

Research Article

A Linear Approximate Algorithm for Earth Mover's Distance with Thresholded Ground Distance

Longjie Li, Min Ma, Peng Lei, Xiaoping Wang, and Xiaoyun Chen

School of Information Science & Engineering, Lanzhou University, Lanzhou 730000, China

Correspondence should be addressed to Xiaoyun Chen; chenxy@lzu.edu.cn

Received 16 July 2013; Accepted 27 November 2013; Published 16 January 2014

Academic Editor: Jian Li

Copyright © 2014 Longjie Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Effective and efficient image comparison plays a vital role in content-based image retrieval (CBIR). The earth mover's distance (EMD) is an enticing measure for image comparison, offering intuitive geometric interpretation and modelling the human perceptions of similarity. Unfortunately, computing EMD, using the simplex method, has cubic complexity. FastEMD, based on min-cost flow, reduces the complexity to $O(N^2 \log N)$. Although both methods can obtain the optimal result, the high complexity prevents the application of EMD on large-scale image datasets. Thresholding the ground distance can make EMD faster and more robust, since it can decrease the impact of noise and reduce the range of transportation. In this paper, we present a new image distance metric, EMD^+ , which applies a threshold to the ground distance. To compute EMD^+ , the FastEMD approach can be employed. We also propose a novel linear approximation algorithm. Our algorithm achieves $O(N)$ complexity with the benefit of qualified bins. Experimental results show that (1) our method is 2 to 3 orders of magnitude faster than EMD (computed by FastEMD) and 2 orders of magnitude faster than FastEMD and (2) the precision of our approximation algorithm is no less than the precision of FastEMD.

1. Introduction

The development of advanced multimedia technology increases the importance of image retrieval, since large-scale image datasets are proliferating. Effective and efficient methods for computing the similarity between images is vitally important to content-based image retrieval (CBIR) systems.

The earth mover's distance (EMD) is a famous measure for image retrieval [1]. It not only takes the corresponding bins into account, but also considers the correlations between noncorresponding bins and thus is robust to histogram shift and rotation. The effectiveness of EMD and/or its variants is witnessed in many application domains, such as image retrieval [1–4], common pattern discovery [5, 6], mathematical symbol retrieval [7], contour matching [8], texture retrieval [1, 9], shape matching [10], graph matching [11], visual event recognition [12], face verification [13, 14], visual tracker [15], and uncertain or probability data [16, 17].

The EMD defines the comparison of two histograms as the *transportation problem*. (In [1], EMD is used to compute the distance between two signatures. For the convenience of statement, we use histogram in this paper. In fact, our

algorithm can be used to both histogram and signature.) The minimum cost to transport one histogram to another one is taken as the distance of the two histograms. Rubner et al. [1] employ the transportation *simplex* algorithm [18] to compute the EMD. Thus the computation of EMD is very expensive, with cubic time complexity $O(N^3 \log N)$. Pele and Werman [19] present the FastEMD algorithm which is faster than the simplex algorithm. FastEMD is based on the min-cost flow algorithm, thus it still suffers from quadratic complexity $O(N^2 \log N)$. Ling and Okada [10] propose $EMD-L_1$ which employs L_1 -norm as the ground distance. $EMD-L_1$ can reduce the number of variables in the transportation simplex algorithm from $O(N^2)$ to $O(N)$, so the time complexity can be reduced to $O(N^2)$. But only using L_1 -norm as ground distance limits the application of $EMD-L_1$. Rahimi and Kiram [20] propose a greedy algorithm with complexity of $O(N^2)$. Shirdhonkar and Jacobs [21] present a linear time approximate EMD algorithm, *Wavelet EMD*, based on a weighted wavelet transform. To approximate EMD, wavelet EMD uses the weighted wavelet coefficients of the difference histogram. Another linear time approximate

EMD algorithm proposed by Jang et al. [22] is MHC which is used for dominant color descriptor (DCD). This algorithm employs Hilbert curves to fill the multidimensional color space and computes cost according to the Hilbert order of dominant color. To improve accuracy, MHC uses $(2^k \cdot k!)/2$ Hilbert curves to fill the k -dimensional space, then computes $(2^k \cdot k!)/2$ cost values, and takes the minimum cost as the distance between two images.

Furthermore, EMD is not a metric if the two histograms have different total weight. To solve this drawback, Ljosa et al. [4] introduce a special bin, called the *bank*, for each image. This extra bin allows the two histograms to exchange mass. The bank has the same ground distance to all other bins. Pele and Werman [3] propose a variant of EMD, named $\widehat{\text{EMD}}$, which takes the difference of two weights as an extra mass. The moving distance of the extra mass is θ times maximum ground distance. If the ground distance is a metric and $\theta \geq 0.5$, $\widehat{\text{EMD}}$ is a metric for all histograms.

The EMD with thresholded ground distance has proven to be more robust, since it can reduce the interference of noise and correspond to the way humans perceive distance [2, 3, 19]. With the help of thresholded ground distance, FastEMD can reduce the number of edges of flow network from $O(N^2)$ to $O(N)$ [19] and thus makes an improvement in both accuracy and speed.

In this paper, we present a new image distance metric, EMD^+ , based on thresholded ground distance. We prove that EMD^+ is a metric by deducing that EMD^+ equals $\widehat{\text{EMD}}$ with the same ground distance threshold. Therefore, we can employ FastEMD [19] to compute the value of EMD^+ . However, its quadratic complexity makes FastEMD a nonideal choice. According to the definition of EMD^+ , we derive a novel linear approximation algorithm which can compute EMD^+ very fast. In experiments we find that our algorithm is 2 to 3 orders of magnitude faster than EMD (computed by FastEMD) and 2 orders of magnitude faster than FastEMD. Moreover, the results of our approximation algorithm are no worse than the exact results.

In the remainder of the paper, we first review the earth mover's distance and $\widehat{\text{EMD}}$ in Section 2. We then formalize the EMD^+ in Section 3. Section 4 proposes the details of our novel linear approximation algorithm. In Section 5, we experimentally evaluate the practical performance of our algorithm. We conclude this paper in Section 6.

2. Earth Mover's Distance

The earth mover's distance is a perceptual and intuitive measure between two histograms. EMD which defines the distance between two histograms as the minimum cost that must be paid to transform one histogram to another one is based on the famous *transportation problem*. We consider one histogram as a *supplier* with a given amount of products distributed at some places and another histogram as a *demand* that has some warehouses with a given limited capacity. To transport products from the *supplier* to the *demand*, the optimal solution is required to minimize the transportation cost. The following is the formal definition of EMD.

