Efficient Query Processing Using the Earth Mover's Distance in Video Databases

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

ABSTRACT

The rapid increase in generation and dissemination of online video data has recently raised the demand on efficient and effective query processing techniques in large video databases. In this paper, we first introduce a novel compact video representation model to achieve high effectiveness, and then propose to alleviate computational time complexity of the well-known *Earth Mover's Distance* by introducing a filter approximation analyzing earth flows locally and restricting the number of flows globally, ensuring *completeness*. Moreover, extensive experimental evaluation performed on high dimensional real world datasets points out high efficiency and effectiveness of the proposals, significantly reducing the number of Earth Mover's Distance computations and outperforming the state of the art by up to two orders of magnitude with respect to selectivity and query processing time.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.2.4 [**Database Management**]: Systems—*Multimedia databases*

General Terms

Theory, Performance, Experimentation

Keywords

Earth Mover's Distance, Lower Bound, Filter Distance, Efficient Query Processing

1. INTRODUCTION

With increasing ubiquity of the internet and rich diversity of multimedia capture devices and social networking and data sharing web sites, recent years have witnessed an explosion in generation and collection of multimedia data, in particular videos. As reported in [26], 100 hours of video are uploaded to YouTube every minute, and over 6 billion hours of video are watched each month on the same website. The resulting enormous amount of video data in the technological world today makes efficient and effective query processing indispensable for large video databases.

The Earth Mover's Distance (EMD) [15] denoting strong human perceptual similarity is proven to be a very effective distance-based similarity measure in various domains. The EMD determines the dissimilarity between two data objects by the minimum amount of work required to transform one feature representation into another one. Each data object, for example a video clip, can be represented by a *signature* denoting individual object-specific features, or by a *histogram* consisting of shared features in the feature space where histograms expose a special case of signatures. Signatures which are also referred to as *adaptive binning* or *individual binning* can be used to represent a wide spectrum of data types, such as uncertain [3], medical [2], probabilistic [25], and multimedia data [15, 24, 21, 22], as well as events [19] and molecules [6]. The major advantage of the utilization of the signatures is the high quality of content approximation coupled with similarity search and query processing in various type of databases.

A signature is basically defined as a set of features, also found as *representatives* in the literature, in a feature space. Each feature is assigned a real-valued weight denoting the number of features related to that corresponding feature. This is carried out by first extracting the features and then clustering them by using a clustering algorithm, such as k-means algorithm. The resulting counters of the features in the clusters form a signature which we refer to as an *absolute signature* exhibiting individual total weights. Absolute signatures are appropriate for applications for which different characteristics and properties of data are of high importance, such as different image resolution or different video clip length. Many applications rely on an additional preprocessing step by which the absolute signatures are normalized, leading to *relative signatures* exposing a uniform total weight among all data objects. Below, we will immediately show the limitations caused by this normalization step, and how important the usage of absolute signatures is, particularly if partial similarity is involved. Overall, this paper aims at efficient similarity query processing for absolute signatures which is supported only little by existing work.

Relative signatures have often been utilized in numerous applications [15, 1, 24, 21], however, absolute signatures and similarity search using them are still unexplored. The diagnosis of various types of cancer or neurological diseases, such as Alzheimer's Disease [9] require absolute signatures

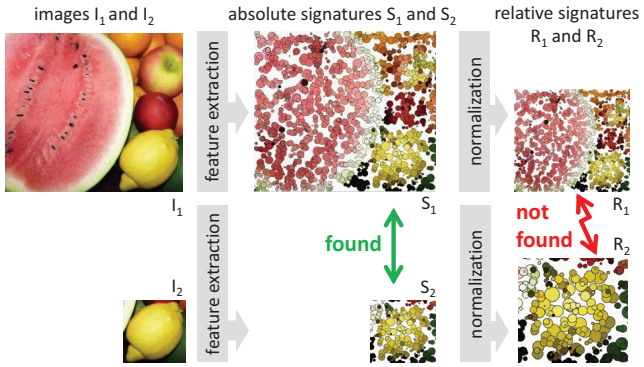


Figure 1: Given images I_1 and I_2, the absolute signature S_2 is found to be a part of the absolute signature S_1 which consider individual absolute weights. However, a normalization step results in relative signatures R_1 and R_2 which are detected as non-similar.

in (bio)medical image classification and similarity search. Rough and coarse boundaries of a cell makes it difficult to determine if it is a cancer/tumor cell. Such biomedical images are commonly stored in fuzzy object databases where each image is partitioned in a specific number of shells where the assignment of a certain probability to each pixel is essential, i.e. the absolute number of pixels for each shell is important to store [20]. An extra normalization step after the feature extraction would lead to inappropriate fuzzy object representation of the cells and, thus, to irrelevant results. In biotechnology, the metabolite identification and quantification is an important task for which data normalization results in obscuring variation of data where the chemical properties cannot be preserved any more [10]. In addition, normalizing the metabolomic data affects its covariance structure which is undesired by the experts.

To contribute to the reader's understanding, we illustrate the absolute and relative signatures in Figure 1. A common task in partial similarity search is to determine if a particular part of a given data object exists in the target dataset. Each signature comprises representatives visualized by circles and is based on the characteristic information of the presented image, such as color. Image I_1 comprises various fruits including also a lemon, while the image I_2 shows only a lemon where the user intends to determine if a lemon exists in the image I_1. When the absolute signatures S_1 and S_2 are considered, S_2 is found to be similar to S_1, since it is detected as a part of S_1. However, if an additional normalization step is applied to attain the relative signatures R_1 and R_2, they are detected as non-similar. A closer look reveals that the two images are evaluated as different images due to the utilization of the normalization step after the feature extraction, i.e. normalizing absolute signatures does not carry out the required partial similarity search task. Another example is the currently attractive and vital domain of video similarity search, in particular near-duplicate video detection, where a typical subclip video detection task [8], required for various purposes as copyright protection and management, entails the need to utilize absolute signatures so that a query subclip can be detected in a given video dataset. Hence, depending on the application, it is explicitly crucial to utilize absolute signatures for queries and datasets, as normalization does not attain partial similarity search tasks formulated and demanded by the user.

Since the empiric time complexity of the EMD is supercubic with respect to feature dimensionality, database community has devised research to propose efficient query processing techniques for the EMD [15, 4, 1, 24, 25, 21]. While existing efficiency improvement techniques for the EMD have been successfully utilized on relative signatures, nevertheless most of them have unfortunately the shortcoming that they can only be applied to fixed-binned relative signatures. Furthermore, they cannot be applied to absolute signatures denoting individual total weights which come up in numerous applications and domains, such as in computer vision [13, 4], multimedia databases [11], fuzzy object databases [20], and biotechnology [10]. While the lower-bounding technique IM-Sig (Independent Minimization for Signatures) [21] is proven to result in efficient results, it can only be applied to relative signatures, not to absolute signatures.

In this paper, we introduce a lower-bounding filter approximation technique *IM-Sig** which is applicable to both fixed-binned and adaptive-binned absolute and relative signatures. In particular, our approach computes the same filter distance as for IM-Sig on relative signatures, and on top of this, our proposal can also be applied to the absolute signatures, hence, filling the gap with respect to lower-bounding the EMD on absolute signatures. To this end, we focus on efficient query processing with the EMD on both relative and absolute signatures in order to introduce a comprehensive solution, which is carried out by analyzing earth flows locally and restricting the number of flows globally. In addition, we take the video domain as an example in this paper, however, it is noteworthy that our efficiency improvement technique can be applied to all domains where complex data objects need to be represented by relative or absolute signatures, as mentioned above. The main contributions of our paper are listed as follows:

- We introduce an adaptive-binning video representation model applicable to the EMD (Section 3).
- We propose an analytic solution IM-Sig* for adaptive-binned signatures without any weight restriction (Section 5.1-5.2).
- We show the optimality of our solution leading to the lower-bounding property of the IM-Sig*, ensuring exact query processing with the EMD (Section 5.3).
- We develop an algorithm for our proposal and analyze the computational time complexity (Section 5.4).
- Experiments on real world data show the efficiency and efficacy of our approach (Section 6).

2. RELATED WORK

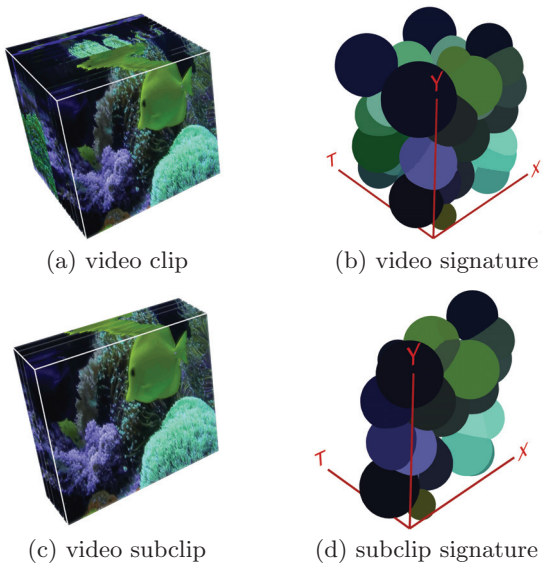
Efficient Query Processing with the EMD. Various efficient query processing techniques have been proposed for the EMD on histograms, i.e. fixed-binned signatures. [1] proposed to lower-bound the EMD via L_p-based distances and constraint relaxation. [24] developed dimensionality reduction techniques for the EMD where reduced cost matrices are utilized relying on the original cost matrix. Furthermore, [25] derived a lower bound of the EMD by utilizing the primal-dual theory in linear programming on top of B^+-trees. In addition, [16] proposed to lower-bound the EMD by

