Circular Earth Mover's Distance for the comparison of local features

Julien Rabin, Julie Delon and Yann Gousseau
Telecom ParisTech, LTCI CNRS
46 rue Barrault, 75013 Paris, France
{rabin,delon,gousseau}@telecom-paristech.fr

Abstract

Many computer vision algorithms make use of local features, and rely on a systematic comparison of these features. The chosen dissimilarity measure is of crucial importance for the overall performances of these algorithms and has to be both robust and computationally efficient.

Some of the most popular local features (like SIFT *[4] descriptors) are based on one-dimensional circular histograms. In this contribution, we present an adaptation of the Earth Mover's Distance to one-dimensional circular histograms. This distance, that we call* CEMD*, is used to compare SIFT-like descriptors. Experiments over a large database of 3 million descriptors show that* CEMD *outperforms classical bin-to-bin distances, while having reasonable time complexity.*

1. Introduction

Many computer vision applications rely on the comparison of local features between images. For instance, object pose estimation [4], image stitching [2] or image classification [9] are quite often based on either descriptor matching or clustering. In a comparative study [5], SIFT-like descriptors [4] have proven to be the most effective and robust methods for image matching. Such descriptors consist of several histograms of gradient orientation, and the comparison of two descriptors boils down to the comparison of one-dimensional circular[1] histograms. Observe that many other features based on one-dimensional circular histograms can be found in the literature, such as hue histograms [7].

In practice, "bin-to-bin" distances, like the Euclidean distance [4, 5] or the χ^2 distance [1, 9], are considered as the simplest way to measure quickly the dissimilarity between two histograms at a low computational cost. However, these distances are obviously not robust to histogram quantization. Therefore, the number of bins of gradient orientation histograms for original SIFTs [4] is limited to $N = 8$ to make a compromise between the discriminative power of the descriptor and the robustness of the representation.

This quantization problem can be avoided by using cross-bin distances, such as the Earth Mover's Distance, initially proposed by Rubner [6] as a metric for multi-dimensional histograms. This distance, often used to compare image signatures, is known to be more robust than bin-to-bin distances, but is computationally far more expensive. A nice variant of this distance is proposed in [3] by Ling *et al.* in order to speed up the comparison. However, this measure remains too expensive to be applied to the matching problem when the number of features increases (see Section 3) and does not address the circularity of orientation histograms.

These limitations led us to propose a new dissimilarity measure, called CEMD, specifically designed to compare one-dimensional circular histograms (see Section 2). This measure, based on the Earth Mover's Distance (as a solution of a transportation problem), is computationally efficient. We show in Section 4 that it behaves well with respect to histogram quantization and outperforms classical bin-to-bin distances for the comparison of SIFT-like features.

2. Circular Earth Mover's Distance (CEMD)

Let $f = (f[i])_{i=1\dots N}$ and $g = (g[i])_{i=1\dots N}$ be two discrete histograms with samples on N bins and normalized, in the sense that $\sum_{i=1}^{N} f[i] = \sum_{i=1}^{N} g[i] = 1$. The Earth Mover's Distance between f and g is defined in [6] as

$$\text{EMD}(f,g) := \min_{(\alpha_{i,j})\in\mathcal{M}} \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_{i,j} c(i,j), \quad (1)$$

[1] Circular means that the first and the last bins of the histogram are neighbors.

where $\mathcal{M} = \{(\alpha_{i,j});\ \alpha_{i,j} \geq 0,\ \sum_j \alpha_{i,j} = f[i],\ \sum_i \alpha_{i,j} = g[j]\}$ and where $c(.,.)$ is a ground distance between bins.

The distance $\text{EMD}(f,g)$ can be understood as a transportation cost. The value $c(i,j)$ measures the cost of moving a unit mass from bin i to bin j, and $\alpha_{i,j}$ is the amount of mass carried from i to j. This definition can be used in any dimension, but histograms of dimension larger than 2 involve heavy computations.

