Approximation Techniques for Indexing the Earth Mover's Distance in Multimedia Databases

Ira Assent[1], Andrea Wenning[2], Thomas Seidl[1]
[1]Data Management and Exploration Group, RWTH Aachen University, Germany
[2]EXAConsult GmbH, Hückelhoven, Germany
{assent,seidl}@cs.rwth-aachen.de, a.wenning@exaconsult.de

Abstract

Todays abundance of storage coupled with digital technologies in virtually any scientific or commercial application such as medical and biological imaging or music archives deal with tremendous quantities of images, videos or audio files stored in large multimedia databases. For content-based data mining and retrieval purposes suitable similarity models are crucial. The Earth Mover's Distance was introduced in Computer Vision to better approach human perceptual similarities. Its computation, however, is too complex for usage in interactive multimedia database scenarios. In order to enable efficient query processing in large databases, we propose an index-supported multistep algorithm. We therefore develop new lower bounding approximation techniques for the Earth Mover's Distance which satisfy high quality criteria including completeness (no false drops), index-suitability and fast computation. We demonstrate the efficiency of our approach in extensive experiments on large image databases.

1 Introduction

There is tremendous growth in data produced and filed by applications in medicine, engineering, biology and numerous others. This is due to new technologies such as computer tomography as well as decreasing storage prices which allows filing of the data produced. These huge amounts of data call for efficient strategies for storing and retrieving multimedia data.

There has been substantial progress, especially in the area of content-based image retrieval (CBIR) where color images are retrieved from multimedia databases using three main concepts: first, a feature extraction model, which maps each image to a set of attributes, the so-called features. Secondly, a distance measure between these features which reflects the underlying similarity model. And finally, storage and retrieval methods for large image databases.

Several models have been proposed for the feature extraction process. The histogram approach, which maps each image to a vector of attributes encoding e.g. color distribution [15, 27, 28], texture [21, 22], shape [2], has proven its usefulness in numerous applications and CBIR systems such as IBM's QBIC [10].

Comparing image histograms using L_p-norms such as the Euclidean distance is very sensitive to minor shifts in feature value distribution. Quadratic Forms use a ground similarity between feature bins to milden this sensitivity (cf. [10]). A recent approach from Computer Vision towards human similarity perception, the Earth Mover's Distance (EMD) [25] models similarity as the amount of changes necessary to transform one image feature into another one.

The quality of the Earth Mover's Distance for image similarity has been demonstrated in several application scenarios: color [17, 25], contour matching [12], texture [20], melodies [30, 31], graphs [8], vector fields [19], etc. Its high quality for image similarity search has been demonstrated in an extensive survey on distances for CBIR [22]. To benefit from the EMD's high quality retrieval in these settings, efficient storage and query processing is crucial.

Formally, the Earth Mover's Distance is defined as a Linear Programming problem which can be solved using the simplex method as in [16]. It is far too complex, however, for application in large multimedia databases.

To cope with this problem, we propose a database indexing and multistep approach similar to the GEMINI framework [11]. Related works following this paradigm have been successfully applied for multimedia data such as time series [6, 11] or ellipsoid queries [1, 28]. For efficient multistep query processing, which uses an approximative filter function to reduce the computation times of the original distance function, a tight and computationally simple approximation of the Earth Mover's Distance is essential. Our filters are motivated by geometric intuition of the Earth Mover's Distance and possible lower bounding distances. This leads to valuable insights on the nature of the EMD

computation, which is used to develop a more complex, but more selective filter suitable for higher dimensionalities. The analysis of our filter properties such as index applicability and efficiency are used to conceptualise a multistep algorithm which combines the advantages of the respective filters. We proof the completeness of our approach to guarantee no false drops in image retrieval. Our experiments on high-dimensional color histograms in large image databases validate our findings.

The Earth Mover's Distance is flexible in adapting to various applications by employing different ground distances encoded in the cost matrix. It can also be used in a number of extensions for additional scenarios such as partial matching (losing its metric property) or generalized histograms, i.e. so-called signatures encoding image pixel clusters. In this paper we concentrate on classical histograms and metric ground distances as they are standard representation in color image retrieval.

This paper is structured as follows: In Section 2 we establish the notions and properties of the Earth Mover's Distance needed for our technique. We then explain possible multistep algorithm concepts and their respective suitability in Section 3. Section 4 presents several novel approximations of the Earth Mover's Distance along with an analysis of their properties and proofs of completeness. In Section 5 we give an extensive experimental evaluation of our multistep approach using real world color image databases. In the conclusions we summarize our proposal and outline perspective future work.

2 The Earth Mover's Distance

The Earth Mover's Distance was introduced in the computer vision community in 1997 by Rubner, Guibas and Tomasi [25, 24] to overcome inconsistencies with perceptual similarity observed in (weighted) L_p-norms and quadratic forms. To understand these difficulties, let us recall how these distance measures are defined.

L_p norms are defined for two histograms $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, where x_i and y_i denote the bin entries, as follows: $L_p(x, y) = \sqrt[p]{\sum_{i=1}^{n} (x_i - y_i)^p}$. From this formula we immediately see that only corresponding 'bins' of the histograms are compared, ignoring relationships between 'neighboring' bins. To see what this means, consider three images and their corresponding color histograms as in Figure 1 (based on [24]). The top left one and the lower left one are different only by a slight shift in color tone. The lower right one, however, exhibits a completely different color distribution. A comparison using (weighted) L_p norms does not accurately reflect this. As difference is calculated bin-by-bin, i.e. only entries corresponding to the exact same color tones are compared, the distance between the left two images is larger than between the two ones on the right. Quadratic forms were introduced

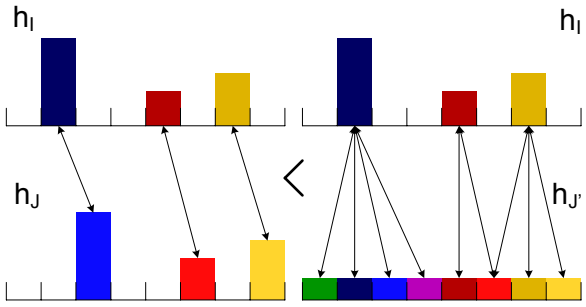

Figure 1. The Earth Mover's Distance

to overcome these difficulties [15, 27]. A similarity matrix $A = [a_{ij}]$ is used to reflect the perceived color distances between bin i and bin j. Formally, $QF_A(x, y) = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - y_i) a_{ij} (x_j - y_j)}$. This means, that all differences between arbitrary bins i and j influence the resulting distance measure weighted by a_{ij}. Reconsidering our example in Figure 1, depending on the similarity matrix A, the differences are merely 'smoothed' over all bins, meaning that still structural differences in images cannot be distinguished from color shifts.

The Earth Mover's Distance also uses a cost matrix $C = [c_{ij}]$ to describe distances between bins. These matrix entries c_{ij} reflect the perceived dissimilarity between the bins i and j which is called ground distance. A best match between the two images represented through their histograms x and y is then calculated by transferring all histogram entries in x to those in y at minimal cost. Technically, this means that matrix entries do not merely weigh the differences of bin entries as in quadratic forms, but instead guide the distance calculation. An assignment of one histogram entry x_i to another y_j is determined based on the cost of their corresponding matrix element c_{ij}. This can be formalized as the search for a minimum in a linear programming scheme.