Table 1: Overview of the lower bounds to the Earth Mover's Distance regarding feature representations

lower bounds to the EMD	signatures	
	adaptive binning	individual signature weight
L_p-based [1]		
RedEMD [24]		
PrimalDual [25]		
Rubner [15]	✓	
IM-Sig [21]	✓	
Pemd [4]	✓	✓
IM-Sig*	✓	✓

projecting histograms on a vector and approximating their distance by a normal distribution. It is noteworthy that the limitation of all aforementioned approaches is that they are applicable to histograms sharing the same features in a feature space, not to adaptive-binned signatures denoting individual features per data object. As mentioned in the previous section, since histograms denote a special case of signatures by utilizing shared features instead of individual features per data object, it is of vital importance to propose comprehensive methods applicable to signatures. [15] proposed to lower-bound the EMD by computing the distance between mean signatures. Although the filter time is remarkably low, the efficiency of the query processing is hampered by the worse selectivity resulting in high refinement time. Moreover, a considerable limitation lies in the fact that it cannot be applied to absolute signatures, as will be presented in Section 4. Another efficiency improvement technique for signatures is proposed by [21], which is based on the relaxation of the target constraint of the EMD by local examination of each feature in the source signature. While this method indicates high efficiency improvement, it is nevertheless not applicable to absolute signatures with individual total weights. Furthermore, [4] proposed to lower-bound the EMD on signatures by computing the EMD values for projected signatures each of which comprises features projected on an individual dimension of the feature space. Note that the latter approach is applicable to both relative and absolute signatures. The overview of the applicability of existing lower-bounding methods and our approach IM-Sig* with respect to feature representations is depicted in Table 1.

Video Similarity Search Models. Video similarity search in video databases has been a challenging research area where there have been numerous attempts to provide effective similarity search techniques. [18] (*vitri*) summarizes each video into a small number of clusters each of which includes similar frames. The similarity between two videos is determined by simply estimating the number of similar frames, neglecting the temporal information. [27] (*fras*) is another approach which is an improvement of [18], symbolizing video sequences on a frame basis. The limitation of both approaches lies in the determination of a threshold which is supposed to specify the similarity between any two frames. [8] (*vdt*) proposes to transform a video from a sequence of histograms denoting frames into a one-dimensional distance trajectory where the distance is determined via a reference point. This model is contingent upon frame elements which may lead to performance limitations regarding generation of



(a) video clip (b) video signature

(c) video subclip (d) subclip signature

Figure 2: Illustration of a video clip and a subclip with the visualization of their *video signatures*.

video representations of a high number of frames, and the segmentation of linear segments requires the determination of a threshold regarding the distance between two consecutive frames, which is not robust to outliers. [7] (*bcs*) is another video clip representation model utilizing the principal component analysis in order to specify a bounded range of data projections along each coordinate axis. This method neglects temporal information causing a restriction, in particular for long video clips. Not least, all aforementioned approaches propose to represent each frame only by a histogram in RGB color space.

3. NOVEL VIDEO REPRESENTATION

In this section, we propose our novel video model ***Vi**deo **S**ignature (vis)* which utilizes the determination of individual features and related weights denoting the number of assigned features, leading to a particular compact video clip representation. Unlike frame-based and sequence-based models [7, 8, 18, 27], our model is not contingent upon frames or keyframes, attaining great flexibility via exploiting any requested feature types, such as color, position, contrast, and coarseness. On top of this, the proposal implicitly takes the temporal information into consideration which does not require extra effort at all, since the temporal information is utilized as an individual dimension of the underlying feature space. Mathematically, let $(\mathbb{F}, \delta)$ be a feature space where $\mathbb{F}$ is a set of features coupled with a ground distance function $\delta : \mathbb{F} \times \mathbb{F} \to \mathbb{R}$. Any video clip is represented by a finite set of features $x_1, \ldots, x_n \in \mathbb{F}$. We refer to a *video signature* as a finite set of features (so-called representatives) each of which is assigned a non-negative real number corresponding to the number of features assigned to that representative. The formal definition is given below.

Definition 1 (Video Signature). *Given a feature space $(\mathbb{F}, \delta)$, a video signature V is defined as $V : \mathbb{F} \to \mathbb{R}^{\geq 0}$, subject to $|\mathrm{R}_V| < \infty$, where $\mathrm{R}_V := \{x \in \mathbb{F} \,|\, V(x) > 0\} \subseteq \mathbb{F}$ denotes the set of representatives of V.*

According to the definition above, each representative contributes to the video representation by taking a positive real number $V(x) \in \mathbb{R}^{\geq 0}$ as weight. Any video signature, hence, includes an individual set of representatives and weights which leads to an appropriate video representation.

The illustration of a video clip and a subclip with their corresponding video signature visualizations are depicted in Figure 2. The signatures comprise 50 representatives which are visualized as spheres based on position, color, texture, and temporal information. In the visualization, positional (X,Y) and temporal (T) dimensions are explicitly utilized to contribute to the reader's understanding, where the size of each sphere refers to the weight of that representative. This figure epitomizes the video approximation and modeling with respect to subclip video detection: Obviously, the individual total weight of the absolute signature of the video clip (b) is greater than that for the subclip video (d) which facilitates the utilization of the EMD to determine the dissimilarity between them in order to solve the desired subclip detection task. If they were normalized, the desired task would not be solved, since the EMD would then determine *total similarity* between the two videos, which does not correspond to the user's intention. In the upcoming sections, the term *signature* refers to a video signature. As will be presented in Section 6, modeling videos as *video signatures* highly contributes to effectiveness results. For the sake of simplicity, in the following sections, we utilize the class of non-negative signatures $\mathbb{S}^+ := \{V | V \in \mathbb{R}^{\mathbb{F}} \wedge 0 < |\mathrm{R}_V| < \infty \wedge \forall x \in \mathbb{F} : V(x) \in \mathbb{R}^{\geq 0}\}$ including video signatures whose representatives denote non-negative weights. In the following section, we present the EMD and the utilized filter-and-refine architecture.

4. EARTH MOVER'S DISTANCE

In this section, we present the well-known Earth Mover's Distance which can be utilized in a filter-and-refine architecture in order to boost the query processing. Initally introduced in the computer vision domain, the Earth Mover's Distance (EMD) [15] computes the dissimilarity between two signatures by transforming one signature into another one. The formal definition is given below.

Definition 2. *Let $X, Y \in \mathbb{S}^+$ be two signatures over a feature space $(\mathbb{F}, \delta)$ and $\delta : \mathbb{F} \times \mathbb{F} \to \mathbb{R}$ be a ground distance function. The* Earth Mover's Distance $\mathrm{EMD} : \mathbb{S}^+ \times \mathbb{S}^+ \to \mathbb{R}$ *between X and Y is defined as a minimum-cost flow of all possible flows $F = \{f | f : \mathbb{F} \times \mathbb{F} \to \mathbb{R}\} = \mathbb{R}^{\mathbb{F} \times \mathbb{F}}$ as:*

$$\mathrm{EMD}(X, Y) = \min_{f \in F} \left\{ \frac{\sum_{x \in \mathbb{F}} \sum_{y \in \mathbb{F}} f(x, y) \cdot \delta(x, y)}{\min\{\sum_{x \in \mathbb{F}} X(x), \sum_{y \in \mathbb{F}} Y(y)\}} \right\},$$

subject to constraints $NC \wedge SC \wedge TAC \wedge FC$ with:
NC: $\forall x, y \in \mathbb{F}\ f(x, y) \geq 0$, *SC:* $\forall x \in \mathbb{F} \sum_{y \in \mathbb{F}} f(x, y) \leq X(x)$,
TAC: $\forall y \in \mathbb{F} \sum_{x \in \mathbb{F}} f(x, y) \leq Y(y)$, *and*
FC: $\sum_{x \in \mathbb{F}} \sum_{y \in \mathbb{F}} f(x, y) = \min\{\sum_{x \in \mathbb{F}} X(x), \sum_{y \in \mathbb{F}} Y(y)\}$.

The EMD is the minimum cost required to transform one signature into another one by guaranteeing the non-negativity (NC), source (SC), target (TAC), and total flow constraints (FC), as given above. Hence, the EMD denotes a linear optimization problem and can be solved by simplex algorithms. Figure 3(a) illustrates an example EMD computation between two signatures X, Y where 1 unit earth is transferred from $x_2 \in R_X$ with a distance of 1, resulting in the EMD value of $1 \times 1 = 1$. It is worth noting that the absolute signatures X and Y exhibit individual total weights of 2 and 1, respectively, where the total flow constraint FC guarantees that the minimum of the total weights is transferred from the source signature X to the target signature Y.

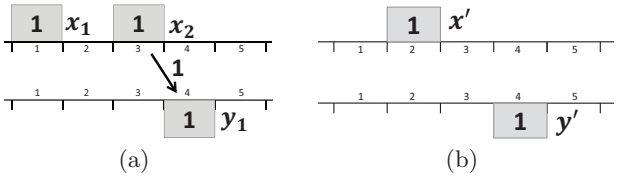


Figure 3: (a) The EMD between absolute signatures X and Y is $EMD(X, Y) = 1$. (b) $Rubner(X, Y) = \delta(x', y') = 2 \not\leq EMD(X, Y)$ holds, i.e. Rubner filter is not a lower bound to the EMD on absolute signatures denoting individual total weights.