Definition 1 (earth mover's distance). Given two histograms $P = \{(p_1, w_{p_1}), \dots, (p_m, w_{p_m})\}$ and $Q = \{(q_1, w_{q_1}), \dots, (q_n, w_{q_n})\}$ with sizes m and n respectively, where w_{p_i} is the weight of bin p_i and w_{q_j} is the weight of bin q_j . Let d_{ij} be the distance between p_i and q_j and let f_{ij} represent the amount of weight transferred from p_i to q_j . The total cost of matching P and Q is defined as

$$\text{WORK} = \sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij}, \quad (1)$$

with the following constraints:

$$\begin{aligned} f_{ij} \geq 0, \quad \sum_{j=1}^n f_{ij} \leq w_{p_i}, \quad \sum_{i=1}^m f_{ij} \leq w_{q_j}, \\ \sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min \left(\sum_{i=1}^m w_{p_i}, \sum_{j=1}^n w_{q_j} \right). \end{aligned} \quad (2)$$

The earth mover's distance is defined as the minimum total cost for solving this problem; that is,

$$\text{EMD}(P, Q) = \min_{\{f_{ij}\}} \frac{\sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}}. \quad (3)$$

The EMD is a many-to-many matching method which can handle variable-size structures. If the ground distance is a metric and the total weights of two histograms are equal, that is, $\sum_{i=1}^m w_{p_i} = \sum_{j=1}^n w_{q_j}$, EMD is a metric [1]. However, in many applications, the total weights of two histograms cannot be equal. To solve this drawback, Pele and Werman [3] present a new EMD variant, $\widehat{\text{EMD}}$, which is defined as

$$\begin{aligned} \widehat{\text{EMD}}_{\theta}(P, Q) = \min_{\{f_{ij}\}} \sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij} \\ + \left| \sum_{i=1}^m w_{p_i} - \sum_{j=1}^n w_{q_j} \right| \times \theta \max_{i,j} \{d_{ij}\}, \end{aligned} \quad (4)$$

subject to the same constraints as the EMD.

If the two histograms are probability distributions (i.e., total weight is one) EMD and $\widehat{\text{EMD}}$ are equivalent. Pele and Werman proved that if $\theta \geq 0.5$ and the ground distance is a metric, $\widehat{\text{EMD}}$ is a metric. The proof can be found in [3].

$\widehat{\text{EMD}}$ is more appropriate for two cases. One case is when the total weight of histograms is important. And the other one is when difference of total weight between histograms is distinguished [3].

3. New Distance Metric: EMD^+

Thresholded distances are distances that are limited by a threshold. Let d_{ij} be the distance between p_i and q_j and let $\mu > 0$ be a threshold; the thresholded distance between p_i and q_j is defined as $d_{ij}^{\mu} = \min(d_{ij}, \mu)$.

As mentioned in [2, 3, 19], thresholding the ground distance can improve the performance of EMD family.

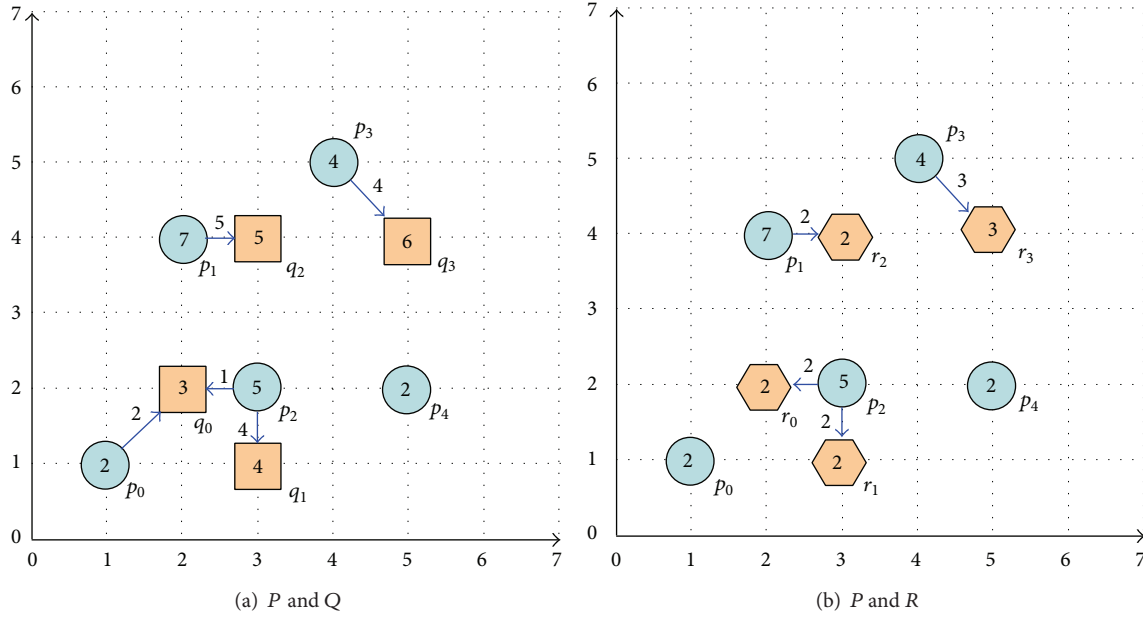


FIGURE 1: Three histograms P , Q , and R . The light cyan circles indicate histogram P , the light yellow squares indicate histogram Q , and the light yellow hexagons indicate histogram R . In both figures, P is the same histogram. One circle, square, and hexagon denote one bin of P , Q , and R , respectively. The number inside the circle or square or hexagon is the weight of that bin. The blue arrows describe the flow of mass; the number beside the arrow denotes the weight of flow.

Motivated by this thought, we design a new variant of EMD with thresholded ground distance. We call this new measure EMD^+ . If the ground distance is a metric, the EMD^+ is a metric. Firstly, we formally define EMD^+ in Definition 2.