Formally, the Earth Mover's Distance between histograms x and y with respect to a cost matrix $C = [c_{ij}]$ is defined as the minimum flow between x and y as follows:

$$EMD_C(x, y) = \min_{i,j} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{c_{ij}}{m} f_{ij}, f_{ij} \geq 0, \right.$$
$$\left. \sum_{j=1}^{n} f_{ij} = x_i, \sum_{i=1}^{n} f_{ij} = y_j \right\}$$

The denominator $m := \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}$ normalizes the EMD by the total flow, i.e. the mass of the histograms. When defined this way, the EMD is metric als long as the ground distance is metric.

This means, that the Earth Mover's Distance measures the effort to match one histogram against the other. All entries in one histogram need to be transferred to entries in the other at minimal cost. Similar images will only cause minor effort as there will be little differences in bin entries. Reconsidering Figure 1, we see that using an adequate cost matrix will indeed lead to a larger difference for the rightmost images as shifting all bin entries by one is less effort than moving some of the entries to entirely different bins.

Figure 2 gives an example of the EMD shape in a two-dimensional feature space. An exemplary iso-contour, i.e. points which have the same EMD-distance value from the center point, is highlighted. Additionally, all image pixels are colored according to their distance, where smaller distances are visualized using darker gray tones. We can see that the EMD is bounded by hyperplanes, thus a good filter approximation should describe a closely surrounding geometry. As we will see in Section 4.2, this is a good starting point for lower bounds of the EMD.

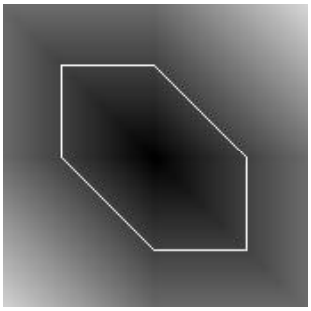

Figure 2. Iso-lines of the EMD

The Earth Mover's Distance can be calculated using a streamlined linear programming scheme, the simplex algorithm as found in numerical mathematics literature (see e.g. [16]). While this provides a solution for pairwise comparison of two histograms, it is much too expensive for large databases where numerous objects have to be compared. We therefore use a multistep algorithm to reduce expensive EMD calculations while guaranteeing correctness and completeness of the result.

3 Multistep Retrieval

The need for speeding up retrieval algorithms has brought forth several approaches for avoiding unnecessary distance calculations on images which are just 'too far away' by filtering them out in advance. This is crucial for interactive similarity retrieval applications.

3.1 Direct index usage vs. Multistep retrieval

Index structures are used to organize the data with the objective that only a small percentage needs to be accessed during query processing. They can be used in any setting where only distances between objects are given (e.g. M-tree [7]). Performance, however, is much better in a (multi-)dimensional space, where index structures such as R-trees or X-trees [3, 13] can be used, since they do not only rely on distances to index the data. Additionally, information on individual dimensions in feature space is exploited.

Index usage can be even more accelerated by calculating faster approximate distances on the index; thus generating candidate results which are then evaluated with the original distance to obtain the true results. Such a multistep approach has been successfully employed in several application scenarios: time series [6, 11], tumor shapes [18], GIS (geographic information systems) [5], ellipsoid queries [1, 28], and dimensionality reduction in general [9].

The efficiency of multistep retrieval depends on the chosen filter distance which has to be tailored for the original distance function. For tumor shapes, the max morphological distance was approximated by a more efficient max granulometric filter distance [18]; for time series a dimensionality-reduced Chebyshev coefficient metric is shown to lower bound the original high-dimensional Euclidean distance [6].

The general idea of multistep retrieval algorithms is summarized in Figure 3. A query is first executed as an approximative query using a suitable index structure. This filter step, based on an approximate filter distance function, generates a set of candidates which is then evaluated using the real distance function to retrieve the desired result from the database.

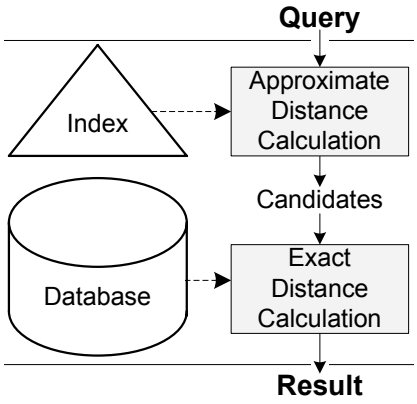

Figure 3. Multistep Retrieval

3.2 Query processing

Range queries are a common query type in multimedia databases to access objects similar to an example object. A range query is defined as the set of all objects from the database DB who have a distance to a query object q of less than or equal to ϵ. Formally: $S_\epsilon(q) = \{i \in DB \mid dist_{exact}(i,q) \leq \epsilon\}$. Range queries can be accelerated using the GEMINI multistep retrieval techniques [9, 11]. An index is queried using the same ϵ threshold, but with an appropriate filter distance: $Cand_\epsilon(q) = \{i \in DB \mid dist_{filter}(i,q) \leq \epsilon\}$. This candidate set is then evaluated to see if the actual object distance is not larger than ϵ. In real world applications, it is often not easy to specify a suitable ϵ threshold. If ϵ is chosen too small, the result set may be empty; if it is chosen too large, there may be too many results - in a worst case scenario the whole database may be output.

This motivates k-Nearest Neighbor (k-NN) queries, which do not require the user to specify such an ϵ. To a query object q the database DB is queried for those k objects which have the smallest distance from q. That is, $S_k(q)$ is the smallest subset of the database DB with $\forall i \in S_k(q)\, \forall j \in DB - S_k(q) : dist_{exact}(i,q) \leq dist_{exact}(j,q)$.

In the GEMINI framework, the index is first queried using a filter distance to retrieve k candidates. The distance ϵ' of the k-th farthest candidate object is then used to determine the actual k-NN result by a range query. The authors show that, provided the lower bounding property holds for the distance function and the respective filter distance, the resulting output is indeed correct and complete.

An optimal competing multistep algorithm for k-NN queries which makes use of the distance information from those candidates already filtered out was devised in previous work [29]. First, candidates are generated using a ranking procedure on the index based on the filter distance. Here, however, as they are generated, for each candidate object the exact distance to the query q is computed and the ϵ' range is lowered accordingly. This guarantees optimality as to the number of candidates generated while still being correct and complete.

3.3 Filter properties

Crucial for the efficiency of multistep algorithms is the quality of the filter distance measure. This measure should meet a number of criteria:

Completeness means that the multistep algorithm does not produce false drops, i.e. in the filter step no actual result object should be discarded. For all of the algorithms revised in Section 3.2, completeness can be assured by a corresponding lower bounding lemma which adheres the following observation:

Completeness of lower bound filters

If a filter distance $dist_{filter}$ lower bounds the actual object distance $dist_{exact}$, i.e. $dist_{filter}(x,y) \leq dist_{exact}(x,y)$ for all $x, y \in DB$ then the multistep algorithms as described in Section 3.2 guarantee completeness.

