The Earth Mover's Distance - Beyond Nearest Neighbor Classification ?

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

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

Histograms are ubiquitous tools in numerous machine learning and computer vision tasks. It is common practice to use distances such as L_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) [12] 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:

$$\mathrm{EMD}(P,Q) = (\min_{\{f_{ij}\}} \sum_{i,j} f_{ij} d_{ij}) / (\sum_{i,j} f_{ij}) \quad s.t:$$
$$f_{ij} \geq 0 \quad \sum_j f_{ij} \leq P_i \quad \sum_i f_{ij} \leq Q_j \quad \sum_{i,j} f_{ij} = \min(\sum_i P_i, \sum_j Q_j)$$

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.

The EMD as defined by Rubner is a metric only for normalized histograms. However, recently Pele and Werman [10] suggested $\widehat{EMD}$:

$$\widehat{EMD}_{\alpha}(P,Q) = (\min_{\{f_{ij}\}} \sum_{i,j} f_{ij} d_{ij}) + |\sum_i P_i - \sum_j Q_j| \alpha \max_{i,j} d_{ij} \quad s.t:$$
$$f_{ij} \geq 0 \quad \sum_j f_{ij} \leq P_i \quad \sum_i f_{ij} \leq Q_j \quad \sum_{i,j} f_{ij} = \min(\sum_i P_i, \sum_j Q_j)$$

Pele and Werman proved that $\widehat{EMD}$ is a metric for any two histograms if the ground distance is a metric and $\alpha \geq 0.5$.

The EMD has some embeddability results. Khot and Naor [5] 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*. [1] showed that for sets with cardinalities upper bounded by a parameter s, the distortion reduces to $O(\log s \log d)$.

The Earth Mover's Distance has been used successfully in many applications such as image retrieval [12, 8, 11], edge and corner detection [13], keypoint matching [10, 2, 7], near duplicate image identification [17], classification of texture and object categories [18, 6], NMF [14], contour matching [3] and as a melodic similarity measure [15]. However, these methods used nearest neighbor classifiers, or SVMs with a transformation of EMD as a kernel (which is not assured to be positive-semidefinite).

The workshop organizers proposed two themes:

1. How can one obtain suitable similarity information from data representations that are more powerful than, or simply different from, the vectorial?

2. How can similarity information be used in order to perform learning and classification tasks?

The Earth Mover's Distance is one possible answer to the first question as it is inherently different from the Euclidean distance. The second question of how to perform learning and classification with EMD is an open question which we would like to discuss with the workshop participants.

The most efficient strongly polynomial algorithm for computing the EMD or $\widehat{EMD}$ with a general ground distance is Orlin's algorithm [9]. It's time complexity is $O(N^3 \log N)$. However, there are specific ground distances or families of ground distances that allow a faster computation of the EMD.

For one dimensional linear or cyclic normalized histograms with L_1 as the ground distance there is a linear time algorithm [16, 10]. Ling and Okada proposed EMD-L_1[7]; *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 $O(N)$. Orlin's algorithm time complexity on this simplified network is $O(N^2 \log N(d + \log N))$; Where d is the dimension. Gudmundsson *et al.* [4] 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 [10] 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.

Finally, Pele and Werman [11] proposed a simplification of the flow network for EMD or $\widehat{EMD}$ with thresholded ground distance. Orlin's algorithm time complexity on this simplified network is $O(N^2 \log N(K + \log N))$; Where K is the average number of edges for each bin with a cost different from the threshold. Pele and Werman [10, 11] also showed that EMD performs much better when it has a thresholded ground distance.

References

[1] A. Andoni, P. Indyk, and R. Krauthgamer. Earth mover distance over high-dimensional spaces. In *SODA*, 2008. 1

[2] R. Collins and W. Ge. CSDD Features: Center-Surround Distribution Distance for Feature Extraction and Matching. In *ECCV*, 2008. 1

[3] K. Grauman and T. Darrell. Fast contour matching using approximate earth mover's distance. In *CVPR*, 2004. 1

[4] J. Gudmundsson, O. Klein, C. Knauer, and M. Smid. Small Manhattan Networks and Algorithmic Applications for the Earth Movers Distance. In *EWCG*, 2007. 2

[5] S. Khot and A. Naor. Nonembeddability theorems via Fourier analysis. *Mathematische Annalen*, 2006. 1

[6] S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions. *PAMI*, 2005. 1

[7] H. Ling and K. Okada. An Efficient Earth Mover's Distance Algorithm for Robust Histogram Comparison. *PAMI*, 2007. 1, 2

[8] Q. Lv, M. Charikar, and K. Li. Image similarity search with compact data structures. In *ICIKM*, 2004. 1

[9] J. Orlin. A faster strongly polynomial minimum cost flow algorithm. In *STOC*, 1988. 2

[10] O. Pele and M. Werman. A linear time histogram metric for improved sift matching. In *ECCV*, 2008. 1, 2

[11] O. Pele and M. Werman. Fast and robust earth mover's distances. In *ICCV*, 2009. 1, 2

[12] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *IJCV*, 2000. 1

[13] M. Ruzon and C. Tomasi. Edge, Junction, and Corner Detection Using Color Distributions. *PAMI*, 2001. 1

[14] R. Sandler and M. Lindenbaum. Nonnegative Matrix Factorization with Earth Movers Distance Metric. In *CVPR*, 2009. 1

[15] R. Typke, P. Giannopoulos, R. Veltkamp, F. Wiering, and R. van Oostrum. Using transportation distances for measuring melodic similarity. In *ISMIR*, 2003. 1

[16] M. Werman, S. Peleg, and A. Rosenfeld. A distance metric for multidimensional histograms. *CVGIP*, 1985. 2

[17] D. Xu, T. Cham, S. Yan, and S. Chang. Near Duplicate Image Identification with Spatially Aligned Pyramid Matching. In *CVPR*, 2008. 1

[18] 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. 1