r_i, r_j)/2$, we obtain the L_2-Signature quadratic form distance [4] suffering from worse effectiveness but computable in linear time.

The rest of the paper is structured as follows: we present the scalable feature signatures and our new reduction technique in the following section, then we experimentally demonstrate in section 3 our new reduction technique can be employed for effective approximate search with the signature quadratic form distance, and finally we conclude the paper and point on the future work in section 4.

2 Scalable Feature Signatures

In this section, we introduce the *scalable feature signatures* – a formal framework based on hierarchical clustering enabling definition of sophisticated reduction strategies for feature signatures. The framework extends and generalizes the maximal component feature signatures [2] primarily designed for approximate filter and refine retrieval. As we experimentally demonstrate, the filter signature quadratic form distance employing the maximal component feature signatures does not approximate the original distance (or its lower bound) well, and thus we focus on new filter distances using new feature signature reduction techniques providing better approximations of the original feature signatures. Unlike the maximal components approach that just removes tuples with small

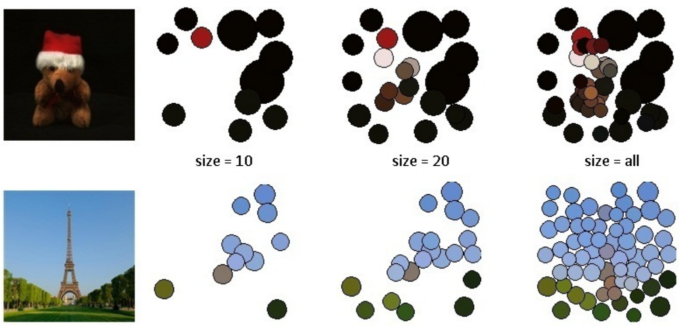

Fig. 2. Maximal component feature signatures

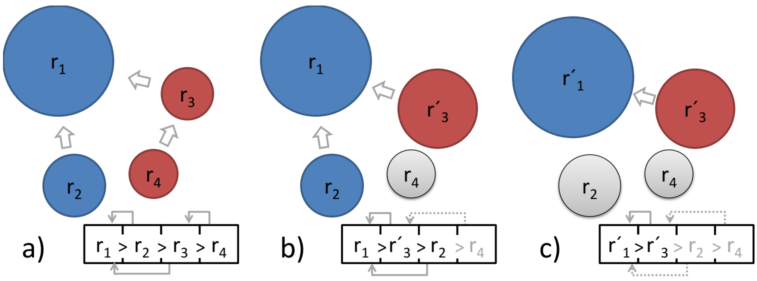

Fig. 3. Scaling feature signature

weights, our new approach aggregates the tuples during the reduction of the feature signatures. Our new approach is motivated by the feature signature extraction process [12], where the adaptive k-means clustering removes centroids with small weights, while the points distributed within the removed centroids are assigned to the remaining centroids. However, after the extraction process is finished and new feature signatures are stored, the points are no longer available and thus only information in the stored tuples can be used for the reduction of feature signatures.

Before we proceed to formal definitions, let us describe a motivation example depicted in Figure 3 where the feature signature FS in Figure 3a is consecutively reduced to the half of the original size in Figure 3c. If the maximal components approach was used, the reduced feature signature would contain only two blue tuples, which would not correspond to the original image. Therefore, instead of removing tuples, we can join them using an aggregation function τ to keep the original information at least in the aggregated form. To determine which tuples are joined in each step, we expect a total ordering $>$ defined over all the tuples in $\mathbb{F} \times \mathbb{R}^+$ and a mapping function ϕ^{FS} defined for all tuples in FS depicted as gray arrows in Figure 3. Using a suitable $>$, ϕ^{FS} and τ, we may observe the reduced feature signature can be a good approximation of the original feature signature (as depicted in Figure 3). Furthermore, if we store one of the original tuples and a pointer to the joined tuple after each join operation, we can later utilize a reverse split operation to reconstruct the original feature signature (or just less reduced feature signature). Such scalability property of the descriptor can be beneficial because we can balance the actual size of the feature signatures according to actual performance needs of a multimedia retrieval system. Let us also emphasize, the reduction process should be deterministic in order to enable preprocessing optimizations for a particular distance functions.