Proofs: see [11, 18, 29].

Good selectivity means that the filter distance should produce a candidate set as small as possible. In a worst case scenario, i.e. the filter distance returns zero for all object distances, the filter simply returns the entire database. Good selectivity is achieved by a filter which approximates the true distance as closely as possible, i.e. return preferably large values without violating the lower bounding property.

Fast single pair computation is important for acceleration of the total response time. In larger databases, numerous filter distance computations are necessary, so efficiency is important for performance of the first search step.

Index enabled: multistep algorithms are generally capable of working on an indexing structure which allows the search process to be guided according to the structure of the index storage. A filter distance should thus be capable of being applied in such a scenario in order to benefit from index support.

4 Lower bounds of the EMD

We will present several new lower bounds for the EMD in this section. Before doing so, we mention a lower bound given in the original EMD paper.

4.1 3D-Averaging lower bound

We first describe a lower bound introduced by Rubner et al. [26]. The basic idea is to average the features in the underlying color space. Let r_i denote the bin representatives (centroids), then the 3D-color averages, a weighted sum of the bin representatives, form a simple lower bound for the EMD:

$$EMD(x,y) \geq \|\sum_{i=0}^{n} \frac{x_i r_i}{m} - \sum_{i=0}^{n} \frac{y_i r_i}{m}\|$$

As shown in [26], this lower bound does help improve the computation time in smaller settings (i.e. in their case 8-dimensional features in a 20,000 color image collection). Since this lower bound (which we will denote by LB_{Avg} is fixed to three dimensions by the color space, there is no flexibility in adapting it to other dimensionalities. As shown in the experiments, the efficiency gains are thus not sufficient for high-dimensional settings. (Section 5).

4.2 L_p-norm-based lower bounds

Our first approach to develop lower bounding filter distances follows the geometric intuition which has been successfully applied to other complex distance functions including quadratic forms [1, 5]. Common conservative approximations such as minimum bounding rectangles or minimum bounding spheres are well suited for spatial objects and, analogously, for similarity range queries that have a rigid geometry bounded by the range parameter ϵ.

For k-nearest neighbor queries, however, the boundary of the query region is not known in advance when query processing starts. What has to be provided as approximations are (unbounded) distance functions instead. The corresponding distance functions for rectangles and spheres are represented by weighted maximum norms (L_∞) and weighted Euclidean norms (L_2), respectively. More precisely, by our geometric approach we aim at providing the lower bounding weighted L_p norms ($p = 1, 2, \infty$) as filter distances.

The advantages of using (weighted) L_p norms are twofold. First, a single comparison is computed in linear time O(n) in an n dimensional histogram space. Secondly, these distance functions are immediately index enabled and supported by any available multi-dimensional indexing structure.

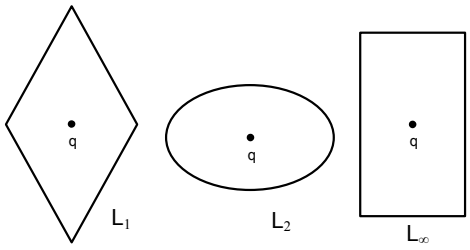

Figure 4. Iso-contours of weighted L_p norms

Geometrically, the iso-surfaces of the weighted L_p norms correspond to rectangles, spheres, and diamonds (Figure 4). Since the EMD iso-surface is bound by hyperplanes as can be observed from Figure 2, a hyperdiamond (L_1) or hyperrectangle (L_∞) is expected to result in a better approximation than a hypersphere (L_2).

4.3 Weighted L_1 lower bound

As discussed in the Introduction, (weighted) L_p norms base their distance calculation on the difference between individual bin entries. The weighted L_1 (Manhattan) distance is defined by $WL_1(x, y) = \sum_{i=1}^{n} w_i|x_i - y_i|$ where the w_i represent weights for the individual dimensions. These weights have to be determined as parameters of the filter. Geometrically, they stretch or compress the corresponding hyperdiamond to optimally enclose the EMD-iso-surface.

To determine weights which best lower bound the EMD, let us take a closer look at how the EMD is computed. Since we only consider metric ground distances, in our costmatrix C, the diagonal entries c_{ii} are zero (definiteness) and $c_{ij} = c_{ji}$ (symmetry). When searching for the minimal flow as described in Section 2, as much mass as possible will be moved at zero cost, i.e. between corresponding bin entries. The amount of mass which can be moved is limited by the corresponding entries in each histogram as defined in Section 2: $m = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij} = \sum_{i=1}^{n} x_i = \sum_{j=1}^{n} y_j$. Figure 5 illustrates the computation procedure between two

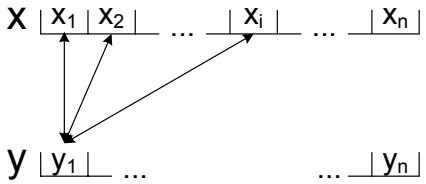

Figure 5. Flows to and from histograms

histograms $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. Let us assume that $x_i \geq y_i$. As mentioned above, zero cost c_{ii} is the minimum, i.e. first, earth of x_i will be moved to y_i until y_i is filled. If there is more earth available in x_i, it cannot be absorbed by y_i. The remaining mass ($|x_i - y_i|$) will thus have to be moved at higher cost. If, on the contrary, y_i is larger than x_i all of x_i will be moved to y_i and the remaining capacity of y_i ($|x_i - y_i|$) will be filled at higher cost by other entries of x. This leads to the following theorem:

Theorem *Weighted L_1 lower bound LB_{Man}*
For any metric ground distance encoded in a costmatrix $C = [c_{ij}]$, the EMD between two histograms $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ with the same total mass, i.e. $m = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$, is always greater than or equal to the weighted Manhattan distance with weights $w_i = min_{j, i\neq j}\{\frac{c_{ij}}{2 \cdot m}\}$. Formally:

$$EMD(x, y) \geq \sum_{i=1}^{n} \min_{\substack{j \\ i\neq j}}\{\frac{c_{ij}}{2 \cdot m}\}|x_i - y_i|$$

Proof:

$$EMD(x, y) = \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} \frac{c_{ij}}{m} f_{ij} \quad (1)$$

$$\geq \sum_{i=1}^{n} \min_{\substack{j \\ i\neq j}}\{\frac{c_{ij}}{m}\}\sum_{\substack{j=1 \\ i\neq j}}^{n} f_{ij} \quad (2)$$

$$\geq \sum_{i=1}^{n} \min_{\substack{j \\ i\neq j}}\{\frac{c_{ij}}{2 \cdot m}\}|x_i - y_i| \quad (3)$$

By definition, the EMD between histograms x and y is the minimal transfer of entries in x to those in y (or vice versa). As described above, entries between corresponding bins are moved at zero cost ($c_{ii} = 0$) which does not contribute to the EMD value. We therefore consider only the remainder $|x_i - y_i|$ which has to be moved to different bins $j \neq i$ (Equation 1).

Approximating each of the cost entries c_{ij} in one row i by their minimum allows us to combine the individual coefficients into a single one for the entire sum (Equation 2).