Definition 2 (distance measure: EMD^+). Given two histograms $P = \{(p_1, w_{p_1}), \dots, (p_m, w_{p_m})\}$ and $Q = \{(q_1, w_{q_1}), \dots, (q_n, w_{q_n})\}$ with sizes m and n , respectively, where w_{p_i} is the weight of bin p_i and w_{q_j} is the weight of bin q_j . Let d_{ij} be the ground distance between p_i and q_j , let f_{ij} be the amount of weight transferred from p_i to q_j , and let $\mu > 0$ be a threshold of ground distance. If $d_{ij} < \mu$, EMD^+ moves as much mass as possible from P to Q ; otherwise, EMD^+ does not move any mass from P to Q . Support ω is the maximum amount that can be transported from P to Q ($\omega \leq \min\{\sum_{i=1}^m w_{p_i}, \sum_{j=1}^n w_{q_j}\}$). The distance between P and Q is defined as

$$EMD^+(P, Q) = \min_{\{f_{ij}\}} \sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij} \quad (5)$$

$$+ \left(\max \left\{ \sum_{i=1}^m w_{p_i}, \sum_{j=1}^n w_{q_j} \right\} - \omega \right) \times \mu,$$

s.t

$$f_{ij} \geq 0, \quad \sum_{i=1}^m \sum_{j=1}^n f_{ij} = \omega, \quad (6)$$

$$\sum_{j=1}^n f_{ij} \leq w_{p_i}, \quad \sum_{i=1}^m f_{ij} \leq w_{q_j}.$$

We denote the total weight of P and Q as $w(P)$ and $w(Q)$ respectively. Without loss of generality, we assume $w(P) \geq w(Q)$. The condition $d_{ij} < \mu$ in Definition 2 limits the range of bins that can be matched by p_i . Only the bins whose distances to p_i are smaller than the threshold have the chance to receive mass from p_i . The optimal solution is the minimum total cost of moving ω amount of mass from P to Q . After that, we ignore the limitation of capacity of Q and transport all the remainder weight of P to Q with cost of μ .

Figure 1 shows two examples of EMD^+ . From Figure 1, we obtain $w(P) = 20$, $w(Q) = 18$, and $w(R) = 9$. Set L_2 -norm to be the ground distance and $\mu = 2$ to be the threshold. We get

$$EMD^+(P, Q) = (10 + 6 \times \sqrt{2}) + (20 - 16) \times 2 \approx 26.5, \quad (7)$$

$$EMD^+(P, R) = (6 + 3 \times \sqrt{2}) + (20 - 9) \times 2 \approx 32.2.$$

If we calculate the EMD values, we get

$$EMD(P, Q) = \frac{10 + 6 \times \sqrt{2} + 2 \times 2}{18} \approx 1.25, \quad (8)$$

$$EMD(P, R) = \frac{6 + 3 \times \sqrt{2}}{9} \approx 1.14.$$

That is to say, $EMD(P, Q) > EMD(P, R)$, while $EMD^+(P, Q) < EMD^+(P, R)$. Actually, Q is more similar to P than R . Thus, EMD^+ has more discernibility.

In fact, the EMD^+ is equivalent to the \widehat{EMD} with a thresholded ground distance and $\theta = 1$ [19]. Because the \widehat{EMD} is a metric, we can deduce that EMD^+ is a metric.

Theorem 3. If the ground distance d is a metric, then EMD^+ is a metric.

Proof. A thresholded ground distance was introduced in [19]. d_{ij} is the ground distance between p_i and q_j , and $\mu > 0$ is a threshold. The thresholded ground distance is defined as $d_{ij}^\mu = \min(d_{ij}, \mu)$. If d is a metric, d^μ is a metric; a proof can be found in [19]. Without loss of generality, we assume $w(P) \geq w(Q)$. Set $\theta = 1$; we rewrite $\widehat{\text{EMD}}$ as

$$\begin{aligned}
& \widehat{\text{EMD}}_1(P, Q) \\
&= \min_{\{f_{ij}\}} \sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij}^\mu + (w(P) - w(Q)) \times \mu \\
&= \min_{\{f_{ij}\}} \sum_{i=1}^m \sum_{\substack{j=1, \\ d_{ij}^\mu < \mu}} f_{ij} d_{ij}^\mu + \sum_{i=1}^m \sum_{\substack{j=1, \\ d_{ij}^\mu = \mu}} f_{ij} \times \mu \\
&\quad + (w(P) - w(Q)) \times \mu \\
&= \min_{\{f_{ij}\}} \sum_{i=1}^m \sum_{\substack{j=1, \\ d_{ij}^\mu < \mu}} f_{ij} d_{ij}^\mu + \left(w(Q) - \sum_{i=1}^m \sum_{\substack{j=1, \\ d_{ij}^\mu < \mu}} f_{ij} \right) \\
&\quad \times \mu + (w(P) - w(Q)) \times \mu \\
&= \min_{\{f_{ij}\}} \sum_{i=1}^m \sum_{\substack{j=1, \\ d_{ij}^\mu < \mu}} f_{ij} d_{ij}^\mu \\
&\quad + \left(w(P) - \sum_{i=1}^m \sum_{\substack{j=1, \\ d_{ij}^\mu < \mu}} f_{ij} \right) \times \mu \\
&= \min_{\{f_{ij}\}} \sum_{i=1}^m \sum_{\substack{j=1, \\ d_{ij}^\mu < \mu}} f_{ij} d_{ij}^\mu + (w(P) - \omega) \times \mu \\
&= \text{EMD}^+(P, Q).
\end{aligned} \tag{9}$$

Due to the fact that $\widehat{\text{EMD}}_1$ is a metric with $d^\mu(\cdot, \cdot)$, thus EMD^+ is a metric. \square

4. Fast Computation of EMD^+

Intrinsically, EMD^+ has the same optimization problem as EMD and $\widehat{\text{EMD}}$. Hence, EMD^+ can be computed by either transportation simplex algorithm [18] (e.g., EMD [1]) or min-cost flow algorithm (e.g., FastEMD [19]). However, the high complexity of simplex and min-cost flow algorithms drives us to develop an optimal method for efficient image retrieval. In this section, we proceed to develop a more efficient algorithm to obtain the value of EMD^+ .

In the definition of EMD^+ (Definition 2), there is a very important condition, $d_{ij} < \mu$, which restricts that each bin in one histogram can only match a small part of bins in another histogram (Figure 1). Taking advantage of this characteristic, we derive a linear approximate algorithm to compute EMD^+ . We will explain our basic idea by an example (Figure 2).

