The Tangent Earth Mover's Distance

Ofir Pele and Ben Taskar

The University of Pennsylvania, Philadelphia PA 19104-6389, USA,
{ofirpele,taskar}@cis.upenn.edu

Abstract. We present a new histogram distance, the Tangent Earth Mover's Distance (TEMD). The TEMD is a generalization of the Earth Mover's Distance (EMD) that is invariant to some global transformations. Thus, like the EMD it is robust to local deformations. Additionally, it is robuster to global transformations such as global translations and rotations of the whole image. The TEMD is formulated as a linear program which allows efficient computation. Additionally, previous works about the efficient computation of the EMD that reduced the number of variables in the EMD linear program can be used to accelerate also the TEMD computation. We present results for image retrieval using the Scale Invariant Feature Transform (SIFT) and color image descriptors. We show that the new TEMD outperforms state of the art distances.

1 Introduction

Histograms are ubiquitous tools in numerous computer vision and machine learning 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, global geometric transformations, etc.

The Earth Mover's Distance (EMD) [1] 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 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, Pele and Werman [2] suggested $\widehat{\text{EMD}}$ and showed that it is a metric for all histograms.

A shortcoming of the EMD is that it does not differentiate between a large number of non-structured deformations and a global deformation. For example, it cannot differentiate between the case that every edgel translated in a random direction and the case that all edgels translated in the same direction.

We propose the Tangent Earth Mover's Distance (TEMD). This new distance does not penalize some global-coherent transformations (*e.g.* global translations). The TEMD is formulated as a linear program which allows efficient computation. Additionally, previous works (*e.g.* [3, 4]) that reduced the number of variables in the EMD linear programming formulation for specific ground distances can be used to accelerate also the TEMD computation.

Notation: Small letters indicate scalars, capital letters indicate matrices, small letters with an arrow indicate vectors and calligraphic font indicate sets.

2 Previous Work

Early Work Early work using cross-bin distances for histogram comparison can be found in [5–8]. Shen and Wong [5] 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. [6] 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. Peleg et al. [8] suggested using the EMD for grayscale images and using linear programming to compute it.

Generalization of EMD for non-normalized histograms Rubner et al. [1] generalized the definition of the EMD to non-normalized histograms and coined the name EMD. Their formulation searches for the optimal partial matching. Pele and Werman [2] proposed a different formulation of the EMD for non-normalized histograms coined $\widehat{\text{EMD}}$. This formulation was shown to outperform Rubner's definition in many cases. Additionally, unlike Rubner's definition, the $\widehat{\text{EMD}}$ is a metric. We give these definitions in section 3.

EMD algorithms Rubner et al. [1] 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 supercubic. Orlin's algorithm [9] time complexity is $O(n^3 \log n)$. Werman et al. [7] 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 [2] to compute the EMD between normalized histograms with a linear time complexity. Rabin et al. [10] showed that the algorithm can be used for any convex cost of the geodesic distance along the circle. Pele and Werman [2] proposed a linear-time algorithm that computes the $\widehat{\text{EMD}}$ with a ground distance of 0 for corresponding bins, 1 for adjacent bins and 2 for farther bins and for the extra mass for cyclic one-dimensional histograms. Delon et al. [11] introduced a class of local indicators that enables the computation of the EMD for general concave ground distances on the line in the unitary case (each histogram bin contains either zero or one) with time complexity $O(n^2)$. For general histograms the time complexity is $O(n^3)$ but often it is more efficient. Ling and Okada proposed EMD-L^1[12]; *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. [13] 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 [9] the EMD-L^1 can be computed with a time complexity of $O(n^2 \log^{2d-1} n)$. Pele and Werman [4] presented a new formulation of the EMD or the $\widehat{\text{EMD}}$ with a robust thresholded ground distance. This formulation reduces the number of variables considerably. Additionally, $\widehat{\text{EMD}}$ distances with thresholded ground distances were shown to considerably outperform EMD or $\widehat{\text{EMD}}$ distances with non-robust, non-thresholded ground distances. The advantages of using robust distances as ground distances were also shown in [1, 14, 15]. Thus, we will use this formulation in our experiments and we give its definition in section 3.