As seen before, the sum of flows from any bin x_i equals its mass: $\sum_{j=1}^{n} f_{ij} = x_i$. Since $c_{ii} = 0$ is always minimal, f_{ii} is as large as possible, i.e. $f_{ii} = y_i$ if $x_i > y_i$ (y_i cannot absorb more). Thus $\sum_{j=1, i\neq j}^{n} f_{ij} = x_i - y_i$ if $x_i > y_i$. Analogously, $\sum_{j=1, i\neq j}^{n} f_{ij} = y_i - x_i$ if $x_i < y_i$. In sum, $2\sum_{j=1, i\neq j}^{n} f_{ij} = |x_i - y_i|$. Thus, we replace the sum of flows by half the absolute bin differences (Equation 3). $\Diamond$

Thus, our weighted Manhattan distance LB_{Man} lower bounds the EMD which guarantees completeness in a multistep retrieval scheme as outlined in Section 3.

4.4 Weighted L_∞ lower bound

For the minimum bounding rectangle approximation, the corresponding filter distance function is represented by a weighted maximum norm (L_∞): $WL_\infty(x, y) = \max_{i=1}^{n}\{w_i|x_i - y_i|\}$. Geometrically, this norm corresponds to a rectilinear hyperrectangle. By similar arguments as in the L_1 case, we obtain:

Theorem *Weighted L_∞ lower bound LB_{Max}*
For any metric ground distance encoded in a costmatrix $C = [c_{ij}]$, histograms $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ with $m = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$, we have:

$$EMD(x, y) \geq \max_i\{\min_{\substack{j \\ i\neq j}}\{\frac{c_{ij}}{m}\}|x_i - y_i|\}$$

Proof:

$$EMD(x, y) \geq \sum_{\substack{i=1 \\ x_i \leq y_i}}^{n} \min_j\{\frac{c_{ij}}{m}\}|x_i - y_i| \quad (4)$$

$$\geq \max_{\substack{i \\ x_i \leq y_i}}\{\min_{\substack{j \\ i\neq j}}\{\frac{c_{ij}}{m}\}|x_i - y_i|\} \quad (5)$$

$$EMD(x, y) \geq \max_i\{\min_{\substack{j \\ i\neq j}}\{\frac{c_{ij}}{m}\}|x_i - y_i|\} \quad (6)$$

The proof of our weighted L_1 norm implies that considering only those cases where $x_i \leq y_i$, we obtain Equation 4. We approximate the sum by its maximum summand, thus yielding Equation 5.

Instead of approximating the EMD by those cases where $x_i \leq y_i$, we could also do so by those cases where $x_i \geq y_i$. As both results lower bound the EMD, we can choose the larger one of them by simply taking the maximum (Equation 6). $\Diamond$

4.5 Weighted L_2 lower bound

Another frequently used L_p norm is the L_2 norm, the Euclidean Distance: $WL_2(x, y) = \sqrt{\sum_{i=1}^{n} w_i(x_i - y_i)^2}$. Geometrically, the iso-surface of the Euclidean Distance is a (hyper-)sphere, if weighted, we obtain iso-oriented ellipsoids.

Theorem *Weighted L_2 lower bound LB_{Eucl}*
For any metric ground distance encoded in a costmatrix $C = [c_{ij}]$, histograms $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ with $m = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$, we have:

$$EMD(x, y) \geq \sqrt{\sum_{i=1}^{n}(\min_{\substack{j \\ i\neq j}}\{\frac{c_{ij}}{2 \cdot m}\})^2(x_i - y_i)^2}$$

Proof:

$$EMD(x, y) \geq \sum_{i=1}^{n} \min_{\substack{j \\ i\neq j}}\{\frac{c_{ij}}{2 \cdot m}\}|x_i - y_i| \quad (7)$$

$$= \sum_{i=1}^{n} \sqrt{(\min_{\substack{j \\ i\neq j}}\{\frac{c_{ij}}{2 \cdot m}\})^2(x_i - y_i)^2} \quad (8)$$

$$\geq \sqrt{\sum_{i=1}^{n}(\min_{\substack{j \\ i\neq j}}\{\frac{c_{ij}}{2 \cdot m}\})^2(x_i - y_i)^2} \quad (9)$$

Equation 7 was proven for the weighted Manhattan lower bound. We then replace the absolute value by the root of the squared value (Equation 8). Then the sum can be pulled in the root, which is done in Equation 9. $\Diamond$

We already see from this proof that the Euclidean lower bound encloses the Manhattan lower bound and is thus of worse selectivity if used in a filter step.

While lower bounds based on L_p norms are well suited for index support and can be efficiently computed, their selectivity decreases for higher dimensions due to the so-called "curse of dimensionality". Whenever the number of bins, and therefore, the number of cost matrix entries, increases, the minimum of these is bound to decrease. Therefore, in high-dimensional settings, the selectivity of bounds based on these minima drops. This is also validated in the experiments in Section 5.

Moreover, index structures fall prey to the curse of dimensionality as well. In high-dimensional settings, sequential scan eventually performs better than any index structure [4, 32]. In these high-dimensional settings, additional techniques are needed for efficient query processing. In the

next section, we will refine the idea gained from the geometrically inspired weighted L_p lower bounds. Computing for each dimension the amount of mass which has to be moved by at least the minimum cost for that dimension can be extended to a tighter, i.e. more selective, lower bound. After that, we will show how these lower bounds can be combined to best benefit from their respective efficiency improvements.

4.6 Independent minimization lower bound

As stated above, we reuse the concept of independently minimizing the cost for moving the mass in individual histogram bins. In contrast to the weighted L_p norms however, we do not stop at picking one minimum for each histogram dimensions. Instead, we re-approach the essential EMD method by taking into consideration that histogram bins can only absorb the mass corresponding to their respective entries ($\sum_{i=1}^{n} f_{ij} = y_j$). That is, for each dimension, we reinstall the constraint that each histogram bin should not receive more mass than specified by its value ($f_{ij} \leq y_j$). The difference to the EMD remaining is then, that this constraint is weakened to the effect that the sum of flows for different dimensions might well be above this limit. To understand this, recall the EMD definition:

$$EMD(x,y) = \min_{i,j} \left\{ \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{c_{ij}}{m} f_{ij}, f_{ij} \geq 0, \sum_{j=1}^{n} f_{ij} = x_i, \sum_{i=1}^{n} f_{ij} = y_j \right\}$$

The Independent Minimization lower bound (LB_{IM}) is then defined as follows:

$$LB_{IM}(x,y) = \min_{i,j} \left\{ \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{c_{ij}}{m} f_{ij}, f_{ij} \geq 0, \sum_{j=1}^{n} f_{ij} = x_i, f_{ij} \leq y_j \right\}$$

This formalizes that the difference between the new LB_{IM} and the Earth Mover's Distance is the constraint on how much mass one histogram bin may receive. While the EMD requires that the sum of flows to a certain bin equals its mass over all dimensions, the LB_{IM} only ensures that for any single dimension the incoming flows do not exceed its mass. We now show that LB_{IM} indeed lower bounds the EMD.

