Fast and Robust Earth Mover's Distances

Ofir Pele
The Hebrew University of Jerusalem
ofirpele@cs.huji.ac.il

Michael Werman
The Hebrew University of Jerusalem
werman@cs.huji.ac.il

Abstract

We present a new algorithm for a robust family of Earth Mover's Distances - EMDs with thresholded ground distances. The algorithm transforms the flow-network of the EMD so that the number of edges is reduced by an order of magnitude. As a result, we compute the EMD by an order of magnitude faster than the original algorithm, which makes it possible to compute the EMD on large histograms and databases. In addition, we show that EMDs with thresholded ground distances have many desirable properties. First, they correspond to the way humans perceive distances. Second, they are robust to outlier noise and quantization effects. Third, they are metrics. Finally, experimental results on image retrieval show that thresholding the ground distance of the EMD improves both accuracy and speed.

1. Introduction

Histograms are ubiquitous tools in numerous computer vision tasks. It is common practice to use distances such as L_2 or χ^2 for comparing histograms. This practice assumes that the histogram domains are aligned. However this assumption is violated through quantization, shape deformation, light changes, etc.

The Earth Mover's Distance (EMD) [29] is a cross-bin distance that addresses this alignment problem. EMD is defined as the minimal cost that must be paid to transform one histogram[1] into the other, where there is a "ground distance" between the basic features that are aggregated into the histogram. The EMD as defined by Rubner is a metric only for normalized histograms. However, recently Pele and Werman [26] suggested $\widehat{EMD}$ and showed that it is a metric for all histograms.

A major issue that arises when using EMD is which ground distance to use for the basic features. This, of course, depends on the histograms, the task and practical considerations. In many cases we would like the distance to correspond to the way humans perceive distances (image retrieval, for example). In other cases we would like the distance to fit the distribution of the noise (keypoint matching, for example). Practical considerations include speed of computation and the metric property that enables fast algorithms for nearest neighbor searches [41, 7], fast clustering [10] and large margin classifiers [15, 36].

We propose using thresholded ground distances. *i.e.* distances that saturate to a constant value. These distances have many desirable properties. First, saturated distances correspond to the way humans perceive distances [34]. Second, many natural noise distributions have a heavy tail; *i.e.* outlier noise. Thresholded distances assign different outliers the same large distance. Finally, we present an algorithm that computes EMD with a thresholded ground distance faster by an order of magnitude than the original algorithm. The algorithm transforms the flow-network of the EMD so that the number of edges is reduced by an order of

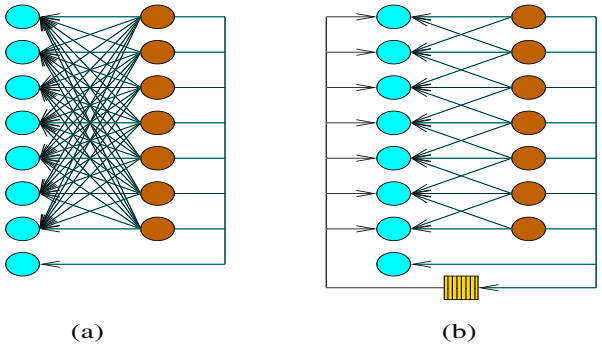

Figure 1. An example of the transformation on a flow network of an EMD or $\widehat{EMD}$ with a ground distance of: $d(a,b) = min(2, |a-b|)$. (a) is the original flow network with $N^2 + N$ edges. Note that $N(N-3)$ of these edges have cost 2. The bottom cyan vertex on the left is the sink that handles the difference between the total mass of the two histograms (ingoing edges cost is 0 for EMD and $\alpha \max_{ij} d_{ij}$ for $\widehat{EMD}$). (b) is the transformed flow network. The striped yellow square is the new transhipment vertex. Ingoing edge cost is the threshold (*e.g.* 2) and outgoing edge cost is 0.

[1] Rubner's noted that EMD can be used with sparse histograms which he coined *signatures*. Our algorithm is applicable to both histograms and signatures.

2009 IEEE 12th International Conference on Computer Vision (ICCV)
978-1-4244-4419-9/09/$25.00 ©2009 IEEE

magnitude (see Fig. 1).

The Earth Mover's Distance has been used successfully in many applications such as image retrieval [29, 23], edge and corner detection [30], keypoint matching [26, 8, 20], near duplicate image identification [40], classification of texture and object categories [42, 19], NMF [31] and contour matching [12]. Many of these works used saturated distances, usually the negative exponent function. The major contribution of this paper is a fast algorithm for the computation of the EMD with thresholded ground distance. We argue that thresholded distances have all the benefits of the negative exponent function that is typically used as a saturated distance; its big advantage is its much shorter computation time.

This paper is organized as follows. Section 2 is an overview of previous work. Section 3 describes the Earth Mover's Distance. Section 4 discusses thresholded distances and proves that they are metrics. Section 5 describes the fast algorithm. Section 6 presents the results. Finally, conclusions are drawn in Section 7.

2. Previous Work

This section first describes EMD algorithms. Second, it describes the use of saturated ground distances in the EMD framework.