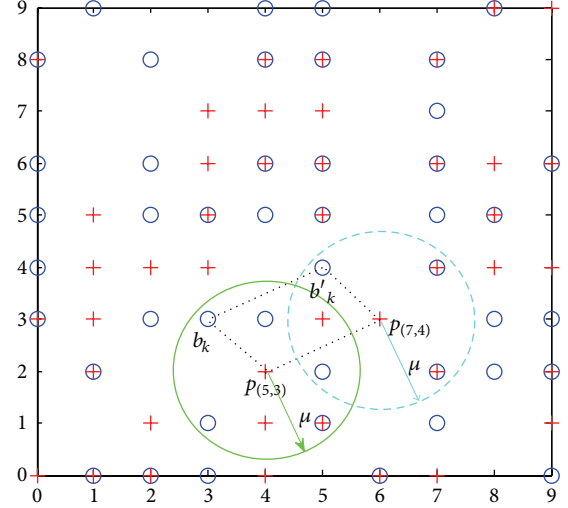


FIGURE 2: Red crosses indicate the distribution of histogram P and blue circles indicate the distribution of histogram Q . The radii of the green solid and cyan dash circles equal the threshold μ .

Input: μ : threshold; d : dimensionality of histogram

Output: Qualified bin set B

- (1) $B \leftarrow \emptyset$; // qualified bin set
- (2) $b \leftarrow (0)^d$; // reference bin, if $d = 2$, $b = (0, 0)$
- (3) $B \leftarrow \text{getBins}(b, B, \mu, d, 1)$; // Algorithm 2
- (4) $\text{Sort}(B)$;

ALGORITHM 1: Get the qualified bin set of reference bin.

In Figure 2, it is clear that only the bins that fall into the solid or dash circle have the qualification of receiving mass from $p_{(5,3)}$ or $p_{(7,4)}$, respectively. We call these bins as *qualified bins*. If b_k is one qualified bin of $p_{(5,3)}$, it is easy to get the qualified bin of $p_{(7,4)}$ at the corresponding position by paralleling b_k . Inspired by this heuristic, we present to precompute all the possible qualified bins of a reference bin (e.g., $(0, 0)$) and then utilize the possible qualified bins to improve the computation of EMD^+ .

To precompute the qualified bins, we propose a novel method which is composed of function *QualifiedBins* (Algorithm 1) and *getBins* (Algorithm 2) (in Algorithm 2, function *getBins*() uses L_p -norm as the ground distance, in other cases it is distance dependent). *QualifiedBins* first calls the function *getBins* to enumerate all possible qualified bins of the reference bin (e.g., $(0, 0)$) (line 3) and then sorts the bins in ascending order according the distances to the reference bin (line 4). Let b_k be the k th qualified bin to reference bin. For each bin p_i , we can get its k th possible qualified bin by paralleling b_k . Notice that the reference bin itself does not appear in qualified bin set; the reason will be shown later.

Let K be the size of qualified bin set B , $K \ll 2^d \cdot \mu^2$ (d is the dimensionality of histogram). Thus the *getBins*() (Algorithm 2) can be done in $O(K)$ time and *Sort*() (line 4) can be done in $O(K \log K)$ time. In fact, K is very small corresponding to the size of histogram and can be treated as

```

Input:  $b$ : one qualified bin;  $B$ : qualified bin set;
 $\mu$ : threshold;  $d$ : dimensionality;
curr: current dimension.
Output: Qualified bin set  $B$ 
(1) for  $i = \text{curr}$  to  $d$  do
(2)    $b' \leftarrow b$ ;
(3)   if  $b'[i] = 0$  then
(4)      $b'[i] \leftarrow b'[i] - 1$ ;
(5)     if  $\|b'\|_p < \mu$  then
(6)        $B.add(b')$ ;
(7)        $\text{getBins}(b', B, \mu, d, i)$ ;
(8)      $b' \leftarrow b, b'[i] \leftarrow b'[i] + 1$ ;
(9)     if  $\|b'\|_p < \mu$  then
(10)       $B.add(b')$ ;
(11)       $\text{getBins}(b', B, \mu, d, i)$ ;
(12)   else
(13)     if  $b'[i] < 0$  then
(14)        $b'[i] \leftarrow b'[i] - 1$ ;
(15)     else
(16)        $b'[i] \leftarrow b'[i] + 1$ ;
(17)     if  $\|b'\|_p < \mu$  then
(18)        $B.add(b')$ ;
(19)        $\text{getBins}(b', B, \mu, d, i)$ ;

```

ALGORITHM 2: Recursive function getBins, Enumerate all possible qualified bins.

a constant. Thus the time of Algorithm 1 can be ignored (in our experiments (Section 5), the time of precomputing qualified bins is no more than 0.001 second).

Next thing is to compute the value of EMD^+ , that is, the core of our idea. The details are depicted in the following.

- (1) The first step is to saturate all the zero-cost flows, since matching corresponding bins does not cause cost. After this step, many bins will be empty (Figure 2).
- (2) The second step focuses on cross-bin matching. We plan to transport as much mass as possible when the cost is as low as possible. We employ the qualified bin set B generated by Algorithm 1 to accelerate the speed of our algorithm. We traverse B from front to end. In the k th iteration, we match each nonempty bin p_i to its corresponding k th qualified bin. If the k th qualified bin of p_i exists, we transfer the highest possible mass; otherwise, we do nothing. Since corresponding bins are already handled in the first step, we have no necessity to store the reference bin in the qualified bin set B .

Algorithm 3 gives the pseudocode of our algorithm. The proposed algorithm is named as *LinearEMD*, because its complexity is linear. The very important parameter B stores the qualified bins of original point (e.g., (0, 0) for 2-dimensional histogram). Notice that B only needs to be computed once.