Theorem *LB_{IM} lower bound*
For any metric ground distance encoded in a costmatrix $C = [c_{ij}]$, histograms $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ with $m = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$, we have:

$$EMD(x,y) \geq LB_{IM}(x,y).$$

Proof:
We start by proving that the constraints in LB_{IM} are indeed a weaker version of those for the EMD, i.e. that for any i:

$$\sum_{i=1}^{n} f_{ij} = y_j \wedge f_{ij} \geq 0 \Rightarrow \; f_{ij} \leq y_j$$

Assume to the contrary that there exists a j such that $f_{ij} > y_j$. Then we have for the sum $\sum_{i=1}^{n} f_{ij} > y_j$ as well, which is a contradiction.
From this we infer for the two sets to be minimized: $S_{EMD} :=$

$$\left\{ \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{c_{ij}}{m} f_{ij}, f_{ij} \geq 0, \sum_{j=1}^{n} f_{ij} = x_i, f_{ij} \leq y_j \right\} \subset \left\{ \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{c_{ij}}{m} f_{ij}, f_{ij} \geq 0, \sum_{j=1}^{n} f_{ij} = x_i, \sum_{i=1}^{n} f_{ij} = y_j \right\} := S_{LB_{IM}}$$

Next, note that for any two sets, where one is a subset of the other, the minimum of one set is always larger than or equal to that of its superset: $s \subset S \Rightarrow$

$$min(S) = min(min(s), min(S \setminus s)) \leq min(s)$$

Consequently, the minimum determined as the Earth Mover's Distance is larger than or equal to that determined as the Independent Minimization lower bound:

$$LB_{IM}(x,y) = min(S_{LB_{IM}}) \leq min(S_{EMD}) = EMD(x,y)$$

$\Diamond$

Intuitively speaking, the proof uses the fact that the EMD is minimized in a more restricted space than the LB_{IM}. The minimum as specified by the EMD constraints is thus always included in the search space of the LB_{IM} which can therefore only result in the same or a smaller minimum.

We have shown that the Independent Minimization indeed lower bounds the EMD and is thus a complete filter. We next analyze how it can be efficiently computed. By restricting the search for a minimum by constraints over the outgoing dimensions i only, we can decompose the LB_{IM} computation into a series of smaller problems for each i. In other words, each of the three constraints, i.e.

- positivity of the flows ($f_{ij} \geq 0$),
- equality of outgoing flows to the mass of each bin ($\sum_j f_{ij} = x_i$) and

- limitation of flows to the target bins mass ($f_{ij} \leq y_j$)

can be checked for individual dimensions. The positivity of all flows f_{ij} is a constraint which does not require the consideration of more than one dimension i at a time. The same is true for the sum over all outgoing flows $\sum_j f_{ij}$: since the constraint needs to be validated for individual x_i, a separated computation yields the same result. And the last constraint on flows f_{ij} to the target bins y_j has been "individualized": the sum of flows to a target bin does no longer have to equal its mass as in the EMD case ($\sum_i f_{ij} = y_j$). Instead, it merely limits the amount of mass for each i ($f_{ij} \leq y_j$). Note that this last constraint is violated by the L_p-norm-based lower bounds, where no check on the receiving bins except the corresponding one ($f_{jj} \leq y_j$) is required; the remainder ($|x_i - y_i|$ is moved at minimal cost $c_{ij}, i \neq j$ independent of the mass in $y_j, i \neq j$). Instead of minimizing flows for all bins simultaneously via the simplex method, we are now able to minimize the cost for each histogram bin separately. This substantially decreases computation times as demonstrated in Section 5.

We can further improve the Independent Minimization lower bounds by adapting two properties seen for L_p-norm-based lower bounds:

(i) Restricting the flows to $f_{jj} + f_{ij} \leq y_j, i \neq j$, instead of simply to $f_{ij} \leq y_j$, can be easily fulfilled before minimization starts: we know that the flow between identical histogram bins is the mass of the smaller entry: $f_{jj} = min(x_j, y_j)$. Since $c_{jj} = 0$, we can subtract these flows f_{jj} from y_j as they do not contribute to the EMD value to improve the selectivity of LB_{IM}. (Cf. the proof of LB_{Man}, Section 4.3).

Let $x = [4, 3, 5, 4, 5]$ and $y = [1, 2, 3, 8, 8]$, compute $x^1 = [x_i - min(x_i, y_i)] = [3, 2, 2, 0, 0]$ and $y^1 = [y_i - min(x_i, y_i)] = [0, 0, 0, 4, 3]$. These histograms have the same EMD value as the original ones, but further restrict for each bin the possibility of weight movement (i.e. the limit y_j is lowered), which results in a better selectivity of our LB_{IM} filter.

(ii) As for the LB_{Max} lower bound (Section 4.4), we can pick the maximum of $LB_{IM}(x, y)$ and $LB_{IM}(y, x)$. Switching the roles of x and y means relaxing the constraint $\sum_{j=1}^{n} f_{ij} = x_i$ to $f_{ij} \leq x_i$ instead of relaxing $\sum_{i=1}^{n} f_{ij} = y_j$ to $f_{ij} \leq y_j$. The lower bounding property is fulfilled in either case, thus we choose the larger, more selective, lower bound, i.e. the maximum of $LB_{IM}(x, y)$ and $LB_{IM}(x, y)$.

Assume $C = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \end{pmatrix}$, Independent Minimization yields for the above x^1 to y^1: $LB_{IM}(x, y) = x_1 * c_{14} + x_2 * c_{24} + x_3 * c_{34} = 3*3+2*2+2*1 = 15$, whereas the computation from y to x: $LB_{IM}(y, x) = 1/2 * y_4 * c_{43} * 1/2 * y_4 + c_{42} + 2/3 * y_5 * c_{53} + 1/3 * y_5 * c_{52} = 2*1+2*2+2*2+1*3 = 13$. By taking the maximum over these two values, we are sure to obtain the LB_{IM} value closest to the EMD.

4.7 Multistep filter concept

We mentioned that L_p-norm-based filters are well suited for index support and have low one-to-one computation times. As they do not provide optimal selectivity for increasing dimensionality, they are best used for low-dimensional problems. As dimensionality increases, our LB_{IM} lower bound is favorable as it yields better selectivity which eventually means that overall computation times are lower in high dimensions where many more candidates might be generated otherwise.

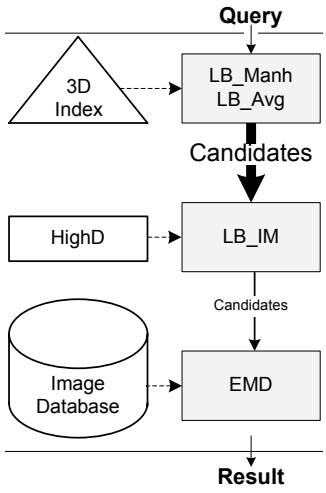

Figure 6. EMD Multistep Concept

The combination of filters as described in Section 3 further reduces the computation time as they dismiss different sets of objects as non-candidates. Therefore, a sequential scan using a good-selectivity-filter in high dimensions on the output of an efficient low-dimensional, index-supported filter makes best use of database technology available.