In the following paragraphs we provide definitions formalizing the key concepts described in the motivation example, starting with the definition of the scalable feature signatures.

Definition 3 (Scalable Feature Signature). *Given a feature signature FS over a feature space $\mathbb{F}$, a total ordering $>$ defined over all tuples in $\mathbb{F} \times \mathbb{R}^+$, a total* *mapping function* $\phi^{FS}: FS \rightarrow FS$ *and an aggregation function* $\tau : (\mathbb{F} \times \mathbb{R}^+)^2 \rightarrow \mathbb{F} \times \mathbb{R}^+$*, then the tuple* $(FS, >, \phi^{FS}, \tau)$ *is called scalable feature signature.*

In the following paragraphs, we show an example of the total ordering and several examples of mapping and aggregation functions. Let $\mathbb{F}$ be the euclidean space over field $\mathbb{R}^n$ and FS be a feature signature over that feature space. We can define a total ordering $>$ using weights and the lexicographic ordering $>_{lex}$ over vectors in $\mathbb{R}^n$ as: $\forall \langle r_i, w_i \rangle, \langle r_j, w_j \rangle \in \mathbb{F} \times \mathbb{R}^+ : \langle r_i, w_i \rangle >_{wl} \langle r_j, w_j \rangle$ if and only if $w_i > w_j \vee (w_i = w_j \wedge r_i >_{lex} r_j)$.

The mapping function ϕ^{FS} can utilize the total ordering $>_{wl}$ and can be defined for each tuple $\langle r_i, w_i \rangle \in FS$ as:

$\phi^{FS}_{min}(\langle r_i, w_i \rangle) = \langle r_i, w_i \rangle$ for $\langle r_i, w_i \rangle = \max_{>_{wl}} FS$, and otherwise as:

$\phi^{FS}_{min}(\langle r_i, w_i \rangle) = \min_{>_{wl}} \{\langle r_j, w_j \rangle : \langle r_j, w_j \rangle \in FS \wedge \langle r_j, w_j \rangle >_{wl} \langle r_i, w_i \rangle\}$.

The mapping function ϕ^{FS}_{min} just maps each tuple from FS to the first greater tuple in FS, except for the maximal tuple that is mapped to itself. The mapping function can consider also a Minkowski distance L_p between the representatives in FS as follows:

$\phi^{FS}_{L_p}(\langle r_i, w_i \rangle) = \langle r_i, w_i \rangle$ for $\langle r_i, w_i \rangle = \max_{>_{wl}} FS$, and otherwise as:

$\phi^{FS}_{L_p}(\langle r_i, w_i \rangle) = \langle r_j, w_j \rangle$ such that $\langle r_j, w_j \rangle \in FS \wedge \langle r_j, w_j \rangle >_{wl} \langle r_i, w_i \rangle \wedge (\forall \langle r_k, w_k \rangle \in FS, k \neq i \neq j : \langle r_k, w_k \rangle >_{wl} \langle r_i, w_i \rangle \implies (L_p(r_j, r_i) < L_p(r_k, r_i) \vee (L_p(r_j, r_i) = L_p(r_k, r_i) \wedge \langle r_k, w_k \rangle >_{wl} \langle r_j, w_j \rangle)))$.

The aggregation operation τ can be defined trivially as a projection:

$\tau_{first}(\langle r_i, w_i \rangle, \langle r_j, w_j \rangle) = \langle r_i, w_i \rangle$,

or as a more complex aggregation:

$\tau_{avg}(\langle r_i, w_i \rangle, \langle r_j, w_j \rangle) = \langle r_i \cdot w_i/(w_i + w_j) + r_j \cdot w_j/(w_i + w_j), w_i + w_j \rangle$.

Having defined scalable feature signature $(FS, >, \phi^{FS}, \tau)$, we can now define an unary reduction operation that replaces the minimum $\langle r, w \rangle$ in FS and the corresponding tuple $\phi^{FS}(\langle r, w \rangle)$ by the join tuple $\tau(\langle r, w \rangle, \phi^{FS}(\langle r, w \rangle))$.