Distances with invariance or robustness to global transformations Simard et al. [16] proposed the tangent distance. The method uses linear approximations of global deformations. That is, a global deformation is approximated with the addition of a tangent vector. The distance is defined as the minimum L^2 norm between the modified vectors. Our method also uses tangent vectors to approximate global deformations. However, unlike the tangent distance, we combine the global deformation with the EMD which is more robust than the L^2 to deformations that are not global. Cohen and Guibas [17] studied the problem of computing EMD which is invariant to transformation sets. The work focused on transformations that change the ground distance and thus the optimization problem becomes non-convex. They also studied the EMD where the histograms scale. Then, the value of the EMD is a decreasing function of the scale factor. In our work, we study transformations that do not change the total masses. Zhang et al. [18] modeled the transformation between two image attributes as a flow problem. They imposed soft affine constraints on the flow itself. Our method does not assume that the transformation should be fully affine. Zhang et al. also proposed adding a soft constraint that helps in handling clutter. Our method is orthogonal to the Zhang's method and the combination of both is an interesting future work.

3 The Earth Mover's Distance

The Earth Mover's Distance (EMD) [1] 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 $\overrightarrow{h1}, \overrightarrow{h2} \in \mathbb{R}^{n+}$ and a ground distance between the histograms bins $D \in \mathbb{R}^{(n \times n)+}$, the EMD as defined by Rubner et al. [1] is:

$$\begin{aligned} &\text{EMD}_D(\overrightarrow{h1}, \overrightarrow{h2}) = \min_F \frac{(\sum_{i,j} F_{ij} D_{ij})}{(\sum_{i,j} F_{ij})} \\ &s.t \quad \sum_j F_{ij} \leq \overrightarrow{h1}_i \quad \sum_i F_{ij} \leq \overrightarrow{h2}_j \quad \sum_{i,j} F_{ij} = \min(\sum_i \overrightarrow{h1}_i, \sum_j \overrightarrow{h2}_j) \quad F_{ij} \geq 0 \end{aligned} \tag{1}$$

where $\{F_{ij}\}$ denotes the flows. Each F_{ij} represents the amount transported from the ith supply to the jth demand.

Pele and Werman [2] suggested $\widehat{\text{EMD}}$, which also depends on a histogram sums difference penalty, c:

$$\begin{aligned}&\widehat{\text{EMD}}_{D,c}(\overrightarrow{h1},\overrightarrow{h2}) = (\min_F \sum_{i,j} F_{ij}D_{ij}) + |\sum_i \overrightarrow{h1}_i - \sum_j \overrightarrow{h2}_j|c \\ &s.t \quad \sum_j F_{ij} \le \overrightarrow{h1}_i \quad \sum_i F_{ij} \le \overrightarrow{h2}_j \quad \sum_{i,j} F_{ij} = \min(\sum_i \overrightarrow{h1}_i, \sum_j \overrightarrow{h2}_j) \quad F_{ij} \ge 0\end{aligned} \tag{2}$$

Pele and Werman [2] proved that $\widehat{\text{EMD}}$ is a metric for any two histograms if the ground distance is a metric and $c \ge 0.5 \max_{i,j} D_{ij}$. The metric property enables fast algorithms for nearest neighbor searches [19, 20] and fast clustering [21]. Because of these advantages we will use the Pele and Werman definition in the remainder of this paper.

As we are using the Pele and Werman [4] formulation of the $\widehat{\text{EMD}}$ with a thresholded ground distance, we give its description here. For each histogram bin i we have a set of neighbors $\mathcal{N}(i)$ where the ground distance between i and j is D_{ij} and smaller than the threshold r. The ground distance between all other bins is the threshold. The formulation has flow variables only for the neighbors and add one transshipment bin, 0, which connects all the bins of histogram $\overrightarrow{h1}$ to all the bins of histogram $\overrightarrow{h2}$ through it:

$$\begin{aligned}&\widehat{\text{EMD}}_{D,\mathcal{N},r,c}(\overrightarrow{h1},\overrightarrow{h2}) = (\min_F \sum_i (\sum_{j\in\mathcal{N}(i)} F_{ij}D_{ij} + F_{i0}r)) + |\sum_i \overrightarrow{h1}_i - \sum_j \overrightarrow{h2}_j|c \\ &s.t \quad \sum_{j\in\mathcal{N}(i)} F_{ij} + F_{i0} \le \overrightarrow{h1}_i \quad \sum_{i\in\mathcal{N}(j)} F_{ij} + F_{0j} \le \overrightarrow{h2}_j \quad \sum_i F_{i0} = \sum_j F_{0j} \\ &\qquad \sum_i (\sum_{j\in\mathcal{N}(i)} F_{ij} + F_{i0}) = \min(\sum_i \overrightarrow{h1}_i, \sum_j \overrightarrow{h2}_j) \quad F_{ij} \ge 0\end{aligned} \tag{3}$$