2.1. EMD Algorithms

Early work using cross-bin distances for histogram comparison can be found in [33, 39, 38, 28]. Shen and Wong [33] suggested unfolding two integer histograms, sorting them and then computing the L_1 distance between the unfolded histograms. To compute the modulo matching distance between cyclic histograms they took the minimum from all cyclic permutations. This distance is equivalent to the EMD between two normalized histograms. Werman *et al*. [39] showed that this distance is equal to the L_1 distance between the cumulative histograms. They also proved that matching two cyclic histograms by only examining cyclic permutations is optimal. Werman *et al*. [38] proposed an $O(M \log M)$ algorithm for finding a minimal matching between two sets of M points on a circle. The algorithm was adapted by Pele and Werman [26] to compute the EMD between two N-bin, normalized histograms with time complexity $O(N)$. Peleg *et al*. [28] suggested using the EMD for grayscale images and using linear programming to compute it. Rubner *et al*. [29] suggested using the EMD for color and texture images and generalized the definition of the EMD to non-normalized histograms. They computed the EMD using a specific linear programming algorithm - the transportation simplex. The algorithm's worst case time complexity is exponential. Practical run time was shown to be super-cubic. Interior-point algorithms or Orlin's algorithm [25] both have a time complexity of $O(N^3 \log N)$ and can also be used.

Ling and Okada proposed EMD-L_1[20]; *i.e*. EMD with L_1 as the ground distance. They showed that if the points lie on a Manhattan network (*e.g*. an image), the number of variables in the LP problem can be reduced from $O(N^2)$ to $O(N)$. To execute the EMD-L_1 computation, they employed a tree-based algorithm, Tree-EMD. Tree-EMD exploits the fact that a basic feasible solution of the simplex algorithm-based solver forms a spanning tree when the EMD-L_1 is modeled as a network flow optimization problem. The worst case time complexity is exponential. Empirically, they showed that this algorithm has an average time complexity of $O(N^2)$. Gudmundsson *et al*. [14] also put forward this simplification of the LP problem. They suggested an $O(N \log^{d-1} N)$ algorithm that creates a Manhattan network for a set of N points in $\mathbb{R}^d$. The Manhattan network has $O(N \log^{d-1} N)$ vertices and edges. Thus, using Orlin's algorithm [25] the EMD-L_1 can be computed with a time complexity of $O(N^2 \log^{2d-1} N)$. Indyk and Thaper [17] proposed approximating EMD-L_1 by embedding it into the L_1 norm. Embedding time complexity is $O(Nd \log \Delta)$, where N is the feature set size, d is the feature space dimension and Δ is the diameter of the union of the two feature sets. Grauman and Darrell [13] substituted L_1 with histogram intersection in order to approximate partial matching. Shirdhonkar and Jacobs [35] presented a linear-time algorithm for approximating EMD-L_1 for low dimensional histograms using the sum of absolute values of the weighted wavelet coefficients of the difference histogram. Lv *et al*. [23] proposed embedding an EMD with thresholded ground distance into the L_1 norm. Khot and Naor [18] showed that any embedding of the EMD over the d-dimensional Hamming cube into L_1 must incur a distortion of $\Omega(d)$, thus losing practically all distance information. Andoni *et al*. [5] showed that for sets with cardinalities upper bounded by a parameter s, the distortion reduces to $O(\log s \log d)$. A practical reduction in accuracy due to the approximation was reported by [12, 23, 35]. In order to increase precision, Grauman and Darrell [12] and Lv *et al*. [23] used the approximation as a filter that returns a set of similar objects, and then used the exact EMD computation to rerank these objects. Khanh Do Ba *et al*. [6] presented optimal algorithms for estimating EMD-L_1 or EMD with a tree-metric as the ground distance.

Pele and Werman [26] proposed $\widehat{EMD}$ - a new definition of the EMD for non-normalized histograms. They showed that unlike Rubner's definition, the $\widehat{EMD}$ is also a metric for non-normalized histograms. In addition, they proposed a linear-time algorithm that computes the $\widehat{EMD}$ with a ground distance of 0 for corresponding bins, 1 for adjacent bins and 2 for farther bins and for the extra mass.

2.2. Saturated Ground Distances with the EMD

Rubner *et al*. [29] and Ruzon and Tomasi [30] used a negative exponent function to saturate their ground distance for the tasks of image retrieval and edge detection, respectively. The negative exponent function practically saturates large distances to a fixed threshold (see Fig. 2). The negative exponent function is used for saturating a metric, because it does not break the triangle inequality. We show that this is true as well for a thresholding function. Note that saturating with a negative exponent might have the drawback of changing the behavior of small distances (see Fig. 2).

Lv *et al*. [23] conducted image retrieval experiments from a database of 10000 images. They showed that thresholding the ground distance improves precision.

Pele and Werman [26] compared several distances for the task of SIFT matching. They proposed an $\widehat{EMD}$ variant. The ground distance of this $\widehat{EMD}$ is 0 for corresponding bins, 1 for adjacent bins and 2 for farther bins and for the extra mass; *i.e*. a thresholded distance. They showed that this distance improves SIFT matching, while $\widehat{EMD}$ with non-thresholded distances negatively affects performance.

3. The Earth Mover's Distance

The Earth Mover's Distance (EMD) [29] is defined as the minimal cost that must be paid to transform one histogram into the other, where there is a "ground distance" between the basic features that are aggregated into the histogram.