Definition 4 (Scalable Feature Signature Reduction). *Given a scalable feature signature* $SFS = (FS, >, \phi^{FS}, \tau)$*, let* $\langle r, w \rangle = \min_{>} FS$*, let* $FS' = (FS - \{\phi^{FS}(\langle r, w \rangle), \langle r, w \rangle\})$ *and let* $\langle r_k, w_k \rangle = \tau(\phi^{FS}(\langle r, w \rangle), \langle r, w \rangle)$*, then the reduction of scalable feature signature SFS denoted as* $\otimes SFS$ *is defined as* $\otimes SFS = (FS_r, >, \phi^{FS_r}, \tau)$*, where* $FS_r = FS' \cup \{\langle r_k, w_k \rangle\}$ *for* $\langle r_k, w_k \rangle \notin FS'$*, and* $FS_r = FS' - \{\langle r_k, w_k \rangle\} \cup \{\langle r_k, 2 \cdot w_k \rangle\}$ *otherwise.*

Let as denote ϕ^{FS_r} is defined in the same way as ϕ^{FS}, but in the context of new feature signature FS_r. We provide also a simple lemma to emphasize the unary reduction operation creates another scalable feature signature. The lemma is without proof because it is a direct consequence of the previous definitions.

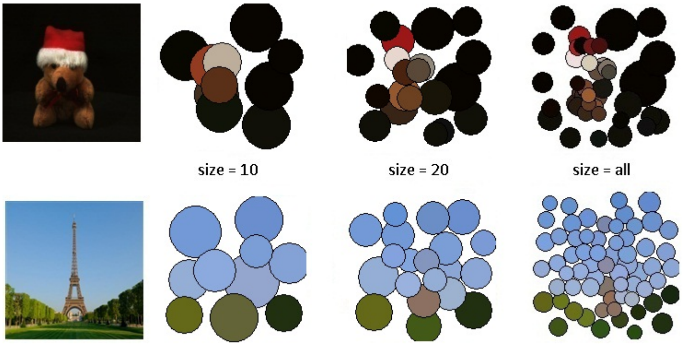

Fig. 4. Scalable feature signatures using $\phi_{L_2}^{FS}$ and τ_{avg}

Lemma 1. *Let $(FS, >, \phi^{FS}, \tau)$ be a scalable feature signature over a feature space $\mathbb{F}$, then $\otimes(FS, >, \phi^{FS}, \tau)$ is also a scalable feature signature over the feature space $\mathbb{F}$.*

So far we have provided a formal framework enabling definition of various feature signature reduction techniques. Using the framework, we can simply define our new reduction technique based on joining of the tuples as a quintuplet $(FS, >_{wl}, \phi_{L_2}^{FS}, \tau_{avg}, \otimes)$, consisting of the scalable feature signature $(FS, >_{wl}, \phi_{L_2}^{FS}, \tau_{avg})$ and the reduction operation $\otimes$. In Figure 4, we may observe our new reduction technique can approximate the distribution of the tuples in the original feature signature well even for smaller number of tuples.

Let us now provide several notes, for the lack of the space without proofs. First, the $\otimes(FS, >, \phi^{FS}, \tau)$ can create new scalable feature signature with $|FS|$, $|FS| - 1$ or $|FS| - 2$ tuples depending on the result of ϕ^{FS} and τ. Second, in case $|FS| = 1$, the reduction operation does not have to be identity, for example, if the aggregation function τ penalizes one of the arguments. Third, $(FS, >_{wl}, \phi_{min}^{FS}, \tau_{first})$ with the reduction operation $\otimes$ corresponds, except for minor differences[2], to the maximal component feature signatures. However, in the text we will strictly use the label *maximal component feature signatures* in order to distinguish the related work from the scalable feature signatures based on joining of the tuples.