Given two signatures X, Y with total weights m_X, m_Y, and a norm-based ground distance function δ, the Rubner filter distance [15] is defined as: $\delta\big((\sum_{x \in \mathrm{R}_X} X(x) \cdot x)/m_X, (\sum_{y \in \mathrm{R}_Y} Y(y) \cdot y)/m_Y\big)$, as depicted in Figure 3(b): x' and y' refer to weighted mean features of X and Y from (a), respectively. The Rubner filter computes $\delta(x', y') = 2$ which, however, does not lower-bound the EMD. As illustrated by this example, the restriction of the Rubner filter is that it cannot be applied to absolute signatures.

Filter-and-refine Architecture. One of the efficiency improvement methods utilized in k-nearest-neighbor (k-nn) query processing is the filter-and-refine architecture model comprising filter and refinement steps [5, 17, 12], as summarized in Figure 4. In the filter step, a filter LB_d generates a set of candidates which is then refined in the refinement step by utilizing the exact distance d (here EMD). A filter ideally fulfills the following properties: First, its computation is attained more efficiently than for the exact distance computation (*efficiency*). Second, LB_d lower-bounds d, i.e. the final refined set includes all objects from the result set, guaranteeing no false dismissals, as it holds $\forall x, y : \mathrm{LB}_d(x, y) \leq d(x, y)$ (*completeness*). Third, the generated set of candidates is smaller if LB_d is tighter, leading to lower computation cost.

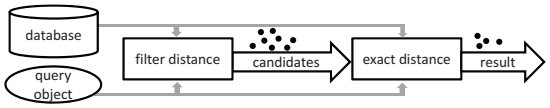


Figure 4: Multistep query processing

In this paper, we utilize a multistep approach which is proven to be optimal in the number of candidates [17]. After a ranking is generated by using a filter distance, it is processed as long as the filter distance does not exceed the exact distance of the current k^{th}-nearest neighbor where the result set and the k^{th}-nearest neighbor distance are continuously updated. After giving the EMD and filter-and-refine architecture, we below present our proposed technique IM-Sig* applicable to both absolute and relative signatures.

5. LOWER-BOUNDING THE EMD ON SIGNATURES

In this section, we first deal with the shortcoming of the existing approach IM-Sig, and then present our proposed comprehensive lower-bounding technique IM-Sig* on signatures, irrespective of any prior information about their total weights. Then, we theoretically show that our analytic solution is both feasible and optimal, which is thus a lower bound to the EMD on all type of signatures including absolute and relative signatures with individual and uniform total weights, respectively. We finally present the computational algorithm of our technique and its complexity analysis.

Our approach has two attractive advantages: First, it is applicable to both relative and absolute signatures, yielding high flexibility with respect to query processing and explicit user-driven tasks, such as subclip video detection, as mentioned before. Second, it is a generic solution regarding efficient query processing which is not restricted to the video domain, and, hence, can be applied to any other domain, such as multimedia, computer vision, and medicine.

5.1 IM-Sig* and the Limitations of IM-Sig

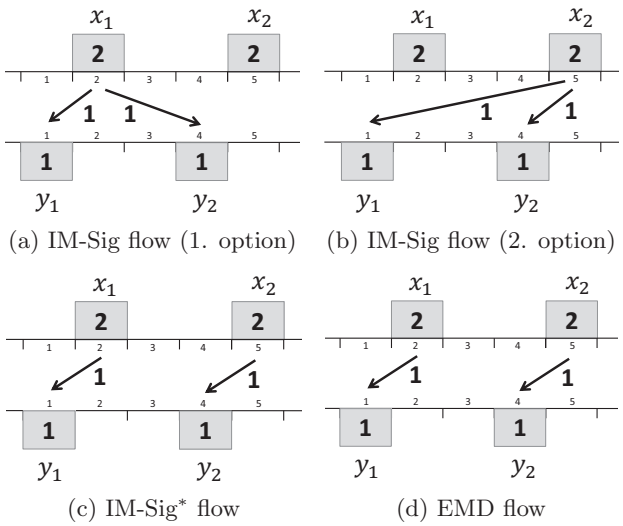
The filter approximation technique Independent Minimization for Signatures (IM-Sig) [21], based on target constraint relaxation of the EMD, was originally proposed to lower-bound the EMD on signatures with uniform total weight. Below, we first give the formal definition of the comprehensive lower-bounding technique IM-Sig*, irrespective of any prior information about the total weights of signatures, and then present the shortcoming of IM-Sig via illustrative examples.

DEFINITION 3 (IM-SIG* FILTER DISTANCE). *Let $(\mathbb{F}, \delta)$ be a feature space with a distance function δ and $X, Y \in \mathbb{S}^+$ be two non-empty positive signatures with weights $m_X = \sum_{x \in \mathbb{F}} X(x)$ and $m_Y = \sum_{y \in \mathbb{F}} Y(y)$. The comprehensive filter distance Independent Minimization for Signatures IM-Sig* : $\mathbb{S}^+ \times \mathbb{S}^+ \to \mathbb{R}^{\geq 0}$ between X and Y is defined as a minimization over all possible flows $F = \{f | f : \mathbb{F} \times \mathbb{F} \to \mathbb{R}\}$:*

$$IM\text{-}Sig^*(X,Y) = \min_{f \in F} \left\{ \sum_{x \in \mathbb{F}} \sum_{y \in \mathbb{F}} \frac{\delta(x,y)}{\min(m_X, m_Y)} f(x,y) \right\},$$

subject to constraints $NC \wedge SC \wedge TC \wedge FC$ with:
NC: $\forall x, y \in \mathbb{F}\ f(x,y) \geq 0$, *SC:* $\forall x \in \mathbb{F} \sum_{y \in \mathbb{F}} f(x,y) \leq X(x)$,
TC: $\forall x, y \in \mathbb{F} : f(x,y) \leq Y(y)$, *and*
FC: $\sum_{x \in \mathbb{F}} \sum_{y \in \mathbb{F}} f(x,y) = \min\{\sum_{x \in \mathbb{F}} X(x), \sum_{y \in \mathbb{F}} Y(y)\}$.

While the non-negativity (NC), source (SC), and total flow constraints (FC) remain unchanged for the EMD and IM-Sig*, the target constraint (TC) of IM-Sig* relaxes that for the EMD by allowing that any single incoming flow (instead of total incoming flows) may not exceed the target capacity. If the signatures exhibit uniform total weights, i.e. if they are relative signatures, the approach IM-Sig [21] can be applied. However, in other cases an appropriate solution is required for the computation of the filter approximation on any kind of signatures with relative or absolute weights. For all possible cases, we propose to compute the comprehensive IM-Sig* by analyzing earth flows locally and restricting the number of flows globally. In other words, our approach IM-Sig* computes the same flow as for IM-Sig for relative signatures and on top of this, IM-Sig* can be applied to the absolute signatures to lower-bound the EMD.



(a) IM-Sig flow (1. option) (b) IM-Sig flow (2. option)

(c) IM-Sig* flow (d) EMD flow

Figure 5: IM-Sig, IM-Sig*, and EMD flows illustrated on two absolute signatures. Since IM-Sig is not defined on absolute signatures exhibiting individual total weights, there is neither a deterministic solution nor a computable minimum-cost flow (a-b). IM-Sig* flow (c) computes an optimal flow which fulfills the minimum-cost property, lower-bounding the EMD (d) on both absolute and relative signatures.

In order to give the underlying basic notion and to clarify the difference between our approach IM-Sig* and IM-Sig, we examine the illustrations given in Figure 5. Numbers 1-5 denote the positions in the 1-dimensional feature space, and the ground distance between any representatives with positions i and j is computed via $|i - j|$, such as $\delta(x_2, y_2) = 1$. The signatures X, Y are illustrated with 2 representatives each, where their weights are denoted in buckets. Since IM-Sig is not defined on absolute signatures, there does not exist a computable minimum-cost flow. Nonetheless, if we try to apply the naive IM-Sig algorithm to these absolute signatures denoting individual total weights, we face two main problems: First, there exists no deterministic solution since IM-Sig basically transfers earth from each representative x_i to its nearest neighbors in the target signature Y. Since there is no specific predefined order of the representatives for the earth transfer from X to Y, one can arbitrarily start with any representative. The first non-deterministic solution of IM-Sig (Fig. 5(a)) would transfer earth from x_1 to its nearest neighbors y_1 and y_2, resulting in IM-Sig$(X,Y) = \frac{1}{2} \times (1 \cdot 1 + 1 \cdot 2) = 1.5$. Another non-deterministic solution of IM-Sig (Fig. 5(b)) would transfer earth from x_2 to its nearest neighbors y_2 and y_1, resulting in IM-Sig$(X,Y) = \frac{1}{2} \times (1 \cdot 1 + 1 \cdot 4) = 2.5$. Second, the computed IM-Sig values do not necessarily yield optimal solutions, which can be inferred from the application of the optimal IM-Sig* on the example (Fig. 5(c)): Our approach first ranks the representative pairs (x_i, y_j) with respect to their ground distance values in ascending order, and then the earth is

transferred from X to Y by taking this order into consideration, as well as all constraints given in Def. 3. Hence, we consider the permutation of $((x_1, y_1), (x_2, y_2), (x_1, y_2), (x_2, y_1))$ with the ground distance values of 1,1,2, and 4. In this way, first 1 unit earth is transferred from x_1 to y_1 and then again 1 unit earth is transferred from x_2 to y_2 after which the optimal solution is attained, fulfilling the IM-Sig* constraints: IM-Sig$^*(X, Y) = \frac{1}{2} \times (1 \cdot 1 + 1 \cdot 1) = 1$. In addition, the EMD (Fig. 5(d)) is computed as $\text{EMD}(X, Y) = \frac{1}{2} \times (1 \cdot 1 + 1 \cdot 1) = 1$, where we observe that the IM-Sig* computes not only the minimum-cost flow but also the feasible solution, lower-bounding the EMD on these absolute signatures. As a result, this example illustrates that IM-Sig is unfortunately not applicable to absolute signatures with individual total weights and there is demand on novel efficient query processing techniques to solve the existing problem.