While index support can substantially improve response times, it has been shown many times, that with increasing dimensionality, starting as low as dimensions six or eight, indexing structures fall prey to the so-called "curse of dimensionality" [4, 32]. This means that with higher dimensionality, large portions of the index have to be accessed, which in turn means that many more I/O activities have to be performed.

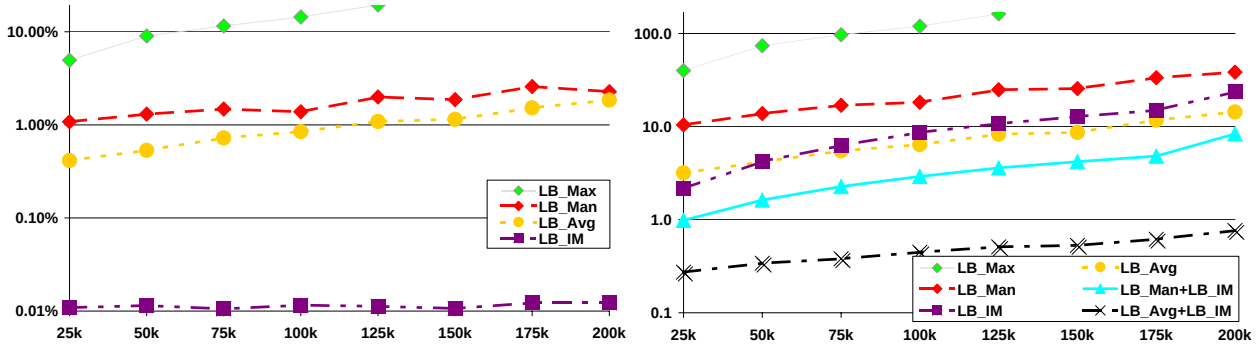

Figure 7. Scalability, selectivity percentage (left), response time in seconds (right)

We therefore propose to integrate a reduction of dimensionality into a two-phase multistep algorithm which combines the advantages of low-dimensional index-support with the good selectivity of our multi-dimensional LB_{IM} filter (Fig. 6). We therefore construct three-dimensional indexes based on the averaging lower bound (which can only handle three dimensions in a three-dimensional color space) or three-dimensional indexes based on the weighted Manhattan lower bound, which, as experiments confirm, is best among our L_p-based filters. Dimensionality reduction for the weighted Manhattan lower bound is performed by simply determining those three dimensions with highest variability and discarding all other dimensions. The resulting distances are necessarily smaller, thus, lower bounding is not jeopardized. Recall the definition of the weighted Manhattan distance from Section 4.3: $WL_1(x,y) = \sum_{i=1}^{n} w_i|x_i - y_i|$. By reducing this to three dimensions (e.g. the first ones) we have $WL_1(x^3,y^3) = \sum_{i=1}^{3} w_i|x_i - y_i|$. Since we only drop positive summands, we are sure to lower the resulting distance value. The three-dimensional index is then built on weighted reduced histograms $w^3 * x^3$.

Similarly, for the $LB_{Avg} = \| \sum_{i=1}^{n} \frac{x_i r_i}{m} - \sum_{i=1}^{n} \frac{y_i r_i}{m} \|$ lower bound, the color averages $a_i = \sum_{i=1}^{n} \frac{x_i r_i}{m}$ have to be precalculated to preserve lower bounding. The corresponding 3D-index is then built on these color averages. The resulting candidate set of the first 3d-filter may be large, but its generation involves relatively few index operations. The candidate set is then further reduced by using our LB_{IM} lower bound which has better selectivity. The final refinement computations using the Earth Mover's Distance are thus reduced as much as possible. As we will see in the experimental evaluation in Section 5 this two-phase multistep algorithm indeed outperforms all other filter techniques.

5 Experiments

We ran extensive experiments on an extended version (200,000 images) of the color image database used in [1]. We implemented all lower bounds and the Earth Mover's Distance in Java (the latter is based on the original C code by Rubner [23]). We used the R-tree index by Hadjieleftheriou [14] from the R-tree Portal. Experiments were run on Pentium 4 2.4 GHz workstations.

Using optimal query processing, we varied the number k of nearest neighbors from 1 to 20, histogram sizes from 16 to 64 dimensions, and tested the scalability of our approach by varying the database size. Using the average of 200 randomly selected query images, we recorded selectivity ratios as well as total response times for query processing.

Scalability. In our first experiment, we evaluated the performance of our lower bounding filters by measuring the selectivity for various database sizes. The smallest database contains 25.000 images, the largest all 200,000 images. We compared our most promising filter techniques for k-Nearest-Neighbor queries, where k was set to 10 and histograms were of highest resolution 64. The selectivity is then given as the average percentage of database images for which EMD computation is necessary.

The LB_{Eucl} weighted L_2 lower bound's performance is, as was already analyzed in Section 4.5, far inferior to LB_{Max}, i.e. weighted L_∞ lower bound and to LB_{Man}, i.e. weighted L_1, lower bound. We therefore did not include it in our diagrams.

In the left diagram in Figure 7 we can see that the weighted L_∞ lower bound LB_{Max} produces many more candidates than the weighted L_1 LB_{Man} or the averaging lower bound LB_{Avg}; its selectivity is inferior by an order of magnitude. This means, that the EMD is more closely surrounded by the hyperdiamond which corresponds to our weighted L_1 lower bound than by the hyperrectangle belonging to our weighted L_∞ lower bound. As LB_{Max} has far worse performance, we no longer include it in our figures.

Our new filter, the LB_{IM} lower bound is of noticeably higher selectivity. It outputs far less than 0.1% of the data as candidates for all database sizes. This is an improvement by more than two orders of magnitude to the second-best lower

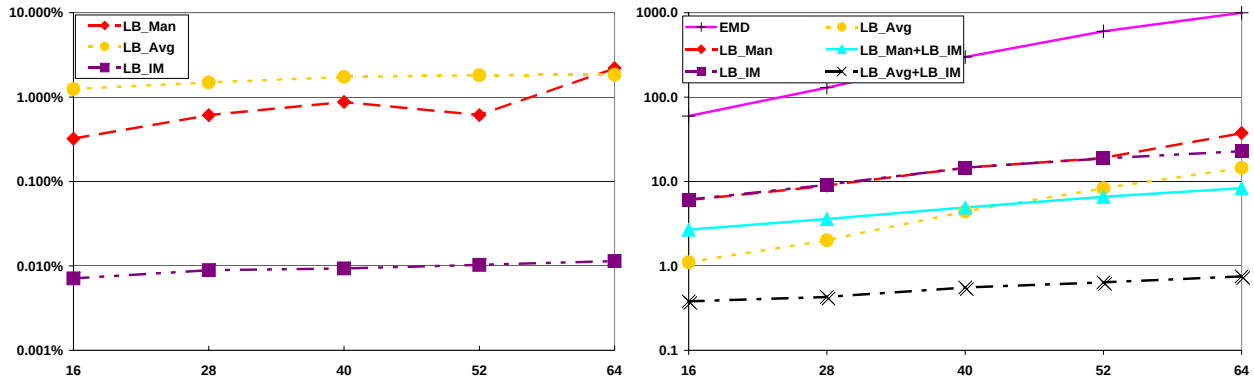

Figure 8. Dimensionality, selectivity percentage (left), response time in seconds (right)

bound LB_{Avg}. In the right part of Figure 7 we see that for the same setup, the response times of LB_{Man} and LB_{Avg} are closely related to their selectivity as their computational overhead is low. LB_{IM} requires more computational effort, thus despite its far superior selectivity, it shows only similar response times. However, the combination with the former two lower bounds as presented in Section 4.7, yields the lowest response times.