For non-circular and one-dimensional histograms, if $c(i,j) = |i-j|/N$, it is known [8] that $\text{EMD}(f,g)$ equals $\frac{1}{N}\sum_{i=1}^{N}|F[i] - G[i]|$, where F and G are the cumulative histograms of f and g. The generalization of this formula to circular histograms is not straightforward (remark that in this case $c(i,j) = \min\{|i-j|/N\,,\, 1-|i-j|/N\}$). Observe in particular that if f is a circular histogram, one can build as many cumulative histograms as there are bins in f (any bin can be chosen as a starting point). However, if f and g are circular and one-dimensional, it can be shown (the proof of this result is omitted in this paper for lack of space) that the Earth Mover's Distance between them equals

$$\text{CEMD}(f,g) = \min_{k\in\{1,\ldots,N\}} \left\{ \frac{1}{N} \sum_{i=1}^{N} |\, F_k[i] - G_k[i]\,| \right\}, \tag{2}$$

where, $\forall\ k \in \{1,\ldots,N\}$ (the definition is similar for G_k by replacing f by g),

$$F_k[i] = \begin{cases} \displaystyle\sum_{j=k}^{i} f[j] & \text{if} \quad i \geq k \\ \displaystyle\sum_{j=k}^{N} f[j] + \sum_{j=1}^{i} f[j] & \text{if} \quad i < k \end{cases}.$$

This means that the distance $\text{CEMD}(f,g)$ is the minimum in k of the L^1 distance between F_k and G_k, the cumulative histograms of f and g starting at the k^{th} quantization bin.

3 Comparing local features

In this section, we first briefly recall the classical way to compare SIFT-like features by using bin-to-bin distances, and then explain how to apply the CEMD introduced in the previous section to the comparison of such local features.

Let us recall [4, 5] that a SIFT-like descriptor a consists of M circular histograms a_m of gradient orientations, weighted by the gradient magnitude and computed for different subregions of a location grid around an interest point. Thus, the comparison of two descriptors a and b boils down to the comparison of histograms a_m and b_m. We suppose here that each histogram is quantized to N bins and that the whole descriptor $a = (a_1,\ldots,a_M)$ is normalized to have unit weight [4].

Bin-to-bin distances The most classical way to compare SIFT-like descriptors is simply to use the L^p distance as in Formula (3), usually with $p = 2$ (Euclidean distance) [4]. Applying this distance requires a global L^p normalization of descriptors a and b. Other bin-to-bin distances that are used to compare local features include the χ^2 distance, as in [9] or the Jeffrey divergence. The definitions of these distances in the framework of SIFT-like descriptors are recalled in Formula (4) and (5) respectively.

$$\text{D}_{L^p}(a,b) := \left(\sum_{m=1}^{M}\sum_{i=1}^{N} |\, a_m[i] - b_m[i]\,|^p \right)^{\frac{1}{p}} \tag{3}$$

$$\text{D}_{\chi^2}(a,b) := \sum_{m=1}^{M}\sum_{i=1}^{N} \frac{(a_m[i]-b_m[i])^2}{a_m[i]+b_m[i]} \tag{4}$$

$$\text{D}_{J}(a,b) := \sum_{m=1}^{M}\sum_{i=1}^{N} a_m[i]\log\left(\frac{2\,a_m[i]}{a_m[i]+b_m[i]}\right) + b_m[i]\log\left(\frac{2\,b_m[i]}{a_m[i]+b_m[i]}\right) \tag{5}$$

Applying CEMD to local features In order to apply CEMD to SIFT-like features, Formula (2), designed to compare normalized histograms, should be applied to the comparison of each histogram pair a_m and b_m. In practice, however, it is by far more robust to globally normalize SIFT-like features to unit weight (as in [4]) than to normalize each histogram a_m individually. Therefore, Formula (2) is applied to non-normalized histograms.

In order to combine distances corresponding to different subregions (different values of m) we choose to use the following distance between two descriptors,

$$D_{\text{CEMD}}(a,b) := \sum_{m=1}^{M} \text{CEMD}(a_m,b_m). \tag{6}$$

Other dissimilarity measures could have been chosen (such as $\sum \text{CEMD}(a_m,b_m)^2$ or $\max \text{CEMD}(a_m,b_m)$). However, we observed experimentally that the distance (6) is more robust.