Having defined an operation reducing the size of a scalable feature signature, we can now proceed to the definition of a new signature quadratic form filter distance, generalizing the filter signature quadratic form distance $SQFD_{filter}$ defined in [2], where the filter distance is utilized for the approximate search in a filter and refine architecture. In order to express multiple superpositions of the unary operation $\otimes$, we use in the following definition $\otimes^n(FS, >, \phi^{FS}, \tau)$ notation as a shortcut for $\underbrace{\otimes \cdots \otimes}_{n-times}(FS, >, \phi^{FS}, \tau)$.

[2] The maximal component feature signatures do not assume a total ordering of the tuples.

Definition 5 (Signature Quadratic Form Filter Distance). *Given two reduced scalable feature signatures* $(FS_{r_1}, >, \phi^{FS_{r_1}}, \tau) = \otimes^{|FS_1|-n}(FS_1, >, \phi^{FS_1}, \tau)$ *and* $(FS_{r_2}, >, \phi^{FS_{r_2}}, \tau) = \otimes^{|FS_2|-n}(FS_2, >, \phi^{FS_2}, \tau)$ *over a feature space* $\mathbb{F}$*, and let* $SQFD$ *be the signature quadratic form distance, then the distance* $SQFD_f^n = SQFD(FS_{r_1}, FS_{r_2})$ *is called the signature quadratic form filter distance according to* $SQFD(FS_1, FS_2)$*.*

The filter distance just simply reduces the original scalable feature signatures to a requested size and evaluates the original distance measure for the two reduced feature signatures. If we define the scalable feature signatures using $>_{wl}$, ϕ_{min}^{FS} and τ_{first} (which corresponds to the maximal component feature signatures), then the signature quadratic form filter distance $SQFD_f^n$ corresponds to the filter distance $SQFD_{filter}$ presented in [2]. The new signature quadratic form filter distance can be also utilized for the approximate search in a filter and refine schemes, where the reduced scalable feature signatures can be either cached or evaluated every time the filter distance is requested. Furthermore, such retrieval system can decide to temporarily use a reduced version of the scalable feature signatures also for the refinement step. In order to prevent from storing multiple versions of the scalable feature signatures, we can implement the reduction operation as a reversible update of the original scalable feature signatures enabling to keep just one actual version of the scalable feature signatures.

For example, the reduction operation in Figure 3ab replaces $\phi_{L_2}^{FS}(\langle r_4, w_4\rangle) = \langle r_3, w_3\rangle$ by $\langle r_{3'}, w_{3'}\rangle = \tau_{avg}(\langle r_3, w_3\rangle, \langle r_4, w_4\rangle)$, removes $\langle r_4, w_4\rangle$, inserts pair $(\langle r_4, w_4\rangle$, pointer to $\langle r_{3'}, w_{3'}\rangle)$ into a stack and sorts the tuples. The corresponding reverse operation removes pair $(\langle r_4, w_4\rangle$, pointer to $\langle r_{3'}, w_{3'}\rangle)$ from the stack, inserts $\langle r_4, w_4\rangle$ into the feature signature, replaces $\langle r_{3'}, w_{3'}\rangle$ by $\tau_{avg}^{rev}(\langle r_{3'}, w_{3'}\rangle, \langle r_4, w_4\rangle) = \langle r_3, w_3\rangle$ and sorts the tuples, where τ_{avg}^{rev} is derived from τ_{avg} as:

$$\tau_{avg}^{rev}(\langle r_i, w_i\rangle, \langle r_j, w_j\rangle) = \langle (r_i - r_j \cdot w_j/w_i) \cdot w_i/(w_i - w_j), w_i - w_j\rangle.$$

3 Experimental Evaluation

For the experiments, we make use of the three different datasets, each with different source of ground truth. Specifically, we use a subset of the ALOI dataset [8] comprising 12,000 images divided into 1,000 classes, each class contain 12 images of a 3D object rotated by 30 degrees; a subset of the Profimedia dataset [5] comprising 21,993 images divided into 100 classes, where the ground truth was collected semi-automatically and verified by users; the TWIC dataset [13] comprising 11,555 images forming 197 classes, where each class represents images obtained by a keyword query to the google images search engine. Each TWIC class was further manually filtered by users. The feature signatures were extracted using a GPU extractor tool [12]. For all three datasets we have used the same extractor parameters except the multiplicative vector that was adjusted to each dataset separately. The average number of tuples in feature signatures was 33 for ALOI dataset and 66 for TWIC and Profimedia datasets. As the query

