A Statistical Approach to the Matching of Local Features

Julien Rabin, Julie Delon, Yann Gousseau

A Statistical Approach to the Matching of Local Features

Julien Rabin, Julie Delon, and Yann Gousseau

Abstract—This paper focuses on the matching of local features between images. Given a set of query descriptors and a database of candidate descriptors, the goal is to decide which ones should be matched. This is a crucial issue, since the matching procedure is often a preliminary step for object detection or image matching. In practice, this matching step is often reduced to a specific threshold on the Euclidean distance to the nearest neighbor.
Our first contribution is a robust distance between descriptors, relying on the adaptation of the Earth Mover's Distance to circular histograms. It is shown that this distance outperforms classical distances for comparing SIFT-like descriptors, while its time complexity remains reasonable. Our second contribution is a statistical framework for the matching procedure, which yields validation thresholds automatically adapted to the complexity of each query descriptor and to the diversity and size of the database. The method makes it possible to detect multiple occurrences, as well as to deal with situations where the target is not present. Its performances are tested through various experiments on a large image database.

Index Terms—Statistical analysis of matching processes, local feature matching, dissimilarity measure, Earth Mover's Distance, a contrario.

1 Introduction

The matching of common structures between digital images is an important issue for a large number of computer vision applications: finding correspondences between images of the same scene [1], image classification [2], image and video retrieval [3], [4], [5], image stitching [6], [7], stereo vision [8], [9], object detection [10] and recognition [11], [12], and 3D object modeling [13]. One of the most classical approaches to this problem consists in using local features around interest points or regions. The locality of the features ensures robustness to occlusion or context change, while the coding of the features should be invariant or robust to various geometrical and radiometrical changes. Numerous local approaches have been proposed in the literature, the exhaustive study of which is beyond the scope of the present paper. In two relatively recent comparative studies [14], [15], the SIFT descriptor [11] has proven to be one of the most robust and invariant representation methods. As a result, the problem of finding correspondences between images often boils down to the matching of such local features. Nevertheless, whereas the extraction and representation of local descriptors has been thoroughly studied (see *e.g.* the references in [14]), their matching has not been the object of a systematic study. In practice, the matching step relies on simple but somehow limited procedures, as detailed further in the paper.

In many applications, this matching procedure is yet a crucial preliminary step. It can for instance be used as a pre-processing stage (before resorting to some geometric consistency algorithm like RANSAC [5], [6], [16] or some mean square error minimization [11]) for finding common objects between images. The matching step is at the core of many recent methods relying on image similarities, see e.g. [3], [5], [6], [7], [8], [9], [11], [12], [13], [16], [17], [18], [19], [20]. At this point, it is worth noticing that this matching step can serve to localize common structures between images, but also to *decide whether a structure is present*. In fact, this is a crucial issue since a computer vision system has to deal with situations where the object of interest is not present. In such cases, it is of great interest to be able to limit the number of false matches, especially in the case of very large databases, see e.g. [3].

Now, as pointed out in [15],

> *Important aspects of matching are metrics and criteria to decide whether two features should be associated, and data structures and algorithms for matching efficiently.*

Indeed, matching features involves two important steps:

- the choice of a dissimilarity measure between features;
- the choice of a matching criterion, used to decide which matches are valid.

The dissimilarity measure should provide relevant comparisons between features and should be robust enough to cope with small variations of these features. The matching criterion should adapt itself to the complexity and diversity of the features. These two aspects (dissimilarity measure and matching criterion) are at the core of this paper. Our first contribution is a dissimilarity measure relying on the adaptation of the Earth Mover's

• *J. Rabin, J. Delon and Y. Gousseau are with the LTCI CNRS, TELECOM ParisTech, Paris France.*
E-mail: {rabin,delon,gousseau}@telecom-paristech.fr

Distance [21] to circular histograms. This measure is proven to behave well with respect to histogram quantization and to outperform classical bin-to-bin distances in the framework of local features comparison. Our second contribution is a matching criterion relying on a statistical framework. This criterion provides thresholds which adapt to the complexity of the features and allow multiple detections over a database, while controlling the total number of matches. In particular, this criterion deals well with situations where we do not know whether the object of interest is present, as will be demonstrated by a specific experimental protocol. Conference proceedings versions of this work have appeared in [22] (dissimilarity measure) and [23] (matching criterion).

1.1 Related works

Dissimilarity measure

As previously mentioned, the choice of a metric is fundamental for the matching of local features. Indeed, the matching criteria that are commonly used (as detailed in the next paragraph) directly rest on a thresholding of the similarity score.

The most classical local features, such as SIFT [11], reduce the geometrical information to one-dimensional circular histograms of local orientations. Usually, "bin-to-bin" distances, such as the Euclidean distance [3], [5], [11], [14], [15] or the χ^2 distance [10], [19], are considered as the simplest way to quickly measure the dissimilarity between such histograms at a low computational cost. The term "bin-to-bin" refers to the fact that, to compare two histograms, each bin of the first histogram is compared exclusively to the bin of same rank of the second histogram. These distances are obviously not robust to histogram quantization. Therefore, one has to choose the number of bins to reach a good compromise between discriminative power and robustness of the comparison. For instance, the number of bins of gradient orientation histograms for original SIFTs [11] is limited to $N = 8$.