Implementation and computational cost Let $X_k[i] = F_k[i] - G_k[i]$ be the difference of the cumulative histograms computed in Formula (2). X_k can be written as a function of X_1 (with the convention $X_1[0] = 0$), $\forall\, k \in \{1, \ldots, N\}$

$$X_k[i] = \begin{cases} X_1[i] - X_1[k-1] & \text{if } i \geqslant k \geqslant 1 \\ X_1[i] - X_1[k-1] + X_1[N] & \text{if } i < k \end{cases}$$

Observe that $X_1[N] = 0$ when f and g are two normalized histograms. Thus, CEMD only necessitates the computation of histograms F_1 and G_1 (just like EMD in the non-circular case), and the minimization of $\|X_k\|_1$ according to k. The complexity of the CEMD computation is therefore approximately N times the complexity of the Euclidean distance computation, where N is the number of bins of each local histogram ($N = 8$ for classical SIFT).

Ling and Okada in [3] proposed a faster implementation of EMD, called EMD-L_1, with the L^1 ground distance in the multidimensional case. In one of their experiments, EMD-L_1 is used to compare SIFT descriptors, considered as 3-dimensional histograms (coding both orientation and localization). However, they do not address the circular aspect of orientation histograms. Moreover, this distance remains empirically too expensive to be applied to large descriptors databases: computing EMD-L_1 is empirically 720 times slower than the Euclidean distance, according to Table VII in [3]. As an order of magnitude, performing the same evaluation as in Section 4 with EMD-L_1 would require more than one year on a standard 2.5 GHz computer.

4. Experiments

This section compares the performances of the different distances defined in Section 3, when used for the matching of SIFT-like descriptors. These descriptors are obtained in a way similar to the original SIFT [4], except that they are extracted from circular masks divided in $M = 9$ subregions. In order to estimate the robustness of the distances to histogram quantization, each circular histogram is computed twice, first with $N = 8$ then with $N = 12$.

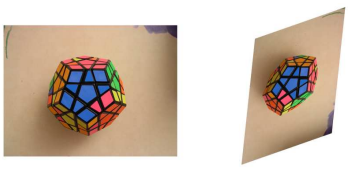

Figure 1. Example of affine original image A (left) and its affine transform A' (right).

The performances of the different distances are compared on a set of 732 images[2] and 3.1 million descriptors. Each image A of the database is compared to an image A', obtained by applying an affine transform (see Figure 1) and adding Gaussian noise (with $\sigma = 5$ for 8-bit coded images) to A. Since our purpose is to compare distances, and since each descriptor a of A should have at most one correct match in A', we choose the simplest criterion to match descriptors: a query descriptor a of A is matched with its nearest neighbor a' in A' if $D(a, a')$ is smaller than a threshold τ.

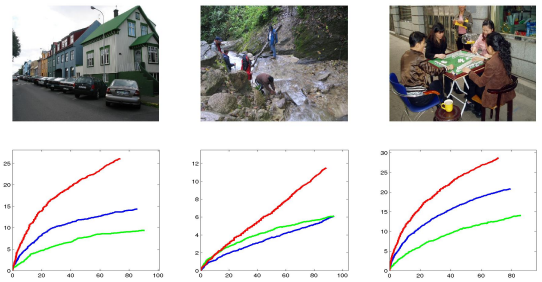

Figure 2. Example of three images and ROC curves from the 732 image database. The red curve corresponds to CEMD, the blue one to the L^1 distance and the green one to the L^2 distance.

Now, a match is declared false (*i.e.* false positive) or correct (*i.e.* true positive) depending on some spatial tolerance (following exactly the same protocol as in [5]) on the relative position of a and a'. For each image A of the database and for each distance D, we obtain a ROC curve showing the ratio of correct matches as a function of the ratio of false matches for different values of the threshold τ on the dissimilarity measure D. More precisely, the ratios of correct matches and false matches are defined as

$$\text{correct matches ratio} = \frac{\#\text{correct matches}}{\#\text{possible matches}},$$

$$\text{false matches ratio} = \frac{\#\text{false matches}}{\#\text{total number of matches}}.$$

Some images and associated ROC curves are shown in Figure 2 -for the sake of clarity, only CEMD, L^1 and L^2 distances are represented, respectively in red, blue and green continuous lines. We can see on these curves that results can be quite different from one experiment to the other. In order to compare the performances of the different distances on the whole database, we choose to draw *average ROC curves*, obtained by averaging all 732 ROC curves, weighted by the number of descriptors