Table 1. The time (in milliseconds) needed to evaluate the filter distances using various number of tuples (1, 2, 4, ..., 64) and the signature quadratic form distance (all)

	1	2	4	8	16	32	64	all
Gaussian	0.0006	0.0007	0.0015	0.0039	0.013	0.050	0.193	0.205
Heuristic	0.0004	0.0006	0.0013	0.0031	0.010	0.038	0.149	0.159

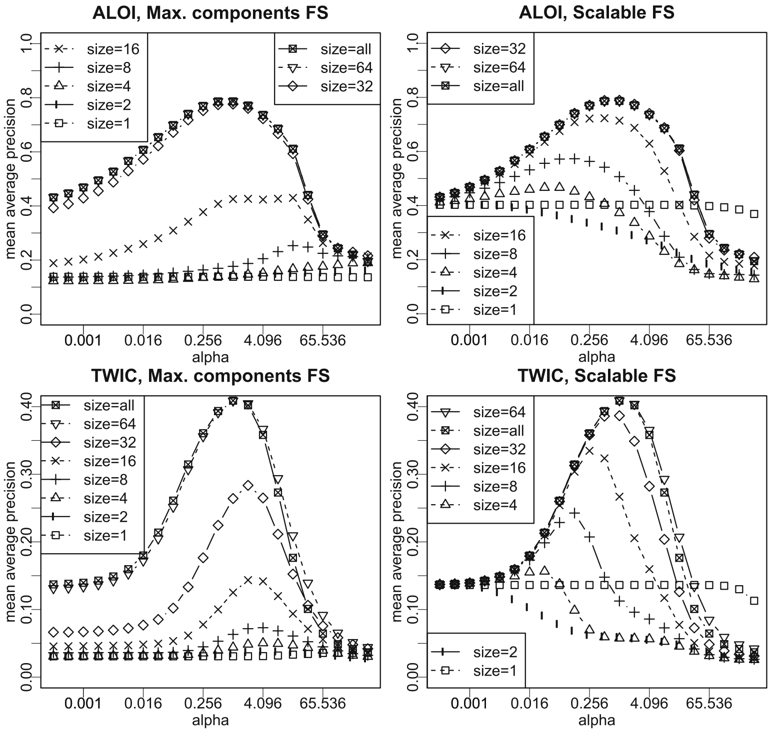

Fig. 5. Mean average precision of the filter distance $SQFD_f^{size}$ utilizing Gaussian similarity function

objects, one representative from each class was selected for all three datasets[3], resulting in 1000 query objects for ALOI, 100 query objects for Profimedia and 197 query objects for TWIC. The experiments have run on 64-bit Windows Server 2008 R2 Standard with Intel Xeon CPU X5660, 2.8 GHz.

In the experiments, we use $(FS, >_{wl}, \phi_{min}^{FS}, \tau_{first})$ as an implementation of the maximal component feature signatures and compare them to the scalable feature signatures using joining of the tuples $(FS, >_{wl}, \phi_{L_2}^{FS}, \tau_{avg})$. For each reduction technique, we utilize six variants of the filter distance $SQFD_f^{size}$ using $size \in \{1, 2, 4, 8, 16, 32, 64\}$ and compare them to the original distance denoted

[3] Profimedia dataset is already provided with a set of query objects.