So far, we have seen the shortcomings of IM-Sig and deduce that there is need for new comprehensive lower-bounding techniques applicable to any kind of signatures without any restriction. Below, we present our novel analytic solution with respect to the computation of IM-Sig*. For the definitions and theoretical analysis in the remainder of the paper, we assume that a feature space $(\mathbb{F}, \delta)$ is given with a distance function δ, and we refer to non-empty positive signatures $X, Y \in \mathbb{S}^+$.

5.2 Analytic Solution

In order to propose our novel analytic solution for IM-Sig*, we first give the definitions of the local feasible set, extensive flow, and global feasible set which are required to define the IM-Sig* flow, as will be given in Definiton 7. The local feasible set of a representative x in the source signature X exhibits the greatest set of nearest neighbors in the target signature Y, where the total weight of its nearest neighbors may not exceed the capacity of x. The formal definition is given below.

Definition 4 (Local Feasible Set). *Given two signatures $X, Y \in \mathbb{S}^+$, let $(y_1, \ldots, y_l)$ be a permutation of R_Y with respect to a feature $x \in \mathbb{F}$ such that $i \leq j \Rightarrow \delta(x, y_i) \leq \delta(x, y_j)$. The local feasible set $S^x_{X,Y} \subseteq R_Y$ is defined as the greatest set of nearest neighbors of x in R_Y whose total weight does not exceed $X(x)$: $y_i \in S^x_{X,Y} \Leftrightarrow \sum_{j=1}^{i} Y(y_j) < X(x)$. Let further $k = |S^x_{X,Y}|$ be the greatest index in $S^x_{X,Y}$, then $\hat{y}_x \notin S^x_{X,Y}$ is defined as the feature directly following y_k regarding the same permutation:*

$$\hat{y}_x = \begin{cases} y_{k+1} & \text{if } k < l = |R_Y| \\ y^* \in \mathbb{F} \backslash R_Y & \text{else} \end{cases}$$

Note that if $|S^x_{X,Y}| = |R_Y|$, i.e. the cardinality of the local feasible set of the representative x is the same as that for the representative set of the target signature Y, $\hat{y}_x \notin R_Y$ is an arbitrarily chosen feature in the feature space which does not belong to R_Y, since the capacity of x exceeds the total signature weight of the target signature, i.e. $X(x) > m_Y$.

The aim of the extensive flow f_e is to transfer earth from the source signature X to the target signature Y by filling up the nearest neighbors of each $x \in R_X$ in the target signature so that for each x all the earth it owns is completely transferred to its nearest neighbors in the target signature. Note that the extensive flow does not take the individual total weights of the signatures into consideration. Technically, the local feasible set $S^x_{X,Y}$ is utilized to define the flow whose definition is given as follows.

Definition 5 (Extensive Flow). *Given two signatures $X, Y \in \mathbb{S}^+$, let $S_{X,Y}(x)$ be the local feasible set for any feature $x \in \mathbb{F}$ (Def. 4). The extensive flow $f_e : \mathbb{F} \times \mathbb{F} \to \mathbb{R}$ is defined as follows:*

$$f_e(x, y) = \begin{cases} Y(y) & \text{if } y \in S^x_{X,Y} \\ X(x) - \sum\limits_{y' \in S^x_{X,Y}} Y(y') & \text{if } y = \hat{y}_x \wedge \hat{y}_x \in R_Y \\ 0 & \text{else} \end{cases}$$

The extensive flow fulfills three constraints of IM-Sig*, namely the non-negativity, source, and IM-Sig* target constraint, which can be summarized in Corollary 1 as follows.

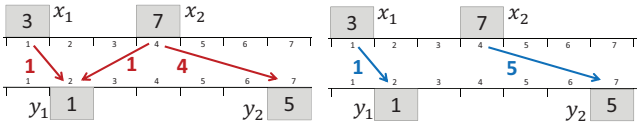
Corollary 1. *Given two signatures $X, Y \in \mathbb{S}^+$, f_e fulfills the following constraints: non-negativity(NC): $\forall x, y \in \mathbb{F}$ $f_e(x, y) \geq 0$, source(SC): $\forall x \in \mathbb{F}$ $\sum\limits_{y \in \mathbb{F}} f_e(x, y) \leq X(x)$, IM-Sig* target(TC): $\forall x, y \in \mathbb{F}$ $f_e(x, y) \leq Y(y)$.*

The corollary above directly follows from Def. 4 and Def. 5. After defining the local feasible set and extensive flow, we now define *global feasible set* which is required for IM-Sig* flow. The global feasible set $S_{X,Y}$ exhibits the greatest set including pairs (x_i, y_j) of representatives from both signatures, where the set includes all pairs for which the corresponding representative y_j receives a positive amount of earth from x_i. An important condition which needs to be fulfilled is that the pairs (x_i, y_j) are ranked according to their ground distance values in ascending order. Hence, the pairs are tracked at a *global* level, i.e. both signatures' representatives are taken into consideration, not only those of the target signature.

Definition 6 (Global Feasible Set). *Given two signatures $X, Y \in \mathbb{S}^+$, let $p = ((x_1, y_1), \ldots, (x_n, y_n))$ be a permutation of $R_X \times R_Y$ such that $i \leq j \Rightarrow \delta(x_i, y_i) \leq \delta(x_j, y_j)$. The global feasible set $S_{X,Y} \subseteq R_X \times R_Y$ is the greatest set satisfying $(x_i, y_i) \in S_{X,Y} \Leftrightarrow \sum_{j=1}^{i} f_e(x_j, y_j) < \min(m_X, m_Y)$, where $k = |S_{X,Y}|$ is the greatest index in $S_{X,Y}$ and $(\hat{x}, \hat{y}) \notin S_{X,Y}$ is defined as $(\hat{x}, \hat{y}) = (x_{k+1}, y_{k+1})$ which directly follows (x_k, y_k) regarding p.*

The global feasible set $S_{X,Y}$ comprises all pairs from $R_X \times R_Y$ sorted according to their ground distances in ascending order where the total amount of the flow coupled with such pairs may not exceed the minimum of the total weights of the signatures. The pair $(\hat{x}, \hat{y})$ is concerned in denoting the last possible flow in the permutation p so that the total flow constraint is guaranteed.

Recall that our goal is to introduce a solution for any pair of signatures, including both relative and absolute total weights, and overcome current limitations. To this end, we explicate our proposed technique IM-Sig* flow which transfers the minimum amount of total weights of given two signatures from the source signature X to the target signature Y. This is achieved by transferring earth by the utilization of the global feasible set which tracks the pairs of representatives allowing for appropriate flows with respect to non-negativity, source, and target constraints. IM-Sig* flow additionally takes the total flow constraint into consideration which is significantly required to yield both feasible and optimal solution to the IM-Sig* minimization problem. The formal definition of IM-Sig* flow is given below.



(a) IM-Sig* flow (b) EMD flow

Figure 6: Illustration of the novel IM-Sig* flow and the EMD flow on signatures.

DEFINITION 7 (IM-SIG* FLOW). *Given two signatures* $X, Y \in \mathbb{S}^+$, *let* $S_{X,Y}$ *be the global feasible set with* $m = \min(m_X, m_Y)$. *For any* $x, y \in \mathbb{F}$ *IM-Sig* flow is defined as:*

$$f_S(x,y) = \begin{cases} f_e(x,y) & \text{if } (x,y) \in S_{X,Y} \\ m - \sum\limits_{(x',y') \in S_{X,Y}} f_e(x',y') & \text{if } (x,y) = (\hat{x},\hat{y}) \\ 0 & \text{else} \end{cases}$$

In order to contribute to the reader's understanding, we illustrate the novel comprehensive IM-Sig* flow by means of an example in Figure 6(a). Numbers 1-7 expose the positions in the 1-dimensional feature space, and the ground distance between any representatives with positions i and j is computed via $|i-j|$. Two signatures X, Y are illustrated with 2 representatives each, where their weights are depicted in buckets. The required permutation is given as $p = ((x_1,y_1),(x_2,y_1),(x_2,y_2),(x_1,y_2))$ with distances 1,2,3,6, respectively. The global feasible set $S_{X,Y} = \{(x_1,y_1),(x_2,y_1)\}$ is then determined as the greatest set whose total flow may not reach the minimum of the total weights of X, Y, i.e. $1+1 < min(10,6) = 6$. Thus, the pair $(\hat{x},\hat{y}) = (x_2,y_2)$ is the last element in the permutation which allows for flow with an amount of only 4 to fulfill the total flow constraint. Hence, IM-Sig* is computed as $\frac{1}{6} \times (1 \times 1 + 2 \times 1 + 3 \times 4) = 2.5$. Not least, when compared with the EMD (Figure 6(b)) computed as $\frac{1}{6} \times (1 \times 1 + 3 \times 5) = 2.66$, it is obvious that the novel IM-Sig* flow leads to a very tight lower bound to the EMD on absolute signatures.

So far, we have seen that the analytic solution IM-Sig* flow can be introduced in order to boost the query processing with the EMD on signatures irrespective of their total weights. Below, we would like to show that our analytic solution is feasible and optimal with respect to the IM-Sig* constraints which leads to the conclusion that the utilization of IM-Sig* flow indeed leads to a lower bound to the EMD on any kind of signatures.