Given two histograms P, Q the EMD as defined by Rubner *et al*. [29] is:

$$\text{EMD}(P,Q) = (\min_{\{f_{ij}\}} \sum_{i,j} f_{ij} d_{ij}) / (\sum_{i,j} f_{ij}) \quad s.t \quad f_{ij} \geq 0$$

$$\sum_j f_{ij} \leq P_i \quad \sum_i f_{ij} \leq Q_j \quad \sum_{i,j} f_{ij} = \min(\sum_i P_i, \sum_j Q_j)$$

where $\{f_{ij}\}$ denotes the flows. Each f_{ij} represents the amount transported from the ith supply to the jth demand. We call d_{ij} the *ground distance* between bin i and bin j in the histograms. Pele and Werman [26] suggested $\widehat{EMD}$:

$$\widehat{EMD}_\alpha(P,Q) = (\min_{\{f_{ij}\}} \sum_{i,j} f_{ij} d_{ij}) + |\sum_i P_i - \sum_j Q_j| \alpha \max_{i,j} d_{ij}$$

$s.t$ EMD constraints

Pele and Werman proved that $\widehat{EMD}$ is a metric for any two histograms if the ground distance is a metric and $\alpha \geq 0.5$ [26]. The metric property enables fast algorithms for nearest neighbor searches [41, 7]), fast clustering [10] and large margin classifiers [15, 36]. Because of these advantages we will use the Pele and Werman definition in the remainder of this paper (with $\alpha = 1$).

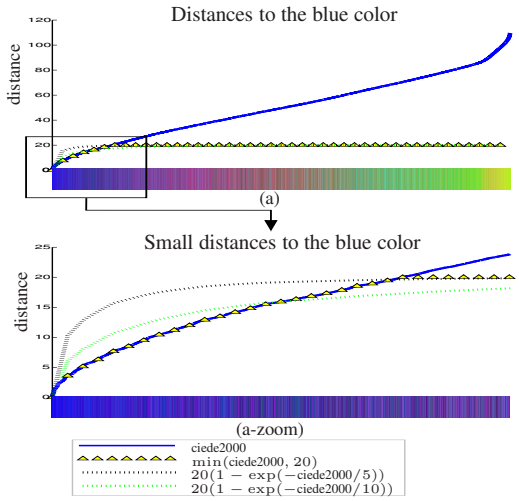

Figure 2. This figure should be viewed in color, preferably on a computer screen. The x-axes are colors, sorted by their distance to the blue color. The distances are the ciede2000 distance [22] and three monotonic saturating transformations applied to it: two negative exponent functions and a thresholding function; *i.e*. a minimum function. There are several observations we can derive from these graphs. First, although ciede2000 is considered as state of the art, it is still far from perfect, especially in the medium to large distance range. Second, color distances should be saturated. For example, although red and yellow are both simply different from blue, the ciede2000 distance between blue and red is 56, while the ciede2000 distance between blue and yellow is 102. This was already noted by Rubner *et al*. [29] and Ruzon and Tomasi [30] who suggested using a negative exponent function because it is a metric, if the distance that is raised to the power is a metric. In this paper we show that the thresholding function is also a metric if the thresholded distance is a metric. Finally, for most of the range, the negative exponent and thresholded functions are very similar. They mostly differ for small distances, where the negative exponent changes the original distance, while the thresholding function does not. Since the ciede2000 was designed and perceptually tested on this range [22], changing the distances on this range can negatively affect performance as was noted by Rubner *et al*. for the Euclidean distance on L*a*b* space [29].

4. Thresholded Distances

Thresholded distances are distances that saturate to a threshold; *i.e*. let $d(a,b)$ be a distance measure between two features - a, b. The thresholded distance with a threshold of $t > 0$ is defined as: $d_t(a,b) = \min(d(a,b), t)$.

We now prove that if d is a metric then d_t is also a metric. Non-negativity and symmetry hold trivially, so we only need to prove that the triangle inequality holds.

$d_t(a,b) + d_t(b,c) \geq d_t(a,c)$ if d is a metric.

We consider three cases:

1. $\Big(d_t(a,b) < t\Big) \wedge \Big(d_t(b,c) < t\Big) \wedge \Big(d_t(a,c) < t\Big) \Rightarrow$
$\Big(d_t(a,b) = d(a,b)\Big) \wedge \Big(d_t(b,c) = d(b,c)\Big) \wedge$
$\Big(d_t(a,c) = d(a,c)\Big) \Rightarrow d_t(a,b) + d_t(b,c) \geq d_t(a,c)$
2. $\Big(d_t(a,b) = t\Big) \vee \Big(d_t(b,c) = t\Big) \Rightarrow$
$d_t(a,b) + d_t(b,c) \geq t \geq d_t(a,c)$
3. $d_t(a,c) = t \Rightarrow$

Assume for contradiction that:

$$d_t(a,b) + d_t(b,c) < d_t(a,c) = t \Rightarrow$$
$$\Big(d_t(a,b) < t\Big) \wedge \Big(d_t(b,c) < t\Big) \Rightarrow$$
$$\Big(d_t(a,b) = d(a,b)\Big) \wedge \Big(d_t(b,c) = d(b,c)\Big) \Rightarrow$$
$$d(a,b) + d(b,c) < t = d_t(a,c) \leq d(a,c) \Rightarrow$$
$$d(a,b) + d(b,c) < d(a,c)$$

The last statement contradicts the triangle inequality of the metric d. The union of cases 2 and 3 is complement of case 1. Thus, cases 1-3 constitute the entire event space. Therefore we proved that d_t is a metric if d is a metric. It is noteworthy that d_t can be a metric even if d is not a metric.