4 The Tangent Earth Mover's Distance (TEMD)

In the TEMD, each global transformation is represented using a tangent vector $\overrightarrow{l} \in \mathbb{R}^n$ where $\sum_i \overrightarrow{l}_i = 0$. Adding $\alpha\overrightarrow{l}$ to a histogram $\overrightarrow{h}$, pre-flows $\alpha \sum_i \overrightarrow{h}_i$ between its bins. For an example of a tangent vector of a histogram see Fig. 1. We use m types of such vectors (*e.g.* translation in x-axis, y-axis) and put them all in a global transformations matrix $L \in \mathbb{R}^{n\times m}$. We use these matrices to encode invariance to global transformations for the TEMD.

Given two histograms $\overrightarrow{h1}, \overrightarrow{h2} \in \mathbb{R}^{n+}$, their corresponding global transformations matrices $L1, L2$, a ground distance between the histograms bins $D \in \mathbb{R}^{n\times n+}$, the histogram sums difference penalty, c, the TEMD is defined as:

$$
\begin{aligned}
&\text{TEMD}_{D,c}(\{\overrightarrow{h1}, L1\}, \{\overrightarrow{h2}, L2\}) = \\
&(\min_{F,\overrightarrow{\alpha1},\overrightarrow{\alpha2}} \sum_{i,j} F_{ij} D_{ij}) + |\sum_i \overrightarrow{h1}_i - \sum_j \overrightarrow{h2}_j| c \\
&s.t \quad \sum_j F_{ij} \leq \overrightarrow{h1}_i + \sum_t \overrightarrow{\alpha1}_t L1_{it} \quad \sum_i F_{ij} \leq \overrightarrow{h2}_j + \sum_t \overrightarrow{\alpha2}_t L2_{jt} \\
&\qquad \sum_{i,j} F_{ij} = \min(\sum_i \overrightarrow{h1}_i, \sum_j \overrightarrow{h2}_j) \quad F_{ij} \geq 0 \quad \overrightarrow{\alpha1}_t \geq 0 \quad \overrightarrow{\alpha2}_t \geq 0
\end{aligned}
\tag{4}
$$

Like the EMD, TEMD is defined as the minimal cost that must be paid to transform one histogram into the other. The difference is that in EMD the transformation was direct from one histogram into the other where there is a "ground distance" between the basic features that are aggregated into the histogram. In TEMD, histograms $\overrightarrow{h1}$ and $\overrightarrow{h2}$ are first transformed into $\overrightarrow{h1}+L1\overrightarrow{\alpha1}$ and $\overrightarrow{h2}+L2\overrightarrow{\alpha2}$ respectively. The flow positivity constraints and the flow feasibility constraints ensure also that these transformed histograms are also non-negative.

The TEMD is a linear program. Previous works (*e.g.* [3, 4]) that reduced the number of variables in the EMD linear programming formulation for specific ground distances (see also Eq. 3) can be used to accelerate also the TEMD computation. To solve the linear program we use the mosek package [22].

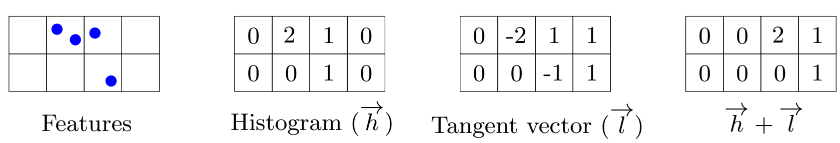

Fig. 1. An example of a histogram and its tangent vector for translation in the x-axis.

5 Results

We present results using the newly defined TEMD and state of the art distances, for image retrieval using SIFT-like descriptors and color image descriptors. We use the Pele and Werman benchmark [4]. They have compared $\widehat{\text{EMD}}$ with a thresholded ground distance to state of the art histogram distances such as $\text{SIFT}_{\text{DIST}}$[2], EMD-$L^1$[12], χ^2 and the L^1 and L^2 norms. They have found that its performance was the best. We show that TEMD outperforms $\widehat{\text{EMD}}$ (both with a thresholded ground distance) and the tangent distance [16].

For completeness we now describe the benchmark that was used by Pele and Werman [4]. The benchmark contains 773 landscape images from the Corel database from 10 classes. The number of images in each class ranges from 50 to 100. The benchmark has 5 query images for each class (images $1, 10, \ldots, 40$). We searched for the 50 nearest neighbors for each query image (without the query). We computed the distance of each image to the query image and its reflection and took the minimum. We present results for two types of image representations: SIFT-like descriptors and small L*a*b* images.