5.3 Theoretical Investigation

Now, we investigate our proposal with respect to its feasibility and optimality regarding IM-Sig* constraints. First, we show that the proposed IM-Sig* flow is a feasible flow fulfilling 4 constraints of non-negativity, source, IM-Sig* target, and total flow, which is given by Theorem 1. Then, we show that IM-Sig* flow is an optimal flow, i.e. there exists no other flow which results in lower overall cost than that for IM-Sig* flow (Theorem 2). Finally, by the utilization of both theorems, we deduce that the utilization of the proposed IM-Sig* flow leads to the lower bound to the EMD on signatures, irrespective of their total weights.

Feasibility Analysis. In order to show the feasibility of our approach, we consider the constraints in Def. 3 and show that these constraints are fulfilled.

THEOREM 1 (FEASIBILITY OF IM-SIG* FLOW). *Given signatures* $X, Y \in \mathbb{S}^+$ *with total weights* $m_X = \sum\limits_{x \in \mathbb{F}} X(x)$ *and* $m_Y = \sum\limits_{y \in \mathbb{F}} Y(y)$, *IM-Sig* flow* f_S *fulfills the following constraints: non-negativity (NC):* $\forall x, y \in \mathbb{F}\ f_S(x,y) \geq 0$, *source (SC):* $\forall x \in \mathbb{F} \sum\limits_{y \in \mathbb{F}} f_S(x,y) \leq X(x)$, *IM-Sig target (TC):* $\forall x, y \in \mathbb{F}\ f_S(x,y) \leq Y(y)$, *and total flow (FC):* $\sum\limits_{x \in \mathbb{F}} \sum\limits_{y \in \mathbb{F}} f_S(x,y) = \min(m_X, m_Y)$.

PROOF. For the proof we consider each constraint given above and show that they are fulfilled regarding the definition of IM-Sig* flow given in Def. 7. For any pair of features x, y, IM-Sig* flow between these features does not exceed the extensive flow between them, i.e. given two signatures $X, Y \in \mathbb{S}^+$, for any $x, y \in \mathbb{F}$ it holds: $f_S(x,y) \leq f_e(x,y)$, following from Def 5 and Def. 7. We denote this fact by the notation $\circledast$ and use it below, where necessary. **NC**: $\forall x, y \in \mathbb{F} : f_S(x,y) \geq 0$. There exist three cases to examine: *Case 1*: $(x,y) \notin S_{X,Y} \wedge (x,y) \neq (\hat{x},\hat{y})$. $\Rightarrow f_S(x,y) = 0$. *Case 2*: $(x,y) \in S_{X,Y}$. $\Rightarrow f_S(x,y) = f_e(x,y) \geq 0$, by Cor. 1. *Case 3*: $(x,y) = (\hat{x},\hat{y})$. Since $\sum\limits_{(x,y) \in S_{X,Y}} f_e(x,y) < \min(m_X, m_Y)$ holds for any two signatures X and Y, we can write the following statement:
$\sum_{(x',y') \in S_{X,Y}} f_e(x',y') < \min(m_X, m_Y) \Rightarrow \min(m_X, m_Y) - \sum_{(x',y') \in S_{X,Y}} f_e(x',y') > 0 \overset{Def.\ 7}{\Rightarrow} f_S(x,y) \geq 0$. **SC**: $\forall x \in \mathbb{F} : \sum\limits_{y \in \mathbb{F}} f_S(x,y) \leq X(x)$. $\sum\limits_{y \in \mathbb{F}} f_S(x,y) \overset{(\circledast)}{\leq} \sum\limits_{y \in \mathbb{F}} f_e(x,y) \overset{\text{SC Cor.1}}{\leq} X(x)$. **TC**: $\forall x, y \in \mathbb{F} : f_S(x,y) \leq Y(y)$. $f_S(x,y) \overset{(\circledast)}{\leq} f_e(x,y) \overset{\text{TC Cor.1}}{\leq} Y(y)$.
FC: $\sum\limits_{x \in \mathbb{F}} \sum\limits_{y \in \mathbb{F}} f_S(x,y) = \min(m_X, m_Y)$. $\sum\limits_{x \in \mathbb{F}} \sum\limits_{y \in \mathbb{F}} f_S(x,y) = \sum\limits_{(x,y) \in R_X \times R_Y} f_S(x,y) = \sum\limits_{(x,y) \in S_{X,Y}} f_S(x,y) + f_S(\hat{x},\hat{y}) \overset{Def.7}{=} \sum\limits_{(x,y) \in S_{X,Y}} f_e(x,y) + \min(m_X, m_Y) - \sum\limits_{(x',y') \in S_{X,Y}} f_e(x',y') = \min(m_X, m_Y)$. □

Optimality Analysis. To prove that IM-Sig* with the proposed flow lower-bounds the EMD, we present that it yields the minimum overall cost among all possible flows by showing that any arbitrarily chosen feasible flow f results in a higher or equal overall cost than that for the IM-Sig* flow.

THEOREM 2 (OPTIMALITY OF IM-SIG* FLOW). *Given signatures* $X, Y \in \mathbb{S}^+$ *with* $m_X = \sum\limits_{x \in \mathbb{F}} X(x)$, $m_Y = \sum\limits_{y \in \mathbb{F}} Y(y)$, *and set of all possible flows* F, f_S *is the minimum-cost flow minimizing the overall cost with respect to IM-Sig*:*

$$f_S = \arg\min_{f \in F} \left\{ \sum_{x \in \mathbb{F}} \sum_{y \in \mathbb{F}} \frac{\delta(x,y)}{min(m_X, m_Y)} f(x,y) \right\}.$$

PROOF. We show that any arbitrarily chosen flow f different from the proposed flow f_S does not lead to lower overall cost. Due to space limitations, we present the idea of the proof instead of giving all the theoretical details. Since f and f_S are feasible, the total amount of earth moved is $\min(m_X, m_Y)$. Let $R_X = E \cup L \cup M$, where E, L, M include features from which f_S transfers an equal, smaller, or greater amount of earth than f, respectively:
$E := \{x \in R_X | \sum\limits_{y \in R_Y} f_S(x,y) = \sum\limits_{y \in R_Y} f(x,y)\}$

$L := \{x \in R_X | \sum_{y \in R_Y} f_S(x,y) < \sum_{y \in R_Y} f(x,y)\}$
$M := \{x \in R_X | \sum_{y \in R_Y} f_S(x,y) > \sum_{y \in R_Y} f(x,y)\}$.

For any $x \in R_X$ and any amount of earth $m \leq X(x)$, the minimum-cost local earth distribution regarding the target constraint is attained by transferring earth from x to its nearest neighbors $y_1, \ldots, y_l$ in R_Y where it holds $1 \leq i \leq j \leq l \Rightarrow \delta(x,y_i) \leq \delta(x,y_j)$, which can also be inferred from [21]. We refer this fact via the notation $\star$ below, where required. Now, we consider 3 cases:

Case 1: For any $x \in E$, the amount of earth transferred by f_S is the same as that of f. By Def.7, f_S transfers earth from any x to its nearest neighbors in R_Y, and by $(\star)$ it is guaranteed that it yields the lowest cost regarding any $x \in E$. Thus, we can conclude:
$\sum_{x \in E} \sum_{y \in R_Y} \delta(x,y) \cdot f_S(x,y) \leq \sum_{x \in E} \sum_{y \in R_Y} \delta(x,y) \cdot f(x,y)$.

Case 2: For any $x \in M$, the total amount of earth f_S transfers from x exceeds that of f. We partition the amount of earth $X(x)$ belonging to the feature x into two parts: $m^{=}_M(x)$ is the amount of earth which each flow transfers to the features in R_Y. The remaining $X(x) - m^{=}_M(x)$ amount of earth is then transferred only by f_S so that it totally transfers more earth than f. Regarding $m^{=}_M(x)$, by $(\star)$ f_S attains the minimum cost by filling up its nearest-neighbors in R_Y (consideration of local distance order of $y \in R_Y$). By Def.7, f_S only transfers earth regarding the pairs (x,y) with $\delta(x,y) \leq \delta(\hat{x},\hat{y})$ (consideration of the global distance order of $(x,y) \in R_X \times R_Y$).

Case 3: For any $x \in L$, the total amount of earth f_S transfers from x is smaller than that of f. We partition the amount of earth $X(x)$ belonging to the feature x into two parts: $m^{=}_L(x)$ is the amount of earth which each flow transfers to the features in R_Y. The remaining $X(x) - m^{=}_L(x)$ amount of earth is then transferred by only f so that it totally transfers more earth than f_S. Regarding $m^{=}_L(x)$, by $(\star)$ f_S attains the minimum cost by filling up its nearest-neighbors in R_Y (consideration of local distance order of $y \in R_Y$). We know f_S can only transfer earth regarding pairs (x,y) with $\delta(x,y) \leq \delta(\hat{x},\hat{y})$, and it does not transfer all the earth from x. In addition, it fills up the features $y \in R_Y$ regarding all pairs (x,y) with $x \in L$ and $\delta(x,y) < \delta(\hat{x},\hat{y})$. In best case, f distributes $m^{=}_L(x)$ amount of earth as f_S distributes, or f distributes it in another way. In the latter case, the last pair (x',y') used for the distribution by f satisfies $\delta(\hat{x},\hat{y}) \leq \delta(x',y')$. Thus, the only place where f transfers the remaining $X(x) - m^{=}_L(x)$ amount of earth involves only the pairs (x,y) satisfying $\delta(\hat{x},\hat{y}) \leq \delta(x',y') \leq \delta(x,y)$ (consideration of the global distance order again).