5. Fast Computation of the EMD with a Thresholded Ground Distance

This section describes an algorithm that computes $\widehat{EMD}$ or EMD with a thresholded ground distance an order of magnitude faster than the original algorithm [2].

$\widehat{EMD}$ can be solved by a min-cost-flow algorithm. Our algorithm makes a simple transformation of the flow network that reduces the number of edges. If N is the number of bins in the histogram, the flow network of $\widehat{EMD}$ has exactly $N^2 + N$ edges (see (a) in Fig. 1). N^2 edges connect all sources to all sinks. The extra N edges connect all sources to the sink that handles the difference between the total mass of the two histograms (we assume without loss of generality that the source histogram total mass is greater or equal to the sink histogram total mass).

The transformation (see Fig. 1) first removes all edges with cost t. Second, it adds a new transhipment vertex. Finally we connect all sources to this vertex with edges of cost t and connect the vertex to all sinks with edges of cost 0.

Let K be the average number of edges going out of each bin that have a cost different than the threshold t. The new flow network has $NK + N$ edges from the original network, N edges connecting all sources to the transhipment vertex and N edges connecting the transhipment vertex to all sinks. Thus the total number of edges is $N(K + 3)$. If K is a constant the number of edges is $O(N)$ as opposed to the original $\Theta(N^2)$. Note that the new flow network is no longer a transportation problem, but a transhipment problem [2]. However, both are special cases of the min-cost-flow problem. Thus any algorithm that solves min-cost-flow can be used for both problems.

Let $K = O(1)$; the min-cost-flow optimization problem can be solved with a worst case time complexity of:

$$O(\min((N^2 \log\log U \log(NC)), \quad (N^2 \log U \sqrt{\log C}),$$
$$(N^2 \log U \log N), \quad (N^2 \log^2 N)))$$

The algorithms are taken from: Ahuja *et al*. [1], Edmonds and Karp [9], used with Ahuja *et al*.'s shortest path algorithm [3], Edmonds and Karp [9], used with Fredman and Tarjan's shortest path algorithm [11] and Orlin [25]. Algorithms with a C term assume integral cost coefficients that are bounded by C. Algorithms with a U term assume integral supply and demands that are bounded by U.

We now prove that the original and the transformed flow networks have the same minimum-cost solution. Let $\mathcal{O}$ be the original flow network and let $\mathcal{T}$ be the transformed flow network. We first show how to create a feasible flow in $\mathcal{T}$, given a feasible flow in $\mathcal{O}$. Both flows will have the same cost. This will prove that the min-cost-flow solution for $\mathcal{T}$ is smaller or equal to the min-cost-flow solution for $\mathcal{O}$.

Given a feasible flow in $\mathcal{O}$, all flows on edges that were not removed are copied to the new flow for $\mathcal{T}$. For flows on edges with the cost of the threshold, we transfer the flow through the transhipment vertex. This gives us a feasible flow in $\mathcal{T}$ with the same cost.

We now show how to create a feasible flow in $\mathcal{O}$, given a feasible flow in $\mathcal{T}$. The flow in $\mathcal{O}$ will have a cost smaller or equal to the cost of the flow in $\mathcal{T}$. This will prove that the min-cost-flow solution for $\mathcal{T}$ is greater or equal to the min-cost-flow solution for $\mathcal{O}$. Together with the previous proof, this shows that the two flow networks have the same min-cost solution.

Given a feasible flow in $\mathcal{T}$, all flows on edges not connected to the transhipment vertex are copied to $\mathcal{O}$. Second, each unit of mass that flows from vertex i to the transhipment vertex and then from the transhipment vertex to vertex j is transferred directly from vertex i to vertex j. We note that this is possible since $\mathcal{O}$ is fully bi-partite. We also note that the cost of the new flow will be smaller or equal to the cost of the flow in $\mathcal{T}$ as all edges in $\mathcal{O}$ have a cost smaller or equal to the threshold. This completes the proof.

It is noteworthy that the algorithmic technique presented in this paper can be applied not only with thresholded ground distance, but in any case where a group of vertexes can be connected to another group with the same cost. For example, we can add a transhipment vertex for all vertexes with a specific color (*e.g*. blue) such that the cost of the transportation between them will be smaller than the cost of the transportation to other colors.

[2]Note that the optimization problems in EMD and $\widehat{EMD}$ are exactly the same. Thus, any algorithm that computes EMD can compute $\widehat{EMD}$ and vice versa with the same time complexity.

5.1. Implementation notes

Flow-netowrk set-up time. For a fixed histogram configuration (*e.g*. SIFT) the flow-network can be pre-computed once. For sparse histograms (*signatures*), the flow-network set-up time complexity is $O(MN)$; where N is the number of non-zero bins and M is the average number of neighbors that need to be checked if the distance is lower than the threshold. M is at most N but it can be lower. For example, let t be the threshold. If we are comparing two images and the ground distance is a linear combination of the spatial distance and the color distance, then the distance computation to vertexes with L_1 distance bigger or equal to t can be skipped. That is, the set-up time in this case is $O(\min(t^2N, N^2))$.

Pre-flowing Monge sequences. A Monge sequence contains edges in the flow-network that can be pre-flowed (in the order of the sequence) without changing the min-cost solution [24, 16]. For example, if the ground-distance is a metric, zero-cost edges are Monge sequence [38]. Alon *et al*. [4] introduced an efficient algorithm which determines the longest Monge sequence.