LinearEMD is composed of 3 parts. Part 1 focuses on the corresponding bins (lines 3~12). Since corresponding bin has no cost ($d_{ij} = 0$), we should move as much mass as possible from P to Q . Then we delete the empty bins (lines 9~12). Once part 1 is done, P and Q will shrink dramatically.

```

Input:  $P, Q$ : histogram;  $\mu$ : threshold;  $B$ : qualified bins
of original point.
Output: Distance between  $P$  and  $Q$ 
(1) cost = 0;
(2)  $\omega = 0$ ;
    // Part 1, corresponding bin
(3) foreach  $p \in P$  do
(4)   if corresponding  $q$  exists then
(5)     mass  $\leftarrow \min(w_p, w_q)$ ; //  $w_p$ : weight of bin  $p$ 
(6)      $\omega \leftarrow \omega + \text{mass}$ ;
(7)      $w_p \leftarrow w_p - \text{mass}$ ;
(8)      $w_q \leftarrow w_q - \text{mass}$ ;
(9)     if  $w_p = 0$  then
(10)      remove  $p$  from  $P$ ;
(11)     if  $w_q = 0$  then
(12)      remove  $q$  from  $Q$ ;
    // Part 2, cross bin
(13) for  $i = 1$  to  $B.size$  do
(14)   foreach  $p \in P$  do
(15)     if  $i$ th qualified bin  $q$  exists then
(16)       mass  $\leftarrow \min(w_p, w_q)$ ;
(17)        $\omega \leftarrow \omega + \text{mass}$ ;
(18)       cost  $\leftarrow \text{cost} + \text{mass} \times \text{distance}(p, q)$ ;
(19)        $w_p \leftarrow w_p - \text{mass}, w_q \leftarrow w_q - \text{mass}$ ;
(20)       if  $w_p = 0$  then
(21)        remove  $p$  from  $P$ ;
(22)       if  $w_q = 0$  then
(23)        remove  $q$  from  $Q$ ;
    // Part 3, compute the  $\text{EMD}^+$ 
(24)  $\text{EMD}^+ \leftarrow \text{cost} + (\max(w(P), w(Q)) - \omega) \times \mu$ ;

```

ALGORITHM 3: EMD^+ between two histograms.

The time complexity of part 1 is $O(N)$ where $N = \min(|P|, |Q|)$. Part 2 adopts greedy idea to match cross bins (lines 13~23). The outer loop (line 13) guarantees that nearest bins get preferential treatment and each bin in P is shifted to the same position. Ideally, many bins in P or Q will be empty after several iterations. The speed will be faster and faster. Assume the average of bin numbers in each iteration is L ($L \ll N$) and the size of B is K . K is a constant. The complexity of part 2 is $O(KL) = O(L)$. Part 3 which gets the result of EMD^+ by a simple calculation has an $O(1)$ expensive (line 31). Therefore, the total cost is $O(N + L) = O(N)$ ($L \ll N$).

5. Experimental Evaluation

This section aims to experimentally test and verify the following things.

- (1) LinearEMD is efficient enough for massive image retrieval.
- (2) Thresholding the ground distance can get a more effective result.
- (3) The result set obtained by LinearEMD is very close the exact value.

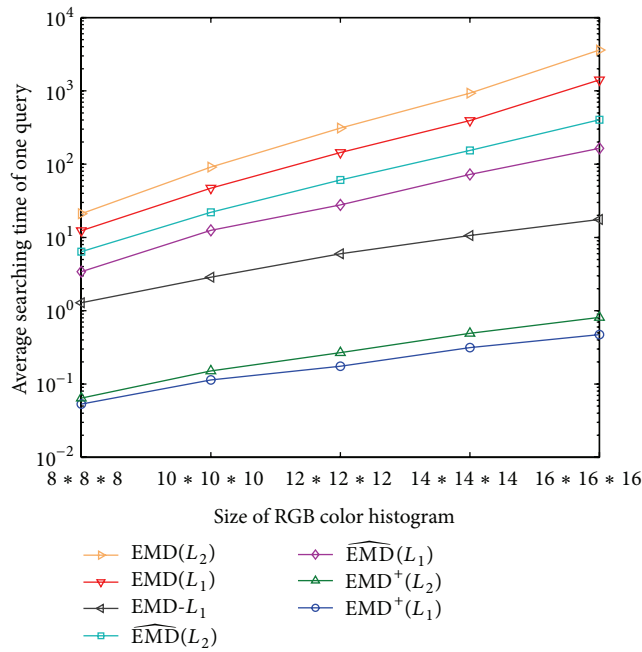


FIGURE 3: The average searching time of each query.