By the facts elucidated in Case 2 and 3, and the constraint **FC**, it is concluded: $\sum_{x \in M} \sum_{y \in R_Y} \delta(x,y) \cdot (f_S(x,y) - f(x,y)) \leq \sum_{x \in L} \sum_{y \in R_Y} \delta(x,y) \cdot (f(x,y) - f_S(x,y))$. As a result, after considering all the facts, we attain the final statement as follows:
$\sum_{x \in R_X} \sum_{y \in R_Y} \delta(x,y) \cdot f_S(x,y) \leq \sum_{x \in R_X} \sum_{y \in R_Y} \delta(x,y) \cdot f(x,y)$,
which indicates that the overall cost induced by any feasible flow f does not lead to lower overall cost than that of f_S. In other words, the proposed flow f_S is proven to be the minimum-cost flow. □

As mentioned above, Theorem 2 states that IM-Sig* flow is an optimal flow by showing that there exists no other flow resulting in lower overall cost than that for IM-Sig* flow.

Lower-bounding the EMD with IM-Sig*. After presenting that IM-Sig* flow is feasible and optimal, we below show the third significant theoretical result (Theorem 3) from which we deduce that the utilization of IM-Sig* flow leads to the lower bound to the EMD on signatures, regardless of their total weights.

THEOREM 3 (LOWER-BOUNDING EMD). *Given any two signatures $X, Y \in \mathbb{S}^+$ with total weights m_X, m_Y, it holds:*

$$\textit{IM-Sig}^*(X,Y) \leq EMD(X,Y).$$

PROOF. By Theorem 1 and 2, f_S (Def. 7) is a feasible and minimum-cost flow regarding constraints of IM-Sig*. Hence, there exists no other flow leading to smaller overall cost. □

Consequently, the theoretical results provide confirmatory evidence that the proposed IM-Sig* flow can be utilized as a filter distance function lower-bounding the EMD on signatures, including both absolute and relative signatures. To this end, we can present a computational algorithm to compute IM-Sig* between any signatures, given as below.

5.4 Computational Algorithm

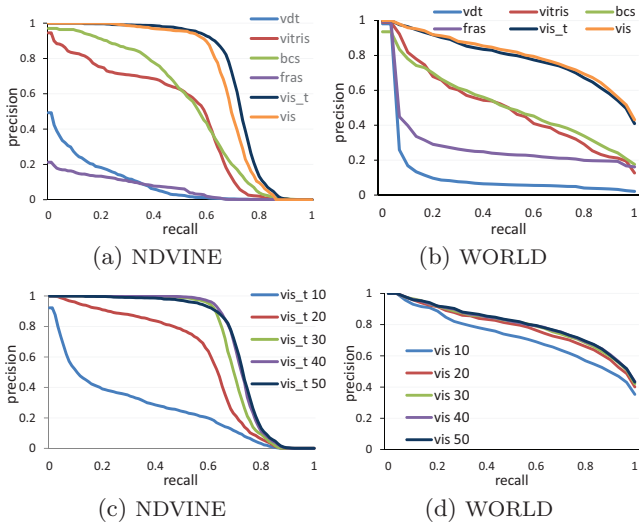
Algorithm 1: IM-Sig* computation

input : signatures X, Y, ground distance δ
output: IM-Sig* between signatures X and Y

```
1  cost = 0
2  construct minHeap for R_X × R_Y regarding δ
3  minWeight = min(m_X, m_Y)
4  initialize sourceCap(x) = X(x) for each x ∈ R_X
5  remainingEarth = minWeight
6  while remainingEarth > 0 do
7      (x, y) = minHeap.poll()
8      if sourceCap(x) > 0 then
9          if Y(y) ≥ remainingEarth then
10             earth = min{remainingEarth, sourceCap(x)}
11         else
12             earth = min{sourceCap(x), Y(y)}
13         end
14         cost = cost + δ(x, y) · earth
15         sourceCap(x) = sourceCap(x) − earth
16         remainingEarth = remainingEarth − earth
17     end
18 end
19 return cost/minWeight
```

Algorithm. The pseudo code of IM-Sig* computation for both absolute and relative signatures with our proposed flow construction is depicted in Algorithm 1. After a min-heap is constructed over $R_X \times R_Y$ with respect to distance values in ascending order (line 2), the algorithm extracts (x,y) with the smallest distance from the min-heap (line 7), until earth in an amount of the minimum weight of both signatures is transferred to Y totally (line 6). Each extracted pair from the min-heap refers to an element in the global feasible set $S_{X,Y}$, and the amount of earth transferred is determined by taking the remaining earth, current source capacity of x, and target capacity of y into consideration (lines 8-17).

Complexity Analysis. Assuming $n = |R_X|, m = |R_Y|$, the min-heap construction is performed in computation time



(a) NDVINE (b) WORLD

(c) NDVINE (d) WORLD

Figure 7: Precision-recall graphs.

complexity $O(n \cdot m)$. For each pair extraction from the min-heap, in worst case $log(n \cdot m)$ many steps are required to ensure the heap property again.

Note that for any given signatures X, Y, it holds that $|S_{X,Y}| < |R_X \times R_Y|$, i.e. the global feasible set $S_{X,Y}$ which is the greatest set whose total flow may not reach the minimum of the total weights of both signatures, and hence, its cardinality is smaller than the number of all possible pairs of representatives in the signatures X and Y. Since the number of extractions from the min-heap is contingent on $min(m_X, m_Y)$ and is bounded by $|S_{X,Y}| < |R_X \times R_Y|$, only t many pair extractions are required with $t < n \cdot m$. Thus, the filter distance computation can be carried out in time complexity $O(t \cdot log(n \cdot m))$ with the proposed IM-Sig* flow.

6. EXPERIMENTAL EVALUATION

Experimental Setup. Results presented in this section expose averages over a query workload of 50 where queries are randomly chosen. All methods are implemented in JAVA and evaluated on a single-core 2.3 GHz machine with Windows Server 2008 and 10 GB of main memory, without parallelization. While we utilize Manhattan distance as ground distance, our approach can, nevertheless, be combined with any ground distance function.

We use three real world datasets. First, we take the data in [14] of 3636 videos and generate approximately 100 near-duplicate copies for each video by altering brightness, contrast, playback speed, resolution, frame order, adding overlay text, borders, and modifying content by frame deletion, yielding a database (NDVINE) of 350,000 videos with 3636 ground truth categories. Second, we use WORLD consisting of 1020 videos we downloaded from *vine.co*(Vine) and *youtube.com*, sorted manually into 34 categories (such as soccer, beach, and forest) of videos which are determined as visually similar according to human perception. For effectiveness experiments, we use the aforementioned databases incorporating video category information. Third, we use the dataset from [23] including 250,000 public social videos of Vine, which we refer as PUBVID. We conduct efficiency experiments with the latter and NDVINE to attain appropriate evaluation regarding data cardinality. We generate video signatures of different dimensionalities as described in Section 3 with position, color, contrast, coarseness and temporal information, while for effectiveness experiments we generate 2 different types of video signatures to recognize the contribution of the temporal information to results: *vis_t* and *vis* denote our novel **vi**deo **s**ignature model with and without **t**emporal information, respectively. In addition, since for any signatures X, Y, IM-Sig$^*(X,Y) \leq max$(IM-Sig$^*(X,Y)$, IM-Sig$^*(Y,X)) \leq$ EMD(X,Y) holds, we implement 3 variations of the algorithm from [17] to evaluate efficiency of query processing, and let *Fqdb, Fdbq, Fmax* refer to filter distance functions with our proposed flow computation where earth is transferred from query signatures to database signatures, from database signatures to query signatures, and where the maximum of both filter distances is utilized in multistep algorithm, respectively. Furthermore, in efficiency experiments we set k=100 for k-nearest-neighbor query processing. The databases used here are available upon request.

Table 2: Single distance computation time (ms)

Dim.	Distance					
	Fdbq	Fqdb	Fmax	Pemd	Rubner	EMD
4	0.0018	0.0023	0.0037	0.0068	0.0005	0.0266
8	0.0051	0.0052	0.0102	0.0070	0.0004	0.0331
16	0.0224	0.0229	0.0452	0.0137	0.0006	0.0858
32	0.1021	0.1035	0.2053	0.0287	0.0012	0.4784
64	0.4740	0.4801	0.9535	0.0651	0.0022	3.9208

Effectiveness Experiments. Figure 7 shows precision-recall graphs of our novel video signature models (*vis_t, vis*), in comparison to four state-of-the-art methods: video triplets (*vitri*) [18], frame-sequence symbolization (*fras*) [27], bounded coordinate systems (*bcs*) [7], and video distance trajectories (*vdt*) [8]. Results summarized in Figure 7(a)-(b) provide confirmatory evidence that our model outperforms the state of the art for both near-duplicate detection (NDVINE) and visual similarity search (WORLD). Considering temporal dimension in the signature model yields better precision for NDVINE, since videos in the same category exhibit a similar temporal ordering. As illustrated in Figure 7(c)-(d), the precision of our models increases with higher signature dimensionality (10-50), as we expected.