Dimensionality. Another important parameter for the performance of lower bounding approximations is the size of the histograms involved. We varied histograms from dimension 16 to 64. We can see in the left diagram in Figure 8, where the largest database with 200,000 was queried for $k = 10$ nearest neighbors, that our LB_{IM} lower bound yields clearly best selectivity results in all cases. For finer histogram resolutions, the LB_{IM} lower bound has selectivity gains of more than two orders of magnitude. The second best lower bound is the weighted L_1 lower bound, followed by the averaging lower bound.

The right diagram in Figure 8 for response times shows that with increasing dimensionality, the computation of LB_{Avg} increases in complexity. The response times of LB_{Man} are more closely related to its selectivity ratios. As in the previous experiment, the overhead of LB_{IM} is greater than that of the other two lower bounds. Once again, its combination with a low-dimensional LB_{Avg}-index yields the best performance improvements. We include the sequential scan EMD computation times as a baseline comparison. Note that the improvement for 64 dimensions comparing EMD and the best multistep concept is from 1000 seconds to less than one second, i.e. more than three orders of magnitude.

Result size. To determine the influence of the number of nearest neighbors queried for, we ran an experiment on the largest database of 200,000 images using 64-dimensional-histograms and searched for 1 to 20 similar images. We see in Figure 9 (left) that the weighted Manhattan lower bound LB_{Man} and the color averaging approach LB_{Avg} show similar selectivity ratios, which are far inferior to those of LB_{IM} lower bound by far. On the right we see that the response times show corresponding behavior, where the multistep concept of Section 4.7 is once again clearly the fastest.

Query processing. Query processing performance, as discussed in Section 3.2, is illustrated for two exemplary lower bounds in Figure 10. As we can see, optimal multistep query processing is noticeably more efficient than the original GEMINI approach, especially for the lower bound LB_{Man} where selectivity is improved by one order of magnitude. As the LB_{IM} lower bound already retrieves substantially fewer candidates, its selectivity gains are smaller, but still noticeable. The response times on the right reflect these selectivity ratios. This means that choosing the optimal algorithm results in efficient query processing.

In summary, our experiments demonstrate that the best strategy for query processing is a combination of assets. We compared computation times of two different approaches: LB_{Man}, LB_{Avg}, and LB_{IM} were tested in "simple" multistep query processing as presented in Section 3, which means that each of these filter functions is immediately followed by the exact EMD calculation. Their response times were only acceptable for low dimensions and non-interactive scenarios. We compared these three with the two proposed combinations of Section 4.7, where three-dimensional LB_{Man} and LB_{Avg} were evaluated in a first step. Their candidate set is then substantially reduced by the best-selectivity-filter, the LB_{IM}. As we can see, this concept indeed results in noticeable speed-up ratios. The high selectivity of LB_{IM} thus results in the expected efficiency improvements. By building a small three-dimensional index based on simpler filter functions and combining it with a highly selective second LB_{IM} filter, index support can be profited from while expensive Earth Mover's Distance computations are minimized.

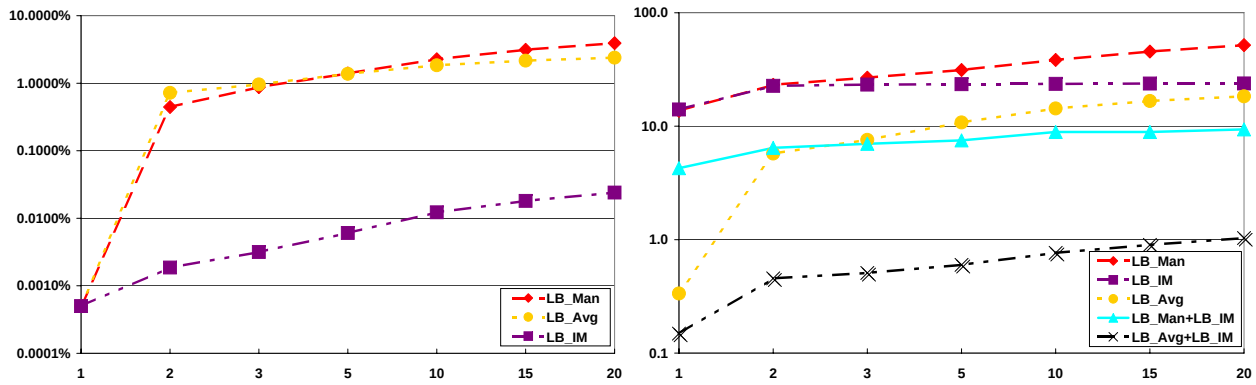

Figure 9. Result size, selectivity percentage (left), response time in seconds (right)

6 Conclusion

Search in large image or other multimedia databases highly dependends on the underlying similarity model. The Earth Mover's Distance, proposed in Computer Vision literature, is an interesting new approach towards achieving high-quality content-based retrieval. Despite its advantages, this distance measure is computed via a linear programming algorithm which is too slow for today's huge and interactive multimedia applications. We therefore developed several new filters designed for usage in a multistep retrieval concept. We analyzed and demonstrated in our experiments that L_p-norm-based filters are easy to compute, they can be used with indexing structures, and achieve satisfactory selectivity for smaller dimensions and non-interactive settings. As dimensionality increases, more complex filters with higher selectivity ratios are needed. We present such an approach which is indeed closer to the Earth Mover's Distance, the LB_{IM} lower bound. We combined our filters in a multistep algorithm making use of the advantages of the filters involved. High dimensions are problematic for most index structures; we therefore use a dimensionality reduction implicit in the averaging lower bound or explicit in our Manhattan reduction. As this still leaves us with far too many candidates, several, increasingly more selective, filters ensue to yield an efficient retrieval procedure. In future work, we plan to extend our multistep retrieval concept techniques from image data to video data. We will apply the powerful perception-oriented Earth Mover's Distance to video data. Since video is more complex, even more efficient retrieval techniques need to be developed.

7 Acknowledgements

The authors would like to thank Till Schulte-Coerne for help implementing the Earth Mover's Distance as well as visualisation tools.

We are grateful to Marc Wichterich for assistance in generating histogram features.

References

[1] Ankerst, M., Braunmüller, B., Kriegel, H.-P. and Seidl, T. Improving adaptable similarity query processing by using approximations. In *Proceedings of the 24th International Conference on Very Large Data Bases, New York, USA*, pages 206–217, 1998.

[2] Ankerst, M., Kastenmüller, G., Kriegel, H.-P. and Seidl, T. Nearest neighbor classification in 3d protein databases. In *Proceedings of the 7th International Conference on Intelligent Systems for Molecular Biology (ISMB)*, pages 34–43. Lecture Notes in Computer Science, Springer, 1999.