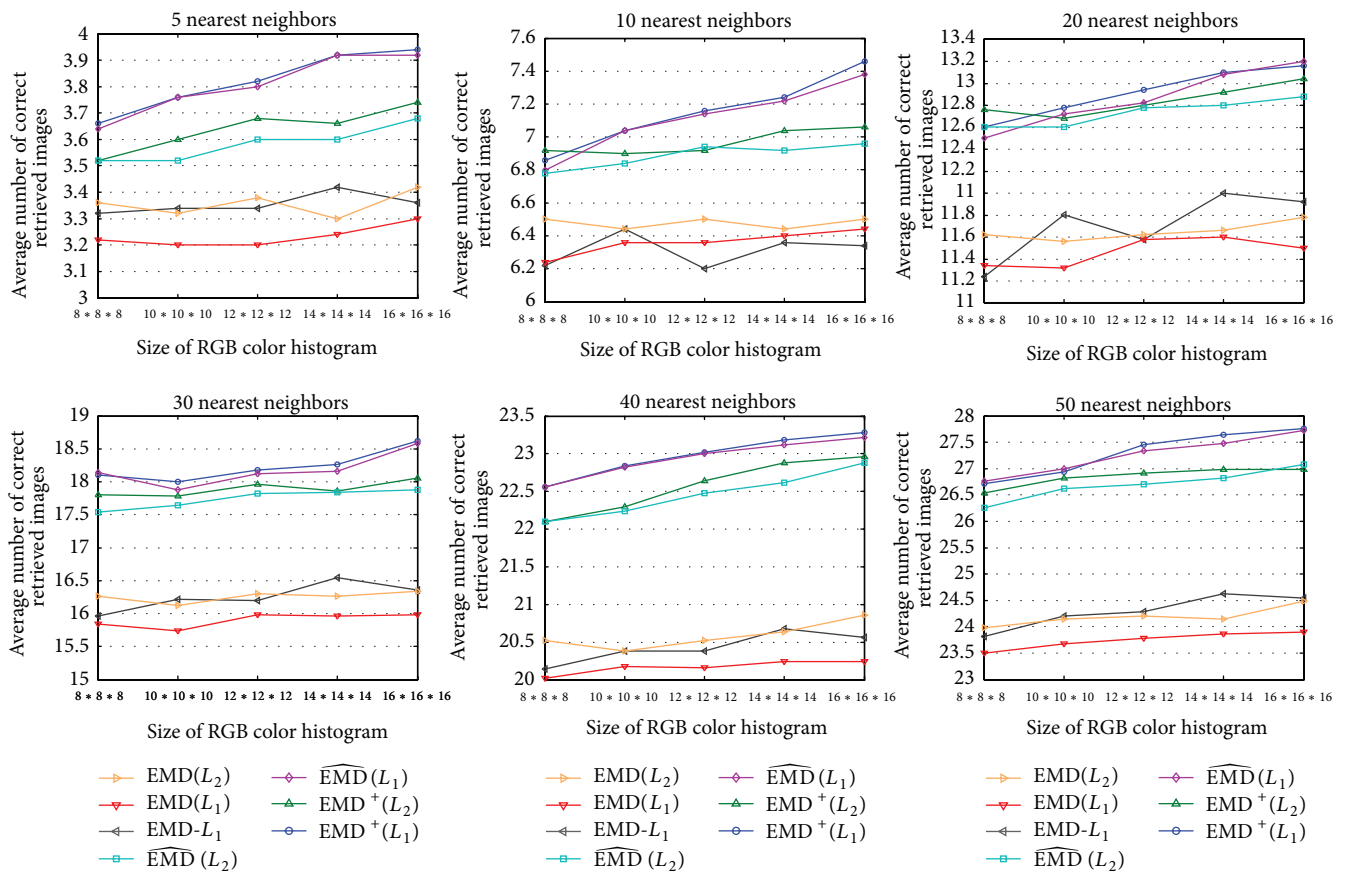


FIGURE 4: Results for image retrieval.

To do that, we use image retrieval as an example to perform the experiments. Section 5.1 gives the setup of our experiments and Section 5.2 describes some evaluation criteria. The experimental results are shown in Section 5.3.

5.1. Experimental Setup. Wang's dataset [23] is a benchmark for image retrieval. This dataset contains 1,000 color images in 10 classes. Each class has 100 images. Since there are some ambiguous images in this dataset, we use a subset of Wang's dataset filtered by [19] (we downloaded the filtered image list from <http://www.seas.upenn.edu/ofirpele/FastEMD/>). This subset is composed of 773 images. We exact the first 5 images from each classe as the queries. We perform our experiments on 5 different size RGB color histograms, that is, $8 * 8 * 8$, $10 * 10 * 10$, $12 * 12 * 12$, $14 * 14 * 14$, and $16 * 16 * 16$, respectively. We use RGB color histogram because it is simple and has the ability to estimate the experimental goals. We do not claim that RGB color histogram is the best descriptor for image retrieval.

We compare $\widehat{\text{EMD}}^+$ with $\widehat{\text{EMD}}$ (in this section, when we say $\widehat{\text{EMD}}$, we mean $\widehat{\text{EMD}}$ with thresholded ground distance), $\text{EMD}-L_1$, and EMD . To compute $\widehat{\text{EMD}}^+$, we use LinearEMD algorithm (we implement this algorithm in C++). For $\text{EMD}-L_1$, we use the authors implementation [10] (we downloaded the code from http://www.dabi.temple.edu/hbling/code_data.htm/). For $\widehat{\text{EMD}}$, we use FastEMD [19] (we downloaded the code from <http://www.seas.upenn.edu/ofirpele/FastEMD/code/>). For EMD , we also use FastEMD [19], since simplex is very expensive. The ground distance of $\text{EMD}-L_1$ is, of course, L_1 -norm. For others, we choose both L_1 - and L_2 -norm. For $\widehat{\text{EMD}}^+$ and $\widehat{\text{EMD}}$, the ground distance thresholds are 3.0, 3.5, 4.0, 4.5, and 5.0 corresponding to the size of RGB color histogram, respectively. We set the coefficient θ of $\widehat{\text{EMD}}$ to be equal to 1. Since $\widehat{\text{EMD}}^+$ equals $\widehat{\text{EMD}}$, the value of $\widehat{\text{EMD}}^+$ (computed by LinearEMD) is approximate value and the value of $\widehat{\text{EMD}}$ (computed by FastEMD) is the exact one.

The experiments are run on a computer with one Intel Xeon E5645 2.40 GHz CPU, 8 GB memory. The OS is CentOS release 6.3 and gcc version is 4.4.6.

5.2. Evaluation Criteria. This section introduces some criteria to fairly evaluate the performance of these distance measures and algorithms.

To evaluate the speed, we compare the average runtime of each query. The effectiveness is measured by the average number of correct retrieved images. We define the mean overlap rate (MOR) to calculate the common images retrieved by LinearEMD and FastEMD. MOR is the average Jaccard coefficient:

$$\text{MOR} = \frac{100\%}{n} \sum_{i=1}^n \frac{|R_A^i \cap R_B^i|}{|R_A^i \cup R_B^i|}, \quad (10)$$

where n is the number of query, R_A^i is the result set of the i th query retrieved by algorithm A, and R_B^i is the result set of the i th query retrieved by algorithm B.

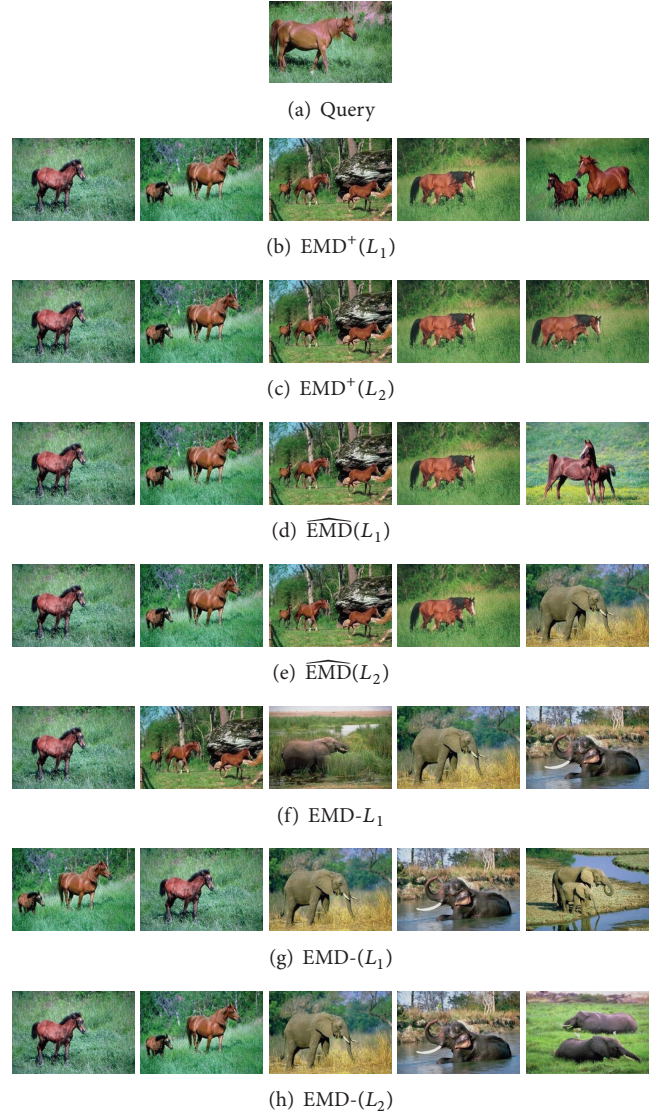


FIGURE 5: An example of retrieved images. The image on the top is the query. (b)–(h) are the 5 nearest neighbors retrieved by different methods. The last two images in (c) look very similar, but they are not the same image. The first 4 images in (b)–(d) are the same.