[2] Images available at: http://www.tsi.enst.fr/~rabin/ICPR08/

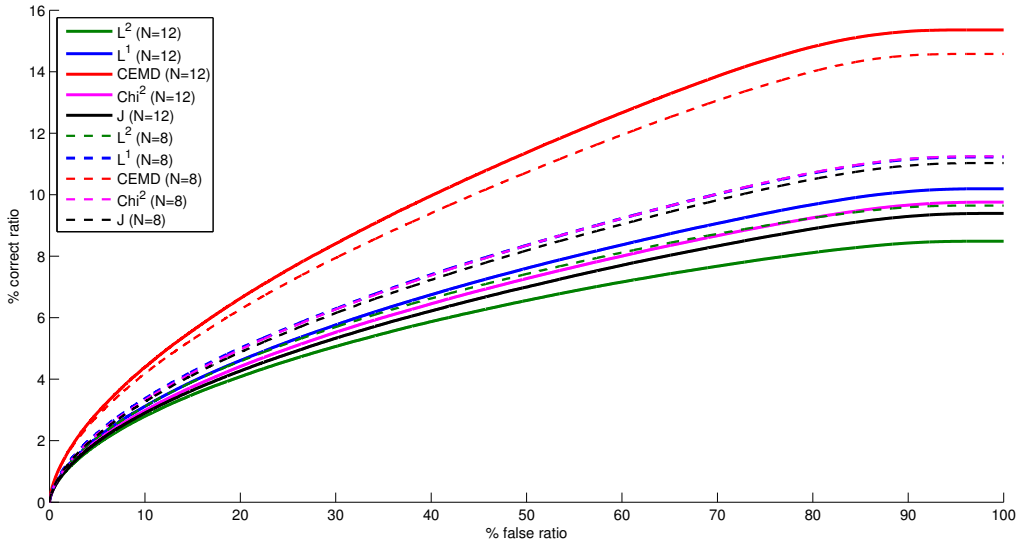

Figure 3. Average ROC curves: (on 732 images and 3.1 million descriptors) for CEMD (red), L^1 (blue), L^2 (green), χ^2 distance (magenta) and Jeffrey divergence (black), with two different quantization steps ($N = 8$ for dashed lines and $N = 12$ for continuous lines).

of each image. Thus, for each distance defined in Section 3, performances are evaluated on the set of 732 images and $3.1\,10^6$ descriptors (involving approximately 25.10^9 descriptor comparisons).

In Fig. 3, the efficiency of the L^1 and L^2 distances, Jeffrey divergence, and χ^2 distance are compared with the proposed CEMD, for two different quantization steps ($N = 8$ and $N = 12$). The average ROC curves clearly show the advantage of CEMD for all quantization choices. Moreover, one observes that increasing N systematically increases the quality of the matching when using CEMD. The number of bins is therefore only driven by computational complexity. This is of course not the case for classical bin-to-bin distances, for which using too many bins yields inefficient comparisons between histograms.

5. Conclusion

In this paper, we propose a new dissimilarity measure called CEMD between circular histograms, relying on an adaptation of Earth Mover's Distance to the circular case. We show, when applied to SIFT descriptors, that this distance clearly outperforms other classical bin-to-bin distances on a large database, while involving low time complexity.

Acknowledgments

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant BLAN07-2_183172, Optimal transport: Theory and applications to cosmological reconstruction and image processing (OTARIE).

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. PAMI*, 24(4):509–522, 2002.

[2] M. Brown, R. Szeliski, and S. Windner. Multi-image matching using multi-scale oriented patches. In *Proc. CVPR*, pages 510–517, 2005.

[3] H. Ling and K. Okada. An efficient Earth Mover's distance algorithm for robust histogram comparison. *IEEE Trans. PAMI*, 29(5):840–853, may 2007.

[4] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[5] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. PAMI*, 27(10):1615–1630, 2005.

[6] Y. Rubner, C. Tomasi, and L. J. Guibas. The Earth Mover's distance as a metric for image retrieval. *IJCV*, 40(2):99–121, 2000.

[7] R. Venkatesh Babu, P. Pérez, and P. Bouthemy. Robust tracking with motion estimation and local kernel-based color modeling. *Image and Vision Computing*, 25(8):1205–1216, 2007.

[8] C. Villani. *Topics in optimal transportation*. American Math. Soc., 2003.

[9] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *IJCV*, 73(2):213–238, 2007.