Pre-flowing to/from isolated nodes. If a source is connected only to the new transhipment vertex, we can pre-flow all its mass to the transhipment vertex and eliminate it. If a sink is connected only to the transhipment vertex, we can add its deficit to the transhipment vertex and eliminate it.

6. Results

In this section we present results for image retrieval. However, note that we do not claim that this is the optimal way to achieve image retrieval. Image retrieval is used here as an example of an application where the EMD has already been used, to show that thresholded distances yield good results. The major contribution is the faster algorithm. We show that by using our algorithm the running time decreases by an order of magnitude.

We use a database that contains 773 landscape images from the COREL database, that were also used in Wang *et al*. [37]. The dataset contains 10 classes[3]: People in Africa, Beaches, Outdoor Buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains and Food. The number of images in each class ranges from 50 to 100.

From each class we selected 5 images as query images (numbers $1, 10, \ldots, 40$). Then we searched for the 50 nearest neighbors for each query image. We computed the distance of each image to the query image and its reflection and took the minimum. We present results for three types of image representations: histograms of orientations, L*a*b* color space and finally linear combinations of the two. We conclude this section with running times.

[3]The original database contains some visually ambiguous classes such as Africa that also contains images of beaches in Africa. We manually removed these ambiguous images.

6.1. SIFT

Our first image representation is orientation histograms. The first representation - SIFT is a $6 \times 8 \times 8$ SIFT descriptor [21] computed globally on the whole image. The second representation - CSIFT is a SIFT-like descriptor. This descriptor tackles two problems related to the SIFT descriptor for color image retrieval. First, it takes into account color edges by computing the SIFT descriptor on the compass edge image [30]. Note that on an edge image there should be no distinction between opposite directions (0 and 180 for example). Thus, opposite directions are considered equal. The second drawback of the SIFT descriptor for color image retrieval is its normalization. The normalization is problematic as we lose the distinctive cue of the amount of edge points in the image. In the CSIFT computation we skip the normalization step. The final CSIFT descriptor has $6 \times 8 \times 8$ bins. We used the following distances for these descriptors: L_1, L_2, χ^2, EMD-L_1 [20], SIFT$_{\text{DIST}}$ [26] and $\widehat{EMD}$. Let M be the number of orientation bins. The ground distances between bins (x_i, y_i, o_i) and (x_j, y_j, o_j) we use are:

$$d_R = ||(x_i, y_i) - (x_j, y_j)||_2 + \min(|o_i - o_j|, M - |o_i - o_j|)$$
$$d_T = \min(d_R, T)$$

The results are given in Fig. 4(a). Due to lack of space we present for each distance measure, the descriptor with which it performed best. Results of all pairs of descriptors and distance measures can be found at: `www.cs.huji.ac.il/%7eofirpele/FastEMD/`. The $\widehat{EMD}$ with a thresholded ground distance performs much better than $\widehat{EMD}$ with a non-thresholded ground distance. In fact, while $\widehat{EMD}$ with a non-thresholded ground distance negatively affects performance, $\widehat{EMD}$ with a thresholded ground distance improves performance. L_1 is equivalent to $\widehat{EMD}$ with the Kroncker δ ground distance [26]. SIFT$_{\text{DIST}}$ is the sum of $\widehat{EMD}$ over all the spatial cells (each spatial cell contains one orientation histogram). The ground distance for the orientation histograms is: $\min(|o_i - o_j|, M - |o_i - o_j|, 2)$. It was shown in [26] and here that this addition of a small invariance to the orientations shifts improves performance. Our distance also adds a small invariance to spatial shifts and thus improves performance even more. However, using a non-thresholded distance adds too much invariance at the expense of distinctiveness, thus reducing performance.

6.2. L*a*b* Color Space

The state of the art color distance is Δ_{00} - ciede2000 on L*a*b* color space[22, 32] (see also Fig. 2). Thus, our second type of image representation is simply a resized color image in the L*a*b* space. We resized each image to 32×48 and converted them to L*a*b* space. Let I_1, I_2 be

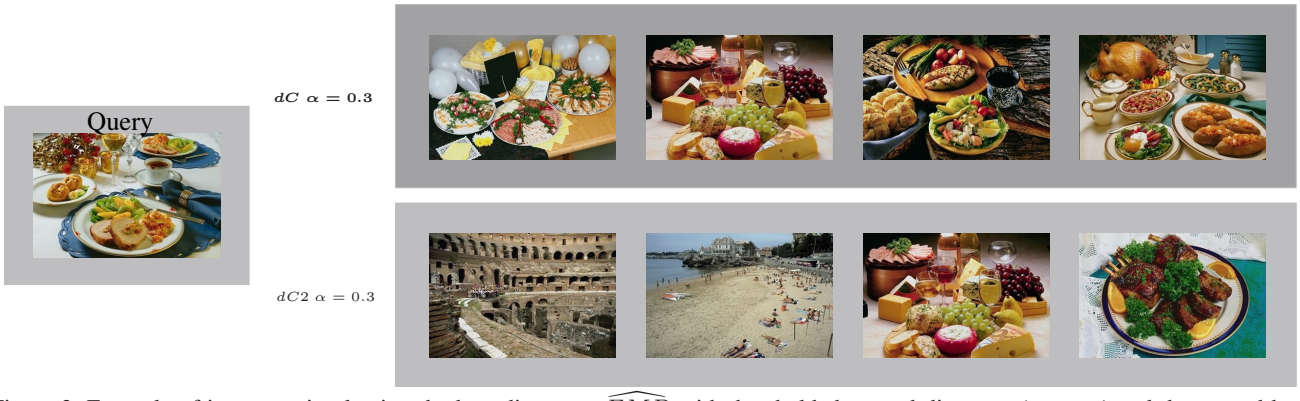

Figure 3. Example of image retrieval using the best distance - $\widehat{EMD}$ with thresholded ground distances (top row) and the second best distance - L_1 like distance (bottom row). The nearest neighbor images are ordered from left to right by their distance from the query image. Note that by allowing small deformations in the $\widehat{EMD}$ we obtain results that are visually similar to the query image. Allowing larger deformations; *i.e*. using a non-thresholded distance negatively affects results.