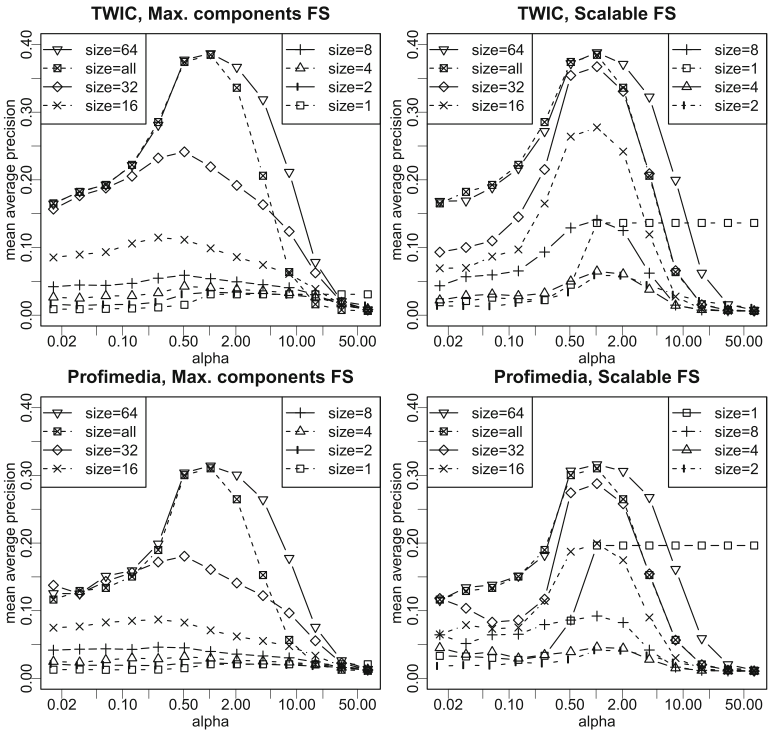

Fig. 6. Mean average precision of the filter distance $SQFD_f^{size}$ utilizing Heuristic similarity function

as *size* = *all*. Before we proceed to the experiments comparing the two reduction techniques, we present a table of average times (in milliseconds) needed to evaluate the utilized filter distances and the original distance, measured for the TWIC dataset for both Gaussian and Heuristic similarity functions. In Table 1, we may observe the Heuristic similarity function is slightly faster then the Gaussian similarity function. We may also observe the expected quadratic time dependency of the signature quadratic form filter distance on the number of tuples in the reduced feature signatures.

Let us now proceed to the following two figures, where the filter distances utilizing the Gaussian similarity function are depicted in Figure 5 and the filter distances utilizing the heuristic similarity function are depicted in Figure 6. In both figures, we have focused on the mean average precision (y-axis) measured for varying parameter α (x-axis). The figures are organized into two columns, where the first column contains results for the maximal component feature signatures (denoted as Max. components FS), while the second column contains the results for the scalable feature signatures using joining of the tuples (denoted simply as Scalable FS). Let us also denote, we have unified the y-axis scaling for each row and thus the reader can directly compare the effectiveness of two corresponding filter distances. In all the graphs we may observe the similar behavior – when decreasing the size of the reduced feature signatures, the filter distances using maximal component feature signatures loose the effectiveness more rapidly than the filter distances using scalable feature signatures based on joining of the tuples. For example, in the second row of Figure 5, we may observe a markable difference between the corresponding pairs of filter distances for signatures comprising 32 and less tuples, where for scalable feature signatures using joining of the tuples the mean average precision is over 30% even for just 16 tuples, while for the same number of tuples and the maximal component feature signatures the mean average precision is just 15%. From the experiments, we may conclude the scalable feature signatures using joining of the tuples provide better filter distances than the maximal component feature signatures.

4 Conclusions and Future Work

In this paper, we have introduced the scalable feature signatures, a formal framework enabling definition of various reduction techniques for feature signatures. As an example, we have defined a new feature signature reduction technique employing joining of the tuples and utilized the technique for definition of effective signature quadratic form filter distances. We have also experimentally demonstrated the filter distances using our new reduction technique significantly outperform the filter distances using maximal component feature signatures. In the future, we plan to examine the scalable feature signatures with other adaptive distance measures and measure the effectiveness of the corresponding similarity models. We would also like to design more complex mapping and joining functions in order to provide more options for the reduction of the scalable feature signatures. We also plan to investigate the performance of the scalable feature signatures on various different features extracted from the images (e.g., SIFT or color SIFT descriptors). We would also like to utilize the scalable feature signatures for more efficient retrieval using new filter and refine schemes or metric/ptolemaic access methods.