To measure how close the values computed by two methods are, we employ the symmetric mean absolute percentage error (SMAPE) [24]. SMAPE is defined as follows:

$$\text{SMAPE} = \frac{100\%}{n} \sum_{i=1}^n \frac{|A_i - B_i|}{(A_i + B_i)/2}, \quad (11)$$

where n is the number of common retrieved images and A_i and B_i are the values of the i th image computed by algorithms A and B.

5.3. Results. Firstly, we compare the speed of each method. Figure 3 shows the average running time of each query. Obviously, our algorithm exceeds other methods in both ground distances. Our algorithm is faster by two to three orders of magnitude faster than EMD, two orders of magnitude faster

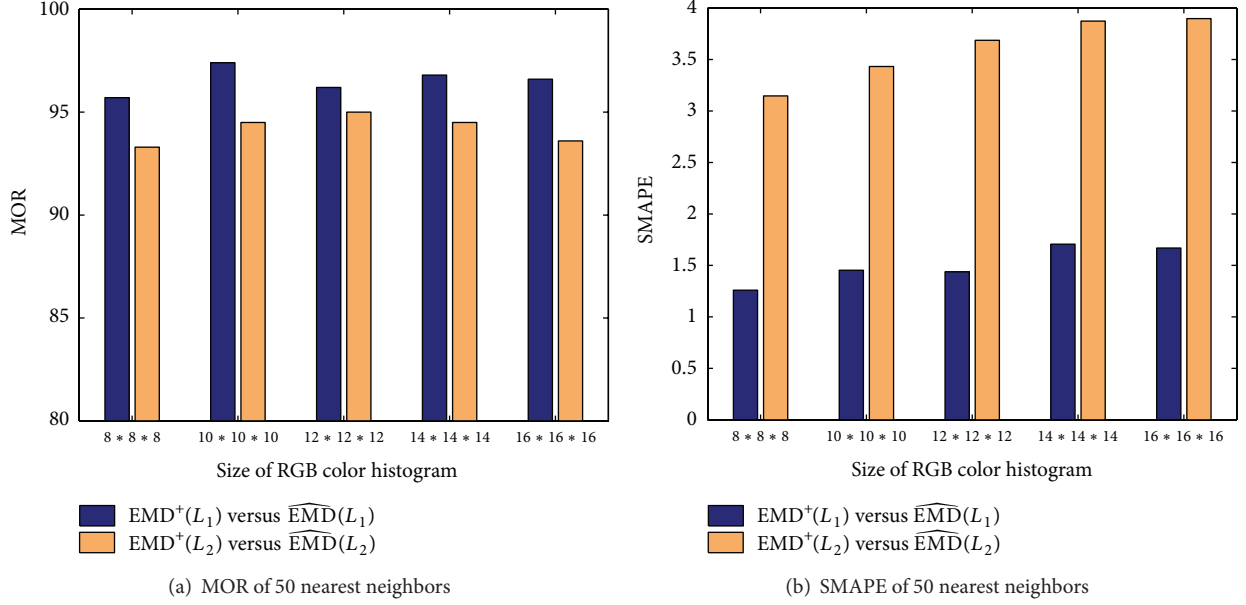


FIGURE 6: The MOR and SMAPE between our algorithm and FastEMD.

than $\widehat{\text{EMD}}$ and an order of magnitude faster than $\text{EMD-}L_1$. The larger the size of RGB color histogram, the better our algorithm. When the size of RGB color histogram is $16 * 16 * 16$, the LinearEMD is 300 to 500 times faster than the FastEMD and nearly 3000 to 4500 times faster than the EMD. Therefore, the LinearEMD is competent enough to handle massive image retrieval task.

The second thing of this section is to compare the retrieved images of each method. Figure 4 shows the average number of correct retrieved images of each query. It is obvious that thresholding the ground distance of EMD (i.e., EMD^+ and $\widehat{\text{EMD}}$) can obtain better retrieval result. That is, EMD^+ and $\widehat{\text{EMD}}$ are more robust and have more discernibility. EMD^+ and $\widehat{\text{EMD}}$ with L_1 norm as ground distance can achieve better results than L_2 -norm as ground distance. However, for EMD, L_2 as ground distance retrieves more images. Generally, $\text{EMD}^+(L_1)$ ($\text{EMD}^+(L_p)$ denotes EMD^+ with L_p -norm as the ground distance; $\widehat{\text{EMD}}(L_p)$ and $\text{EMD}(L_p)$ have similar meanings) gets the best results and $\text{EMD}^+(L_2)$ is better than $\widehat{\text{EMD}}(L_2)$. In addition, Figure 4 illustrates the precision of EMD^+ and $\widehat{\text{EMD}}$ will be improved while the size of histogram is increased. Although our method is the approximate algorithm, the precision of our method is no less, even higher, than the exact algorithm (FastEMD). Figure 5 shows an example of retrieved images. The images retrieved by $\text{EMD}^+(L_1)$, $\text{EMD}^+(L_2)$, and $\widehat{\text{EMD}}(L_1)$ are all correct, yet the last image retrieved by $\widehat{\text{EMD}}(L_2)$ is incorrect. However, there are only 2 correct images in the results retrieved by $\text{EMD-}L_1$, $\text{EMD}(L_1)$, and $\text{EMD}(L_2)$, respectively. This result shows the advantage of thresholding the ground distance once again.