Efficiency Experiments. It is noteworthy to remind again that our approach computes the same filter distance as for IM-Sig on relative signatures, and it corresponds to Fdbq in the experimental results presented in this section. First, we evaluate the influence of signature dimensionality on processing time of single distance computation on PUBVID dataset (Table 2). Note that *Pemd* and *Rubner* refer to the existing methods of projected EMD filter [4] and Rubner filter [15]. Since EMD can be computed in super-cubic time in signature dimensionality, it exhibits the highest values, while Fmax' performance is almost 2 times slower than Fqdb abd Fdbq, corresponding to our expectation. Rubner shows the lowest time cost by only computing the ground distance among average signatures.

Figure 8 presents efficiency results of the state of the art (Rubner, Pemd) and our methods (Fqdb, Fdbq, Fmax). With increasing data cardinality, Rubner exhibits the highest time cost, directly followed by Pemd, which are substantially outperformed by our methods regarding selectivity and overall query time. In particular, Fmax computes 49.9

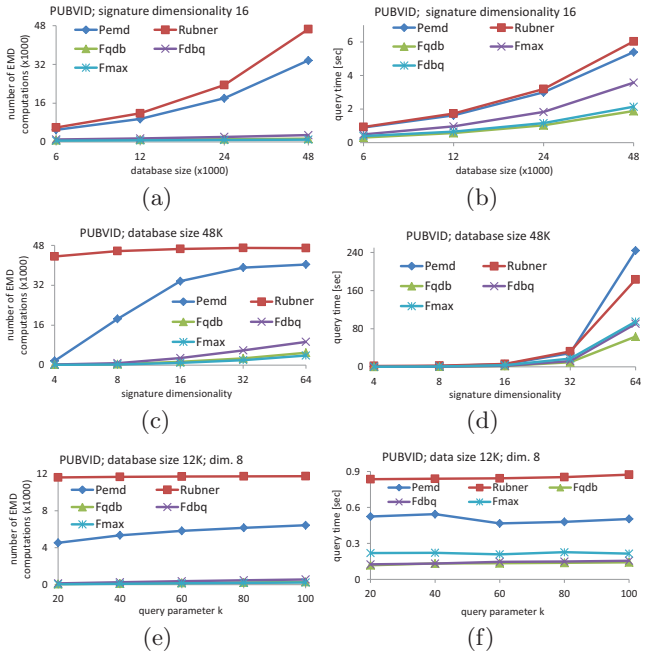


Figure 8: Selectivity and efficiency results with the state of the art.

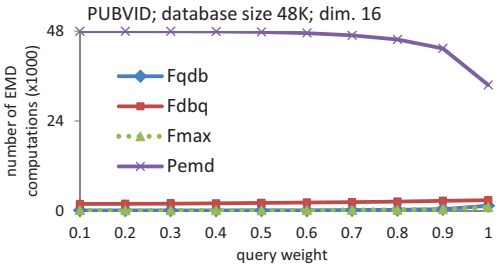


Figure 9: Results with the state of the art regarding selectivity vs. individual total weight of query signatures (absolute signatures).

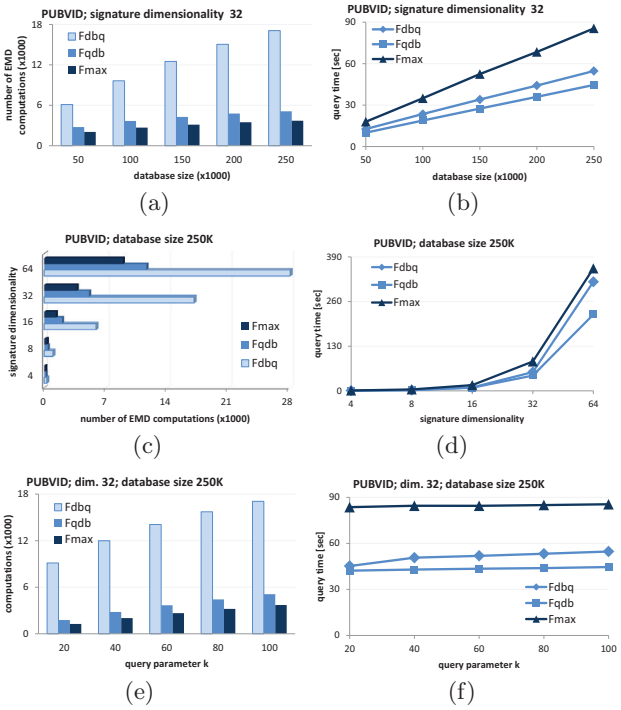


Figure 10: Results regarding selectivity and efficiency on PUBVID database.

and 36.1 times less EMD computations than Rubner and Pemd at data cardinality 48K, respectively. Another result matching our expectation is the constant behavior of Rubner with the worst selectivity, irrespective of dimensionality, while efficiency deterioration of Pemd is remarkable at a very high rate with increasing dimensionality, where, again, our methods outperform both existing approaches. The reason behind these observations is that Rubner simply utilizes distance between average signatures, and Pemd considers single EMD computations performed on each projected dimension, neglecting flow approximation, as given in Section 2. In contrast, our methods explicitly approximate the original EMD flow at a global level by tracking source and target capacity of representatives, and ensuring constraints given in Def. 3, attaining very high efficiency improvement. Furthermore, experiments analyzing the effect of query parameter k for k-nn query processing point out higher efficiency improvement of our proposals. Moreover, we conduct experiments to investigate the applicability to absolute signatures by using video subclip query signatures with varying total weights (0.1-1.0), while ensuring that the weight of any database video signature remains as 1 (Figure 9). Recall that Rubner is not a lower bound here, as illustrated in Section 4, and we observe considerably high selectivity difference between Pemd and our methods, confirming our methods' successful application on absolute signatures. In particular, Pemd refines 99% of all videos between query weights 0.1-0.4, while Fmax refines only at most 0.1%, performing 432 times less EMD computations than Pemd, attaining a selectivity improvement by two orders of magnitude. Note that query time results regarding Figure 9 are omitted, since they expose very similar behavior as those for the number of EMD computations.

After observing that our methods outperform the state of the art, we below perform extensive experiments for our methods, by varying signature dimensionality, database cardinality, and query parameter k for k-nn query processing.

Figure 10 summarizes the efficiency improvement achieved on PUBVID: Fmax indicates the best selectivity, in comparison to other proposed variations, exhibiting the lowest number of EMD computations with increasing data cardinality (50K-250K), and signature dimensionality(4-64), resulting from the computation of a tighter lower bound in the ranking phase of the multistep filter-and-refine algorithm [17]. Since Fmax shows higher filter time, its overall time cost is higher than that for the other two methods. Fqdb shows similar selectivity results, but its lower filter time cost than that for Fmax leads to the fact that Fqdb achieves the highest efficiency improvement regarding database size, dimensionality, and query parameter.

As depicted in Figure 11, evaluation results on NDVINE first indicate an almost constant selectivity behavior for Fqdb and Fmax regardig increasing data size, when compared to Fdbq. This can be elucidated by the intrinsic essence of

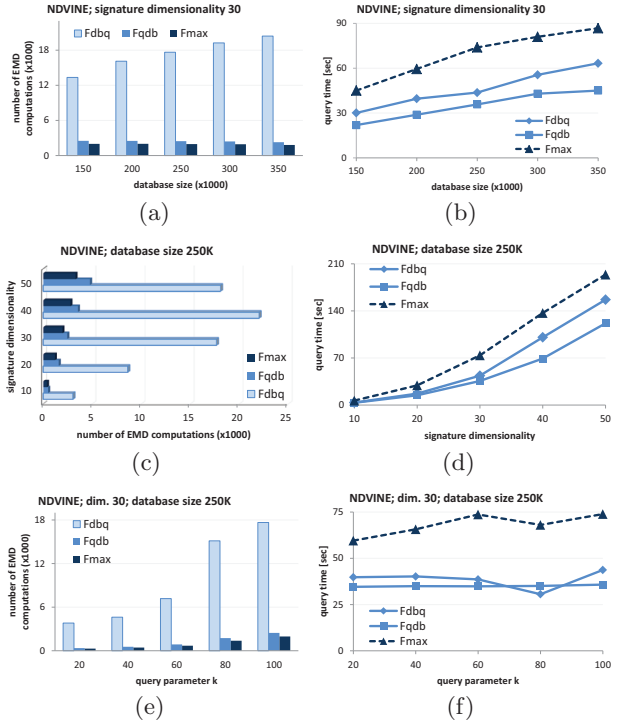


Figure 11: Results regarding selectivity and efficiency on NDVINE database.

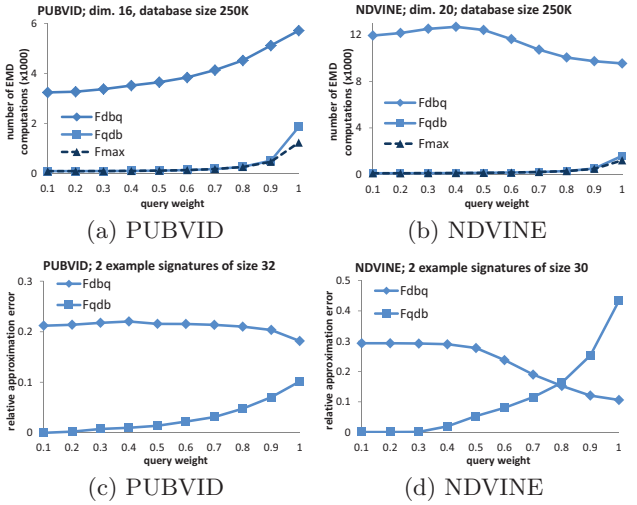


Figure 12: (a)-(b): Selectivity vs. individual total weight of query signatures. (c)-(d): Relative approximation error vs. individual total weight of query signatures for example signature pairs.

this database: With increasing data size near-duplicates of videos emerging via various editing tasks can have similar distances to each other, which affects the filter step. Hence, the higher the data cardinality of near-duplicate video databases, it is more worth using Fqdb to attain an almost constant, low time cost, yielding a considerable advantage. Second, with increasing signature size at a constant data cardinality (250K), the number of promising objects passing Fdbq decreases after dimensionality 40, while Fmax and Fqdb show lower time cost, in particular 1.3 % of selectivity at dimensionality 50 for Fmax. Third, we observe that for all query parameters k (20-100), Fqdb and Fdbq show almost constant behavior for query time due to their lower filter time, matching our expectation.