the two L*a*b* images. We used the following distances:

$$L_1\Delta_{00} = \sum_{x,y}(\Delta_{00}(I_1(x,y), I_2(x,y)))$$

$$L_1\Delta_{00}^T = \sum_{x,y}(\min(\Delta_{00}(I_1(x,y), I_2(x,y)), T))$$

$$L_2\Delta_{00} = \sum_{x,y}(\Delta_{00}(I_1(x,y), I_2(x,y)))^2$$

$$L_2\Delta_{00}^T = \sum_{x,y}(\min(\Delta_{00}(I_1(x,y), I_2(x,y)), T))^2$$

We also used $\widehat{EMD}$, where the ground distance between two pixels $(x_i, y_i, L_i, a_i, b_i)$, $(x_j, y_j, L_j, a_j, b_j)$ is:

$$dc_T = \min(||(x_i, y_i) - (x_j, y_j)||_2 + \Delta_{00}((L_i, a_i, b_i), (L_j, a_j, b_j)), T)$$

The results for $\widehat{EMD}$ with a non-thresholded ground distance are not reported here since the experiments have not finished running (more than ten days). Results are presented in Fig. 4(b). As shown, $\widehat{EMD}$ with a thresholded ground distance outperforms all other distances.

6.3. Color and SIFT Combined

In the experiments described in this section, we used linear combinations of the orientation histograms and color. We combined the two best distances for each of the methods (note that each distance was normalized so that its average is 1):

$$dC = \alpha\left(\widehat{EMD}\ \ d_{T=2}, \text{CSIFT}\right) + (1-\alpha)\left(\widehat{EMD}\ \ dc_{T=20}\right)$$

We also used a combination of the two L_1-like distances:

$$dC2 = \alpha\left(L_1, SIFT\right) + (1-\alpha)\left(L_1\Delta_{00}^{T=20}\right)$$

We used three different α values: 0.1, 0.3, 0.5. The results appear in Fig. 4(c)-(f). The combination of two $\widehat{EMD}$ with thresholded distances performs best, especially for hard classes such as People in Africa and Food. The image results examples for one "Food" query image are given in Fig. 3. More image results can be found at:
`www.cs.huji.ac.il/%7eofirpele/FastEMD/`

6.4. Running Time Results

The algorithm we used for the computation of the $\widehat{EMD}$ is successive shortest path [2]. This algorithm has a worst time complexity of $O(N^2U\log N)$. All runs were conducted on a Pentium 2.8GHz. A comparison of the practical running time of our algorithm and other methods are given in tables 1,2 and in Fig. 5. It is noteworthy that EMD-L_1 accuracy is much lower than our method and even lower than the simple L_1 norm (see Fig. 4(a)). Indyk and Thaper [17] and Shirdhonkar and Jacobs [35] approximate EMD-L_1, so their accuracy is even lower. $\text{SIFT}_{\text{DIST}}$ gives good accuracy (second best, see Fig. 4(a)) and is faster than our method. Our method has better accuracy. More importantly, $\text{SIFT}_{\text{DIST}}$ is limited to a thresholded L1 norm with a threshold of 2 for 1-dimensional histograms. Our method is much more general. For example, $\text{SIFT}_{\text{DIST}}$ cannot be applied to colors. Lv *et al*. [23] report that their approximation runs 5 times faster than Rubner's. Our method runs 75-700 times faster and returns the exact distance.

Our	Rubner's [29]	EMD-L_1[20]	$\text{SIFT}_{\text{DIST}}$[26]
0.04s	3s	0.04s	0.00007s

Table 1. 384-dimensional SIFT-like descriptors matching time

Our	Rubner's [29]	EMD-L_1[20]	$\text{SIFT}_{\text{DIST}}$[26]
6s	4400s	N.A.	N.A.

Table 2. L*a*b* images matching time. Note that EMD-L_1 can be applied only to regular grids and $\text{SIFT}_{\text{DIST}}$ can be applied only to 1-dimensional histograms.