The third thing is to count the common images retrieved by our algorithm and the FastEMD. To this end, we employ the mean overlap rate (MOR). Figure 6(a) shows the MOR

of the two 50 nearest neighbors obtained by our algorithm and the FastEMD (we only count the correct retrieved images). The $\text{MOR}(L_1)$ s ($\text{MOR}(L_p)$ denotes the MOR value of LinearEMD and FastEMD with L_p -norm as the ground distance; $\text{SAMPE}(L_p)$ has the same meaning.) are between 95% and 98% and the $\text{MOR}(L_2)$ s are between 93% and 95%. Although the $\text{MOR}(L_2)$ s are a little bit lower than the $\text{MOR}(L_1)$ s, they are still very high. Figure 6(a) indicates that the results between our algorithm and the FastEMD share many common images.

Finally, we compute the SMAPE of our algorithm and the FastEMD. Like the MOR, we use the two 50 nearest neighbors and only count the correct retrieved images. The SMAPE results are shown in Figure 6(b). Obviously, the $\text{SMAPE}(L_1)$ s are lower than the $\text{SMAPE}(L_2)$ s. In spite of this, the maximum value is lower than 4%. Figure 6(b) illustrates that the values computed by our approximate algorithm highly approach the values computed by the FastEMD.

In one word, our approximate algorithm is an optimal choice for comparing images.

6. Conclusion

In this paper, we have presented a new variant of EMD with thresholded ground distance, named as EMD^+ . We proved that EMD^+ is a metric for any kind of histogram. To compute the value of EMD^+ , we have proposed a linear approximate algorithm. This algorithm is an effective and efficient method for image retrieval. Experimental results show that our method break through the comparing methods. And thus, our method is competent enough for massive image retrieval task. Our method may also be able to compare other histograms (e.g., HSV color histogram) and descriptors (e.g., SIFT).

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.
- [2] Q. Lv, M. Charikar, and K. Li, "Image similarity search with compact data structures," in *Proceedings of the 13th ACM Conference on Information and Knowledge Management (CIKM '04)*, pp. 208–217, Washington, DC, USA, November 2004.
- [3] O. Pele and M. Werman, "A linear time histogram metric for improved sift matching," in *Computer Vision*, D. Forsyth, P. Torr, and A. Zisserman, Eds., vol. 5304 of *Lecture Notes in Computer Science*, pp. 495–508, Springer, Berlin, Germany, 2008, Proceedings of the 10th European Conference on Computer Vision, 2008.
- [4] V. Ljosa, A. Bhattacharya, and K. A. Singh, "Indexing spatially sensitive distance measures using multi-resolution lower bounds," in *Advances in Database Technology*, Y. Ioannidis, M. H. Scholl, J. W. Schmidt et al., Eds., vol. 3896 of *Lecture Notes in Computer Science*, pp. 865–883, Springer, Berlin, Germany, 2006, Proceedings of the 10th International Conference on Extending Database Technology, 2006.
- [5] H. K. Tan and C. W. Ngo, "Common pattern discovery using earth mover's distance and local flow maximization," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, vol. 2, pp. 1222–1229, Beijing, China, October 2005.
- [6] H. K. Tan and C. W. Ngo, "Localized matching using earth mover's distance towards discovery of common patterns from small image samples," *Image and Vision Computing*, vol. 27, no. 10, pp. 1470–1483, 2009.
- [7] S. Marinai, B. Miotti, and G. Soda, "Using earth mover's distance in the bag-of-visual-words model for mathematical symbol retrieval," in *Proceedings of the 11th International Conference on Document Analysis and Recognition (ICDAR '11)*, pp. 1309–1313, Beijing, China, September 2011.
- [8] K. Grauman and T. Darrell, "Fast contour matching using approximate earth mover's distance," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 1, pp. 1220–1227, July 2004.
- [9] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using affine-invariant regions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, vol. 2, pp. 319–324, June 2003.
- [10] H. Ling and K. Okada, "An efficient earth mover's distance algorithm for robust histogram comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 840–853, 2007.
- [11] M. F. Demirci, A. Shokoufandeh, S. J. Dickinson, Y. Keselman, and L. Bretzner, "Many-to-many feature matching using spherical coding of directed graphs," in *Proceedings of the 8th European Conference on Computer Vision (ECCV '04)*, vol. 1, pp. 322–335, 2004.
- [12] D. Xu and S. F. Chang, "Visual event recognition in news video using kernel methods with multi-level temporal alignment," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'07)*, pp. 1–8, Minneapolis, Minn, USA, June 2007.
- [13] F. Wang and L. J. Guibas, "Supervised earth mover's distance learning and its computer vision applications," in *Computer Vision*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds., vol. 7572 of *Lecture Notes in Computer Science*, pp. 442–455, Springer, Berlin, Germany, 2012, Proceedings of the 12th European conference on Computer Vision, 2012.
- [14] D. Xu, S. Yan, and J. Luo, "Face recognition using spatially constrained earth mover's distance," *IEEE Transactions on Image Processing*, vol. 17, no. 11, pp. 2256–2260, 2008.
- [15] I. Leichter, "Mean shift trackers with cross-bin metrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 695–706, 2012.
- [16] B. E. Rutenber and A. K. Singh, "Indexing the earth mover's distance using normal distributions," *Proceedings of the VLDB Endowment*, vol. 5, no. 3, pp. 205–216, 2011.
- [17] J. Xu, Z. Zhang, A. K. H. Tung, and G. Yu, "Efficient and effective similarity search over probabilistic data based on earth mover's distance," *The VLDB Journal*, vol. 21, no. 4, pp. 535–559, 2011.
- [18] F. Hillier and G. Lieberman, *Introduction to Mathematical Programming*, McGraw-Hill, New York, NY, USA, 2nd edition, 1977.
- [19] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 460–467, Kyoto, Japan, October 2009.
- [20] A. Rahimi and R. Kiram, "How earth mover's distance comprises two bags," Tech. Rep., Intel Labs Berkeley, 2007.
- [21] S. Shirdhonkar and D. W. Jacobs, "Approximate earth mover's distance in linear time," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.
- [22] M.-H. Jang, S.-W. Kim, C. Faloutsos, and S. Park, "A linear-time approximation of the earth mover's distance," in *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM '11)*, pp. 505–514, Scotland, UK, October 2011.
- [23] J. Z. Wang, J. Li, and G. Wiederhold, "Simplicity: semantics-sensitive integrated matching for picture libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947–963, 2001.
- [24] B. E. Flores, "A pragmatic view of accuracy measurement in forecasting," *Omega*, vol. 14, no. 2, pp. 93–98, 1986.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