References

1. Beecks, C., Kirchhoff, S., Seidl, T.: Signature matching distance for content-based image retrieval. In: Proc. ACM International Conference on Multimedia Retrieval (ICMR 2013), Dallas, Texas, USA, pp. 41–48. ACM, New York (2013)
2. Beecks, C., Uysal, M.S., Seidl, T.: Efficient k-nearest neighbor queries with the signature quadratic form distance. In: Proc. 4th International Workshop on Ranking in Databases (DBRank 2010) in Conjunction with IEEE 26th International Conference on Data Engineering (ICDE 2010), Long Beach, California, USA, pp. 10–15. IEEE, Washington (2010)
3. Beecks, C., Uysal, M.S., Seidl, T.: Signature quadratic form distance. In: Proc. ACM CIVR, pp. 438–445 (2010)

4. Beecks, C., Uysal, M.S., Seidl, T.: $L2$-signature quadratic form distance for efficient query processing in very large multimedia databases. In: Lee, K.-T., Tsai, W.-H., Liao, H.-Y.M., Chen, T., Hsieh, J.-W., Tseng, C.-C. (eds.) MMM 2011 Part I. LNCS, vol. 6523, pp. 381–391. Springer, Heidelberg (2011)
5. Budikova, P., Batko, M., Zezula, P.: Evaluation platform for content-based image retrieval systems. In: Gradmann, S., Borri, F., Meghini, C., Schuldt, H. (eds.) TPDL 2011. LNCS, vol. 6966, pp. 130–142. Springer, Heidelberg (2011)
6. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. ACM Comput. Surv. 40(2), 5:1–5:60 (2008)
7. Deselaers, T., Keysers, D., Ney, H.: Features for image retrieval: an experimental comparison. Information Retrieval 11(2), 77–107 (2008)
8. Geusebroek, J.-M., Burghouts, G.J., Smeulders, A.W.M.: The Amsterdam Library of Object Images. IJCV 61(1), 103–112 (2005)
9. Hetland, M.L., Skopal, T., Lokoč, J., Beecks, C.: Ptolemaic access methods: Challenging the reign of the metric space model. Inf. Syst. 38(7), 989–1006 (2013)
10. Jegou, H., Douze, M., Schmid, C.: Hamming embedding and weak geometric consistency for large scale image search. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 304–317. Springer, Heidelberg (2008)
11. Jegou, H., Perronnin, F., Douze, M., Sánchez, J., Perez, P., Schmid, C.: Aggregating local image descriptors into compact codes. IEEE Trans. Pattern Anal. Mach. Intell. 34(9), 1704–1716 (2012)
12. Kruliš, M., Lokoč, J., Skopal, T.: Efficient extraction of feature signatures using multi-GPU architecture. In: Li, S., El Saddik, A., Wang, M., Mei, T., Sebe, N., Yan, S., Hong, R., Gurrin, C. (eds.) MMM 2013, Part II. LNCS, vol. 7733, pp. 446–456. Springer, Heidelberg (2013)
13. Lokoč, J., Novák, D., Batko, M., Skopal, T.: Visual image search: Feature signatures or/and global descriptors. In: Navarro, G., Pestov, V. (eds.) SISAP 2012. LNCS, vol. 7404, pp. 177–191. Springer, Heidelberg (2012)
14. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
15. MPEG-7. Multimedia content description interfaces. Part 3: Visual. ISO/IEC 15938-3:2002 (2002)
16. Perronnin, F., Liu, Y., Sánchez, J., Poirier, H.: Large-scale image retrieval with compressed fisher vectors. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3384–3391 (2010)
17. Rubner, Y., Tomasi, C.: Perceptual Metrics for Image Database Navigation. Kluwer Academic Publishers, Norwell (2001)
18. Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003), vol. 2, p. 1470. IEEE Computer Society, Washington, DC (2003)
19. Wu, L., Hoi, S.C.H., Yu, N.: Semantics-preserving bag-of-words models and applications. Trans. Img. Proc. 19(7), 1908–1920 (2010)