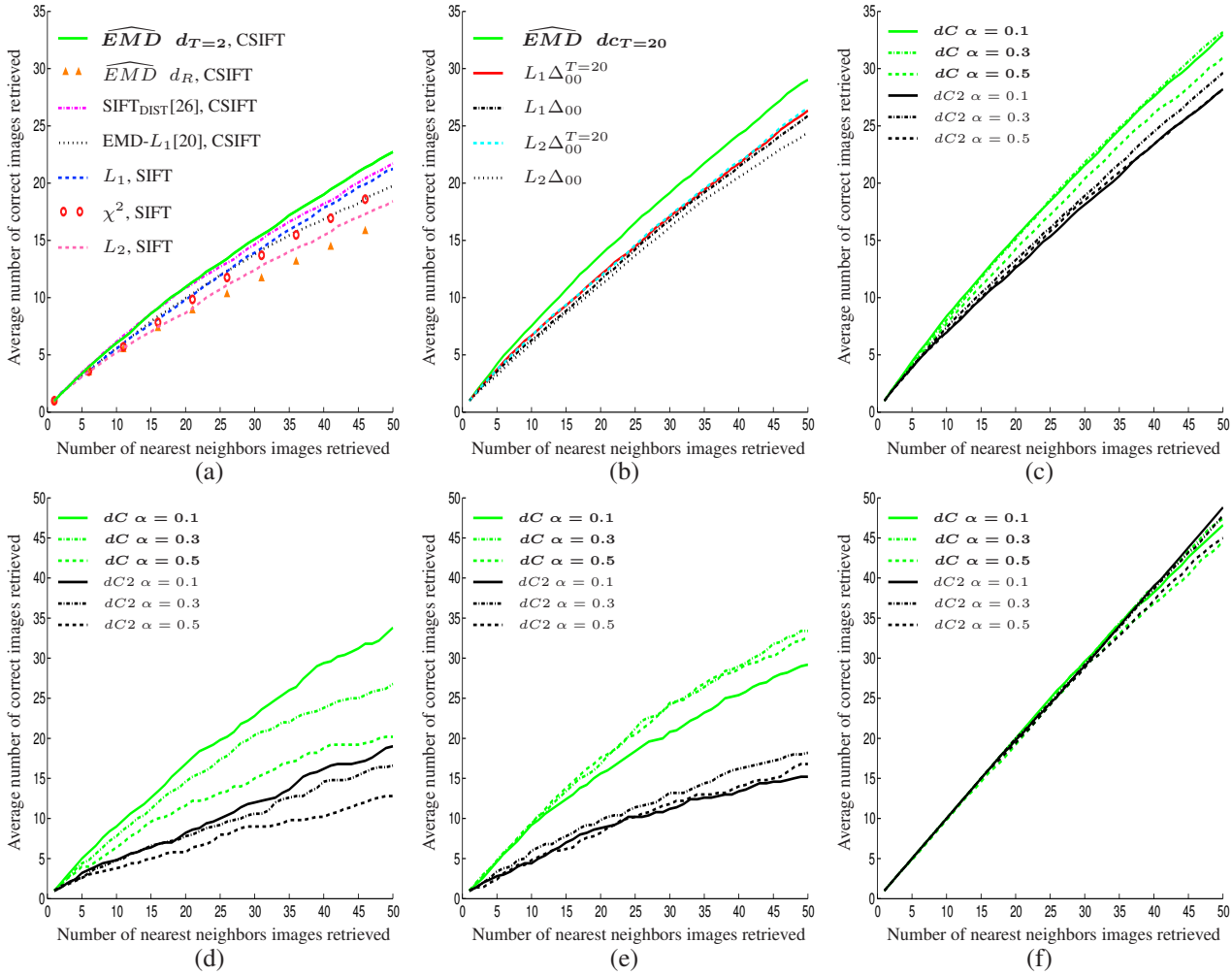

Figure 4. Results for image retrieval. Our method is in bold font. (a) Orientation histogram results. Due to lack of space we present for each distance measure, the descriptor with which it performed best. Results of all pairs of descriptors and distance measures can be found at:
`www.cs.huji.ac.il/%7eofirpele/FastEMD/`. (b) L*a*b* color space results. (c)-(f) Linear combinations of color and orientation results, where (c) is an average over all classes and (d)-(f) are for specific classes: (d) People in Africa (e) Food (f) Flowers. There are two key observations. First, $\widehat{EMD}$ with thresholded ground distances performs best. Second, some of the classes are easy, *e.g.* (f) Flowers. For these classes $\widehat{EMD}$ does not improve performance. However, for harder classes, such as (d) People in Africa and (e) Food, $\widehat{EMD}$ significantly improves results. Image result examples for one "Food" query image are given in Fig. 3.

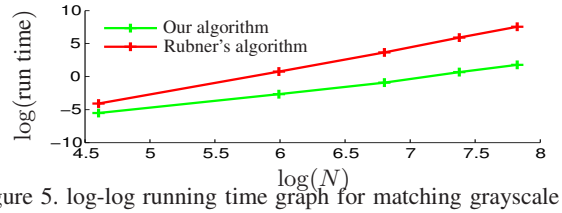

Figure 5. log-log running time graph for matching grayscale images with N pixels.. Linear fit for our algorithm: $\mathbf{2.3} \times \log(N) - 16$. Linear fit for Rubner's algorithm: $\mathbf{3.6} \times \log(N) - 21$.

7. Conclusions

We presented a new family of Earth Mover's Distances. Members of this family correspond to the way humans perceive distances, and are robust to outlier noise and quantization effects. We proved that they are metrics. We also proposed a fast algorithm. The algorithm runs an order of magnitude faster than the original algorithm, which makes it possible to compute the EMD on large histograms and databases. Experimental results show that EMD has the best performance when it is used with thresholded distances. This has also been shown by Rubner *et al*. [29] and Ruzon and Tomasi [30] for saturated distances, which are essentially thresholded. This has also been demonstrated for thresholded distances by Lv *et al*. [23] and Pele and Werman [26]. Our results strengthen these findings. Most im-

portantly, our paper shows that using a thresholded distance not only improves accuracy, but reduces the run time, using our algorithm. The speed can be further improved using techniques such as Bayesian sequential hypothesis testing [27]. The project homepage, including code (C++ and Matlab wrappers) is at:
`www.cs.huji.ac.il/%7eofirpele/FastEMD/`.