For both databases, interestingly, Fdbq results in a higher number of EMD computations than for Fqdb. To expound this result, we perform experiments on another real world dataset which we omit due to space limitations, and recognize that selectivity results of Fdbq and Fqdb do not necessarily differ from each other at a high rate. Hence, the observed difference in selectivity and query time can be attributed to the feature distributions of these databases and the utilized queries, which cause Fqdb flow to be more similar to the EMD flow than that for Fdbq. A future research direction involves in further investigating this issue in detail.

Figure 12(a)-12(b) exhibit the effect of individual query signature weights on the number of EMD computations by fixing the total weight of any database video signature to 1. In particular for Fqdb, the number of EMD computations decreases with decreasing query weight, expounded by the fact that the smaller the query weight, the closer Fqdb approximates the EMD flow. To understand it in more detail, we analyze relative approximation error of Fqdb and Fdbq on example pairs of video signatures, summarized in Figure 12(c)-12(d). For smaller query weights, EMD distributes the weight of query representatives more locally among representatives of the database video, since capacities of target representatives are far greater than those for query representatives. Accordingly, since Fqdb distributes earth optimally for each query representative, its flow is similar to the EMD flow, allowing for a better approximation of the EMD than for Fdbq, which meets our expectation. Note that the flow approximation error increases especially after the query weight of 0.5 for both variations. Analogously, the higher the query weight, the better Fdbq approximates the EMD, as a higher query weight causes the EMD to distribute the weight of the database representatives more locally among those of query, resulting in a similar flow to that for Fdbq.

Figure 13 summarizes absolute filter and refinement distances for 10-nn queries on an example from PUBVID at a data cardinality of 1000 and dimensionality 16. Unlike comprehensive overall results for PUBVID, we observe that Fqdb leads to a higher number of refinements (75 in ranking according to filter distance), while Fdbq and Fmax perform 50 and 43 EMD computations, respectively. A significant result gathered from these figures is Fmax leads to the lowest number of exact distance computations, irrespective of the fact how well the other variants approximate the EMD flow. All query time cost results depicted in aforementioned figures point out the advantage of Fqdb with respect to dimensionality, data cardinality, and query parameter k.

7. CONCLUSION

In this paper, we presented how efficient and effective query processing can be performed on high dimensional video databases. We introduced a new compact video representation model, and proposed to alleviate computational time complexity of the Earth Mover's Distance (EMD) by a novel filter approximation guaranteeing completeness (no false dismissals). Furthermore, we presented both an extensive theoretical analysis of our techniques and a computational al-

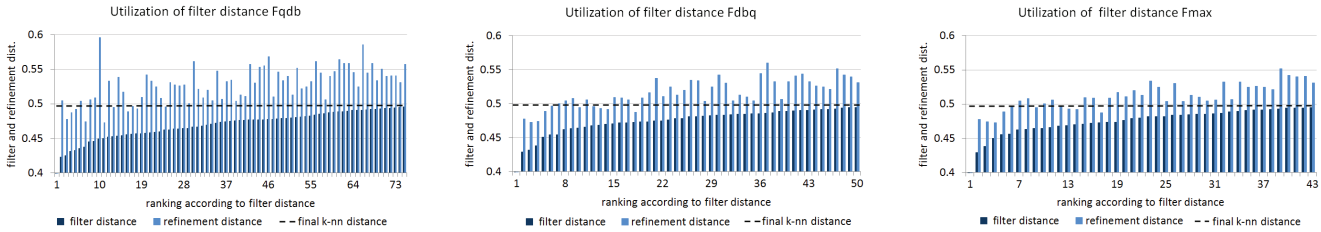


Figure 13: Filter and refinement distance values on an example from PUBVID dataset for 10-nn queries.

gorithm. Moreover, proposed techniques expose two vital advantages: First, they are applicable to both relative and absolute signatures exhibiting uniform and individual total weights, respectively, yielding high flexibility with respect to query processing and explicit user-driven tasks, such as subclip video detection. Second, they exhibit a comprehensive solution which is not restricted to the video domain, and, hence, can be applied to other domains, such as biotechnology and biomedicine. Extensive experimental evaluation on real world data indicates high efficiency, significantly reducing the number of EMD computations and outperforming the state of the art by up to two orders of magnitude regarding selectivity and query time. As future work, we plan to integrate our filter approximation in relational databases.

8. ACKNOWLEDGMENTS

This work is funded by DFG grant SE 1039/7-1.

9. REFERENCES

[1] I. Assent, A. Wenning, and T. Seidl. Approximation techniques for indexing the earth mover's distance in multimedia databases. In *ICDE*, page 11, 2006.

[2] R. S. Chavez and T. F. Heatherton. Representational similarity of social and valence information in the medial pfc. *J. Cogn. Neuroscience*, 27(1):73–82, 2015.

[3] R. Cheng, L. Chen, J. Chen, and X. Xie. Evaluating probability threshold k-nearest-neighbor queries over uncertain data. In *EDBT*, pages 672–683, 2009.

[4] S. D. Cohen and L. J. Guibas. The earth mover's distance: Lower bounds and invariance under translation. Technical report, Stanford Univ., 1997.

[5] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos. Fast subsequence matching in time-series databases. *SIGMOD*, 23(2):419–429, May 1994.

[6] A. Hinneburg et al. Database support for 3d-protein data set analysis. In *SSDBM*, pages 161–170, 2003.

[7] Z. Huang, H. T. Shen, J. Shao, X. Zhou, and B. Cui. Bounded coordinate system indexing for real-time video clip search. *Trans.I.S.*, 27(3):17:1–33, 2009.

[8] Z. Huang, L. Wang et al. Online near-duplicate video clip detection and retrieval: An accurate and fast system. In *ICDE*, pages 1511–1514, 2009.

[9] K. Kantarci. Molecular imaging of Alzheimer disease pathology. *AJNR*, 35:12–17, 2014.

[10] M. Katajamaa and M. Oresic. Data processing for mass spectrometry-based metabolomics. *Journal of Chromatography A*, 1158(1–2):318–328, 2007.

[11] C.-R. Kim and C.-W. Chung. A multi-step approach for partial similarity search in large image data using histogram intersection. *Inf. S.T.*, 45(4):203–215, 2003.

[12] H.-P. Kriegel, P. Kröger, P. Kunath, and M. Renz. Generalizing the optimality of multi-step *k*-nearest neighbor query processing. In *SSTD*, p.75-92, 2007.

[13] O. Pele and M. Werman. A linear time histogram metric for improved sift matching. In *ECCV*, pages 495–508, 2008.

[14] M. Redi, N. OHare, R. Schifanella, M. Trevisiol, and A. Jaimes. 6 seconds of sound and vision: Creativity in micro-videos. In *CVPR*, pages 4272–4279, 2014.

[15] Y. Rubner, C. Tomasi, and L. Guibas. A metric for distributions with applications to image databases. In *ICCV98*, pages 59–66, 1998.

[16] B. E. Ruttenberg and A. K. Singh. Indexing the earth mover's distance using normal distributions. *PVLDB*, 5(3):205–216, 2011.

[17] T. Seidl and H. Kriegel. Optimal multi-step k-nearest neighbor search. In *SIGMOD*, pages 154–165, 1998.

[18] H. T. Shen, B. C. Ooi, and X. Zhou. Towards effective indexing for very large video sequence database. In *SIGMOD*, pages 730–741, 2005.

[19] J. Strötgen, M. Gertz, and C. Junghans. An event-centric model for multilingual document similarity. In *ACM SIGIR*, pages 953–962, 2011.

[20] D. Uskat, T. Emrich et al. Similarity search in fuzzy object databases. In *SSDBM*, pages 32:1–32:6, 2015.

[21] M. S. Uysal, C. Beecks, J. Schmücking, and T. Seidl. Efficient filter approximation using the Earth Mover's Distance in very large multimedia databases with feature signatures. In *CIKM*, pages 979–988, 2014.

[22] M. S. Uysal, C. Beecks, J. Schmücking, and T. Seidl. Efficient Similarity Search in Scientific Databases with Feature Signatures. In *SSDBM*, p. 30:1–30:12, 2015.

[23] B. Vandersmissen et al. The rise of mobile and social short-form video: an in-depth measurement study of vine. In *SoMuS*, v. 1198, pages 1–10, 2014.

[24] M. Wichterich et al. Efficient emd-based similarity search in multimedia databases via flexible dimensionality reduction. In *SIGMOD*, pages 199–212, 2008.

[25] J. Xu, Z. Zhang et al. Efficient and effective similarity search over probabilistic data based on earth mover's distance. *PVLDB*, 3(1):758–769, 2010.

[26] YouTube. Statistics. Retrieved Sep. 1, 2015 from https://www.youtube.com/yt/press/statistics.html.

[27] X. Zhou, X. Zhou, and H. T. Shen. Efficient similarity search by summarization in large video database. In *ADC*, pages 161–167, 2007.