[3] Berchtold, S., Keim, D. and Kriegel, H.-P. The x-tree : An index structure for high-dimensional data. In *Proceedings of 22th International Conference on Very Large Data Bases, Mumbai (Bombay), India*, pages 28–39, 1996.

[4] C. Böhm, S. Berchtold, and D. A. Keim. Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases. *ACM Comput. Surv.*, 33(3):322–373, 2001.

[5] Brinkhoff, T., Kriegel, H.-P., Schneider, R. and Seeger, B. Multi-step processing of spatial joins. In *Proceedings of the 1994 ACM SIGMOD international conference on Management of data*, pages 197–208, 1994.

[6] Cai, Y. and Ng, R. Indexing spatio-temporal trajectories with chebyshev polynomials. In *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data*, pages 599–610, New York, 2004.

[7] Ciaccia, P., Patella, M. and Zezula, P. M-tree: An efficient access method for similarity search in metric spaces. In *Proceedings of 23rd International Conference on Very Large Data Bases, Athens, Greece*, pages 426–435, 1997.

[8] Demirci, M., Shokoufandeh, A., Dickinson, S., Keselman, Y. and Bretzner, L. Many-to-many feature matching using spherical coding of directed graphs. In *Computer Vision - ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic*. Lecture Notes in Computer Science, Springer, 2004.

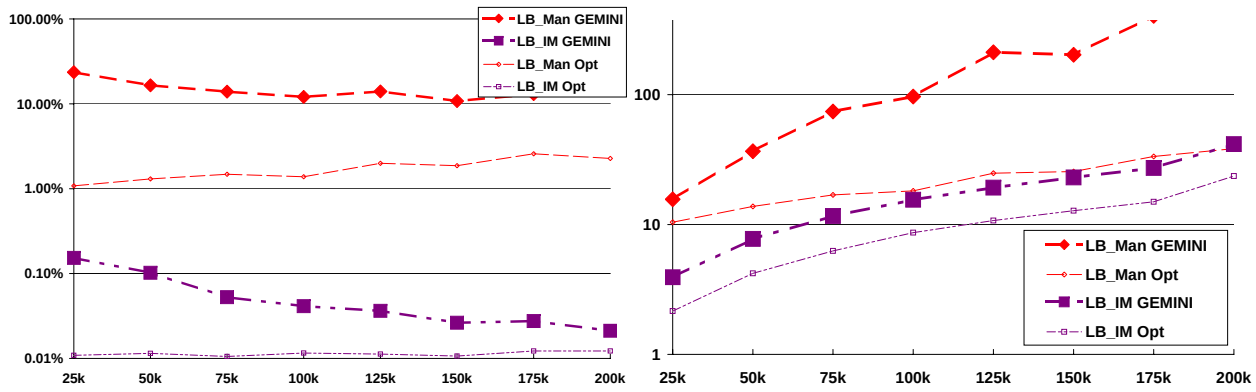

Figure 10. Multistep algorithm, selectivity percentage (left), response time in seconds (right)

[9] Faloutsos, C. *Searching Multimedia Databases by Content*. Kluwer Academic Publishers, 1996.

[10] Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D. and Equitz, W. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3:231–262, 1994.

[11] Faloutsos, C., Ranganathan, M. and Manolopoulos, Y. Fast subsequence matching in time-series databases. In *Proceedings 1994 ACM SIGMOD Conference, Mineapolis, MN*, pages 419–429, 1994.

[12] Grauman, K. and Darrell, T. Fast contour matching using approximate earth movers distance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[13] Guttman, A. R-trees: A dynamic index structure for spatial searching. In *Proceedings of the 1994 ACM SIGMOD international conference on Management of data, Boston, Massachusetts*, pages 47–57, 1984.

[14] Hadjieleftheriou, M. The R-tree-portal. *http://www.cs.ucr.edu/ marioh/spatialindex/*, 2002.

[15] Hafner, J., Sawhney, H., Equitz, W., Flickner, M. and Niblack, W. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):729–736, 1995.

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

[17] Jing, F., Li, M., Zhang, H., Zhang, B. An Efficient and Effective Region-Based Image Retrieval Framework. *IEEE Transactions on Image Processing*, 13(5):699–709, 2004.

[18] Korn, F., Sidiropoulos, N., Faloutsos, C., Siegel E. and Protopapas, Z. Fast nearest neighbor search in medical image databases. In *Proceedings of the 22th International Conference on Very Large Data Bases*, pages 215–226, 1996.

[19] Lavin, Y., Batra, R. and Hesselink, L. Feature comparisons of vector fields using earth mover's distance. In *Proceedings of the conference on Visualization*, pages 103–109, 1998.

[20] Lazebnik, S., Schmid, C. and Ponce, J. Sparse texture representation using affine-invariant neighborhoods. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[21] Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Petkovic, D., Yanker, P., Faloutsos, C., Taubin, G. and Heights, Y. Querying images by content, using color, texture, and shape. *SPIE Conference on Storage and Retrieval for Image and Video Databases*, 1908:173–187, April 1993.

[22] Puzicha, J., Buhmann, J., Rubner, Y. and Tomasi, C. Empirical evaluation of dissimilarity measures for color and texture. In *Proceedings of the International Conference on Computer Vision-Volume 2*, pages 1165–1173, 1999.

[23] Rubner, Y. Earth Mover's Distance source code in C. *http://www.cs.duke.edu/ tomasi/software/emd.htm*, 1998.

[24] Rubner, Y. and Tomasi, C. *Perceptual Metrics for Image Database Navigation*. Kluwer Academic Publishers, 2001.

[25] Rubner, Y., Guibas, L. and Tomasi, C. The Earth Mover's Distance, Multi-Dimensional Scaling, and color based image retrieval. In *Proceedings of the ARPA Image Understanding Workshop*, pages 661–668, 1997.

[26] Rubner, Y., Tomasi, C., and Guibas, L. A metric for distributions with applications to image databases. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 59–66, 1998.

[27] Sawhney, H. and Hafner, J. Efficient color histogram indexing. In *Proceedings of the IEEE International Conference on Image Processing*, pages 66–70, 1994.

[28] Seidl, T. and Kriegel, H.-P. Efficient user-adaptable similarity search in large multimedia databases. In *Proceedings of the 23rd International Conference on Very Large Data Bases, Athens, Greece*, pages 506–515, 1997.

[29] Seidl, T. and Kriegel, H.-P. Optimal multi-step k-nearest neighbor search. In *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, pages 154–165, 1998.

[30] Typke, R., Giannopoulos, P., Veltkamp, R., Wiering, F. and Oostrum, R. van. Using transportation distances for measuring melodic similarity. In *Proceedings of the Fourth International Conference on Music Information Retrieval (ISMIR)*, 2003.

[31] Typke, R., Veltkamp, R. and Wiering, F. Searching notated polyphonic music using transportation distances. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 128–135, 2004.

[32] R. Weber, H.-J. Schek, and S. Blott. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In *VLDB*, pages 194–205, 1998.