SIFT-like Descriptors The representation is a $6 \times 8 \times 8$ SIFT descriptor [23] computed globally on a color-edge [14] image called CSIFT. See [4] for more details. For EMD and TEMD we used the same ground distance that was used in [4]. That is, let $b = 8$ be the number of orientation bins, the ground distance between two CSIFT bins (x_i, y_i, o_i) and (x_j, y_j, o_j) is:

$$D_{ij} = \min\Big(\big(||(x_i, y_i) - (x_j, y_j)||_2 + \min(|o_i - o_j|, b - |o_i - o_j|) \big) \ , \ 2 \Big) \tag{5}$$

We experimented with two values of histogram sums difference penalty $c = \max_{ij} d(i,j) = 2$ (as in [4]) and $c = \frac{\max_{ij} d(i,j)}{2} = 1$. The EMD is a metric for the ground distance that was chosen and for both penalties. We also experimented with the tangent distance [16] denoted TAN-L^2. Finally, we also experimented with a tangent distance which is coupled with the L^1 norm instead of the L^2 norm. This distance is denoted as TAN-L^1. TAN-L^1 is almost identical to TEMD with histogram sums difference penalty $c = 1$ and a ground distance of 2 times the Kroncker delta. The only difference is that the flow constraints in TEMD ensure that the histograms will not become negative after the tangent transformation. In practice the results were identical for both of these distances. For all tangent-type distances (TEMD, TAN-L^2, TAN-L^2) we used 4 tangent vectors for each CSIFT: the CSIFT descriptor minus a CSIFT descriptor where the x/y coordinates were shifted by ± 1 (we added empty bins to the border). The results are presented in Fig. 2(a). TEMD with histogram sums difference penalty $c = 1$ outperformed all other distances.

L*a*b* Images The second type of image representation is a small L*a*b* image (as in [4]). Each image was resized to 32×48 and converted to the L*a*b* space. Each possible color for each possible pixel is a histogram bin. Thus the histograms are sparse histograms with 32×48 ones for the image pixels and $((32 \times 48) \times 256^3) - (32 \times 48)$ zeros. For such sparse histograms the results using bin-to-bin distances such as L^1, L^2, TAN-L^1 or TAN-L^2 are very poor. For EMD and TEMD we used the same ground distance that was used in [4]. That is, 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:

$$D_{ij} = \min(||(x_i, y_i) - (x_j, y_j)||_2 + \Delta_{00}((L_i, a_i, b_i), (L_j, a_j, b_j)), 20) \tag{6}$$

where Δ_{00} is the CIEDE2000 color difference [24, 25]. As the histogram sums equal 32×48, the histogram sums difference penalty has no effect. For TEMD we used 4 tangent vectors for each image histogram: the original histogram minus a histogram which is the result of shifting the x/y coordinates of the original histogram by ± 1 (we added empty bins to the border). We present results in Fig. 2(b). As shown, TEMD outperformed EMD.

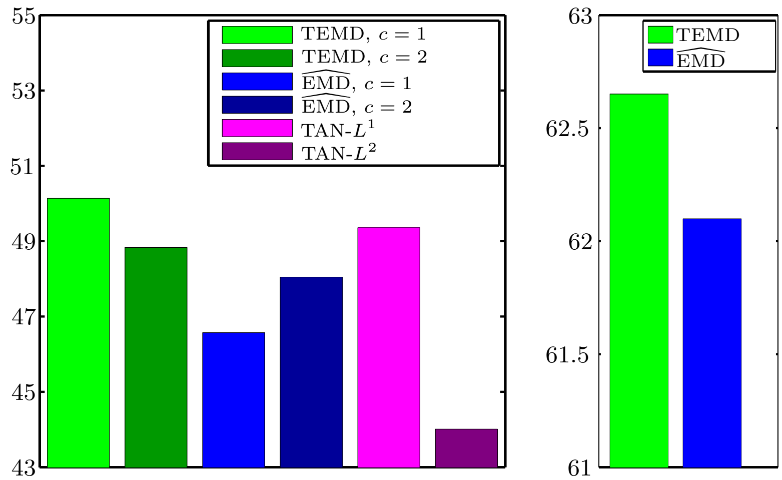

Fig. 2. Normalized area under the curve for retrieval (between 0 and 100, smaller than 50 for randomized classifier as there are 10 classes) using the CSIFT descriptors (left) and the L*a*b* images (right). Results for any bin-to-bin distance (*e.g.* TAN-L^1, TAN-L^1) using the L*a*b* images are not shown as they are very poor.