References

[1] R. Ahuja, A. Goldberg, J. Orlin, and R. Tarjan. Finding minimum-cost flows by double scaling. *Mathematical Programming*, 1992.

[2] R. Ahuja, T. Magnanti, and J. Orlin. *Network flows: theory, algorithms, and applications*. 1993.

[3] R. Ahuja, K. Mehlhorn, J. Orlin, and R. Tarjan. Faster algorithms for the shortest path problem. *JACM*, 1990.

[4] N. Alon, S. Cosares, D. S. Hochbaum, and R. Shamir. An algorithm for the detection and construction of monge sequences. *LAA*, 1989.

[5] A. Andoni, P. Indyk, and R. Krauthgamer. Earth mover distance over high-dimensional spaces. In *SODA*, 2008.

[6] K. D. Ba, H. L. Nguyen, H. N. Nguyen, and R. Rubinfeld. Sublinear time algorithms for earth mover's distance. *CoRR*, 2009.

[7] P. Ciaccia, M. Patella, and P. Zezula. M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. In *ICVLDB*, 1997.

[8] R. Collins and W. Ge. CSDD Features: Center-Surround Distribution Distance for Feature Extraction and Matching. In *ECCV*, 2008.

[9] J. Edmonds and R. Karp. Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems. *JACM*, 1972.

[10] C. Elkan. Using the Triangle Inequality to Accelerate k-Means. In *ICML*, 2003.

[11] M. Fredman and R. Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *JACM*, 1987.

[12] K. Grauman and T. Darrell. Fast contour matching using approximate earth mover's distance. In *CVPR*, 2004.

[13] K. Grauman and T. Darrell. The pyramid match kernel: Efficient learning with sets of features. *JMLR*, 2007.

[14] J. Gudmundsson, O. Klein, C. Knauer, and M. Smid. Small Manhattan Networks and Algorithmic Applications for the Earth Movers Distance. In *EWCG*, 2007.

[15] M. Hein, O. Bousquet, and B. Schölkopf. Maximal margin classification for metric spaces. *JCSS*, 2005.

[16] A. Hoffman. On simple linear programming problems. In *SPM*, 1963.

[17] P. Indyk and N. Thaper. Fast image retrieval via embeddings. In *IWSCTV*, 2003.

[18] S. Khot and A. Naor. Nonembeddability theorems via Fourier analysis. *Mathematische Annalen*, 2006.

[19] S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions. *PAMI*, 2005.

[20] H. Ling and K. Okada. An Efficient Earth Mover's Distance Algorithm for Robust Histogram Comparison. *PAMI*, 2007.

[21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.

[22] M. Luo, G. Cui, and B. Rigg. The Development of the CIE 2000 Colour-Difference Formula: CIEDE2000. *CRA*, 2001.

[23] Q. Lv, M. Charikar, and K. Li. Image similarity search with compact data structures. In *ICIKM*, 2004.

[24] G. Monge. Déblai et remblai. *Mémoires de l'Académie des Sciences*, 1781.

[25] J. Orlin. A faster strongly polynomial minimum cost flow algorithm. In *STOC*, 1988.

[26] O. Pele and M. Werman. A linear time histogram metric for improved sift matching. In *ECCV*, 2008.

[27] O. Pele and M. Werman. Robust real time pattern matching using bayesian sequential hypothesis testing. *PAMI*, 2008.

[28] S. Peleg, M. Werman, and H. Rom. A unified approach to the change of resolution: Space and gray-level. *PAMI*, 1989.

[29] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *IJCV*, 2000.

[30] M. Ruzon and C. Tomasi. Edge, Junction, and Corner Detection Using Color Distributions. *PAMI*, 2001.

[31] R. Sandler and M. Lindenbaum. Nonnegative Matrix Factorization with Earth Movers Distance Metric. In *CVPR*, 2009.

[32] G. Sharma, W. Wu, and E. Dalal. The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *CRA*, 2005.

[33] H. Shen and A. Wong. Generalized texture representation and metric. *CVGIP*, 1983.

[34] R. Shepard. Toward a universal law of generalization for psychological science. *Science*, 1987.

[35] S. Shirdhonkar and D. Jacobs. Approximate earth movers distance in linear time. In *CVPR*, 2008.

[36] U. von Luxburg and O. Bousquet. Distance–Based Classification with Lipschitz Functions. *JMLR*, 2004.

[37] J. Wang, J. Li, and G. Wiederhold. SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture LIbraries. *PAMI*, 2001.

[38] M. Werman, S. Peleg, R. Melter, and T. Kong. Bipartite graph matching for points on a line or a circle. *JoA*, 1986.

[39] M. Werman, S. Peleg, and A. Rosenfeld. A distance metric for multidimensional histograms. *CVGIP*, 1985.

[40] D. Xu, T. Cham, S. Yan, and S. Chang. Near Duplicate Image Identification with Spatially Aligned Pyramid Matching. In *CVPR*, 2008.

[41] P. Yianilos. Data structures and algorithms for nearest neighbor search in general metric spaces. In *SODA*, 1993.

[42] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *IJCV*, 2007.