6 Conclusions and Future Work

We presented a new distance, the Tangent Earth Mover's Distance (TEMD). This distance is both robust to small non-structured deformations and is invariant to some global deformations. Experimental results show that TEMD outperforms state of the art distances on the well studied Corel dataset.

Future works include finding the TEMD Barycenter (as was done for the EMD by Rabin et al. [26]), TEMD clustering (as was done for the EMD by Ricci et al. [27] and Wagner and Ommer [28] and for the tangent distance by Hastie et al. [29]) and learning the TEMD parameters (as was done for the EMD by Cuturi and Avis [30] and Wang and Guibas [31]). Finally, assuming a-priori invariance to some global deformations might be a too restrictive prior. Instead, it will be interesting to penalize these transformations and thus achieve robustness instead of invariance. This will also allow us to learn global transformations penalties. The method code is available at the first author homepage.

References

1. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. IJCV (2000)
2. Pele, O., Werman, M.: A linear time histogram metric for improved sift matching. In: ECCV. (2008)
3. Ling, H., Okada, K.: An Efficient Earth Mover's Distance Algorithm for Robust Histogram Comparison. PAMI (2007)

4. Pele, O., Werman, M.: Fast and robust earth mover's distances. In: ICCV. (2009)
5. Shen, H., Wong, A.: Generalized texture representation and metric. CVGIP (1983)
6. Werman, M., Peleg, S., Rosenfeld, A.: A distance metric for multidimensional histograms. CVGIP (1985)
7. Werman, M., Peleg, S., Melter, R., Kong, T.: Bipartite graph matching for points on a line or a circle. JoA (1986)
8. Peleg, S., Werman, M., Rom, H.: A unified approach to the change of resolution: Space and gray-level. PAMI (1989)
9. Orlin, J.: A faster strongly polynomial minimum cost flow algorithm. In: STOC. (1988)
10. Rabin, J., Delon, J., Gousseau, Y.: Transportation distances on the circle. JMIV (2011)
11. Delon, J., Salomon, J., Sobolevski, A.: Local matching indicators for transport problems with concave costs. JDM (2012)
12. Ling, H., Okada, K.: An Efficient Earth Mover's Distance Algorithm for Robust Histogram Comparison. PAMI (2007)
13. Gudmundsson, J., Klein, O., Knauer, C., Smid, M.: Small Manhattan Networks and Algorithmic Applications for the Earth Movers Distance. In: EWCG. (2007)
14. Ruzon, M., Tomasi, C.: Edge, Junction, and Corner Detection Using Color Distributions. PAMI (2001)
15. Lv, Q., Charikar, M., Li, K.: Image similarity search with compact data structures. In: ICIKM. (2004)
16. Simard, P., LeCun, Y., Denker, J., Victorri, B.: Transformation invariance in pattern recognitiontangent distance and tangent propagation. NN (1998)
17. Cohen, S., Guibas, L.: The earth mover's distance under transformation sets. In: ICCV. (1999)
18. Zhang, W., Srinivasan, P., Shi, J.: Discriminative image warping using attribute flow. In: CVPR. (2011)
19. Yianilos, P.: Data structures and algorithms for nearest neighbor search in general metric spaces. In: SODA. (1993)
20. Ciaccia, P., Patella, M., Zezula, P.: M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. In: ICVLDB. (1997)
21. Elkan, C.: Using the Triangle Inequality to Accelerate k-Means. In: ICML. (2003)
22. MOSEK: The MOSEK optimization software. http://www.mosek.com/
23. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV (2004)
24. Luo, M., Cui, G., Rigg, B.: The Development of the CIE 2000 Colour-Difference Formula: CIEDE2000. CRA (2001)
25. Sharma, G., Wu, W., Dalal, E.: The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. CRA (2005)
26. Rabin, J., Peyré, G., Delon, J., Bernot, M.: Wasserstein barycenter and its application to texture mixing. SSVM (2012)
27. Ricci, E., Zen, G., Sebe, N., Messelodi, S.: A prototype learning framework using emd: Application to complex scenes analysis. (2012)
28. Wagner, J., Ommer, B.: Efficiently clustering earth movers distance. In: ACCV. (2010)
29. Hastie, T., Simard, P., Säckinger, E.: Learning prototype models for tangent distance. NIPS (1994)
30. Cuturi, M., Avis, D.: Ground metric learning. arXiv (2011)
31. Wang, F., Guibas, L.: Supervised earth movers distance learning and its computer vision applications. In: ECCV. (2012)