Bin-to-bin distances are intrinsically limited since they only compare the intensity of modes and not their relative positions. This limitation can be overcome by using cross-bin distances. A classical cross-bin distance, the Mahalanobis distance, requires the computation of the covariance matrix of descriptors over a training database. This distance has been used in the context of local features comparison, but without meaningful gain: as pointed out in [15], *although the Mahalanobis distance is more general than the Euclidean distance, most relative performances were not modified.* Other cross-bin distances, such as the so-called quadratic distance [24] or the diffusion distance [25], rely on smoothings of the histograms. These methods necessitate non-trivial parameter adjustments, such as the choice of a kernel or the scale of smoothings.

The Earth Mover's Distance, proposed by Rubner *et al.* [21] and often used to compare image signatures, is probably one of the most elegant and robust ways of comparing histograms. However, it is computationally far more expensive than bin-to-bin distances when the dimension of histograms becomes strictly greater than one. A nice variant of this distance has been proposed by Ling *et al.* [26] as a way to speed up the comparison. This distance is applied in [26] to the comparison of local features. However, this measure remains too expensive to be applied to the matching problem when the number of features increases (as will be detailed in Section 2) and does not explicitely 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, is computationally efficient and deals with circular histograms, such as orientation histograms in SIFT descriptors [11] for instance. In the experimental section, CEMD is used for the comparison of SIFT descriptors. This distance is shown to be more robust to quantization effects and small geometric perturbations than bin-to-bin distances.

Matching criterion

In order to introduce the most classical criteria that are used to match local descriptors, it is useful to give some vocabulary and notations that will be used throughout this paper. We consider a situation where one seeks for correspondences between N_Q *query descriptors* $\{a^i\}$ and a database of N_C *candidate descriptors* $\{b^j\}$. We assume that distances have been computed between each a^i and each b^j. This step can sometimes be replaced by approximate allocation algorithms, as in [27]. Two different criteria are used in practice to validate matches, as detailed in [14], [15], both relying on user-selected thresholds. Ideally, these thresholds should be set automatically and should depend on both the query and candidate descriptors.

The simplest matching criterion, that we call **DT** (Distance Threshold), relies on a global threshold on distances. That is, each query a^i is simply matched with candidates $\{b^j\}$ that are at a distance $d(a^i, b^j)$ smaller than the threshold. Usually, matches are restricted to the nearest neighbor [7], [17] for each query descriptor, in order to limit multiple false detections that often affect some query descriptors. We will refer to this criterion as **NN-DT** (Nearest Neighbor Distance Threshold). Three main drawbacks inherent to this approach restrict its use in practice. First, the nearest neighbor restriction limits the number of correct matches so that, in some applications, one prefers to select the K nearest neighbors: $K = 3$ in [12], $K = 4$ in [6] for image stitching, and K between 5 and 10 in [5]. The price to pay is then a higher proportion of false matches. Secondly, the nearest neighbor restriction is also problematic in cases where there are multiple occurrences of the structure of interest, for instance when the target object is present more than once in the database (see for instance [28]), when dealing with objects having repetitive parts, such as buildings (this issue is studied in [20]), or when the

interest point detector yields spurious repetitions of the structure to be coded. Lastly, the great variability of distances between descriptors from images to images (as shown in Section 4) makes it particularly difficult to set the right threshold for a particular application.

In order to reduce the variability of the chosen threshold, Lowe [11] introduces another criterion by comparing the distances between a^i and its closest and second-closest neighbors respectively. If the ratio between the two distances is below a threshold r, the match with the closest neighbor is validated. This popular criterion, that we call **NN-DR** (Nearest Neighbor Distance Ratio), benefits from its simplicity and the fact that it is by far more robust than a simple threshold on distances. However, the choice of the "optimal" threshold r is strongly dependent on both the application and the database: $r = 0.8$ in [11], $r = 0.6$ in [3], $r = 0.95$ in [16], or r between 0.56 and 0.7 in [15] for instance. In practice, the NN-DR criterion behaves very well (and in particular significantly better than the NN-DT criterion as shown in [15]) when the target to be matched is present exactly once in the candidate database. Indeed, in this case, it makes sense to assume that the distance to the nearest neighbor is small compared to distances to other candidates and in particular to the second nearest neighbor. Now, the reason why this criterion should work when the structure of interest is not present is less clear. This situation will be considered in the experimental section. It is of great practical importance, because in real situations a computer vision system relying on the matching of local features has to deal with situations when the target is present as well as with situations when the target is missing. Moreover, this criterion is by nature limited to the nearest neighbor, and, as NN-DT, may fail in the case where the structures of interest appear more than once, as already mentioned.

Several variants of these matching criteria have been proposed. In [29], it is suggested to adapt the NN-DR criterion by averaging the distance to the second neighbor over several images for panorama stitching. In [9], a variant of NN-DT consists in keeping only matches (a, b) for which a is also the nearest neighbor of b. More specific matching criteria with geometric constraints have been proposed (see e.g. [13], [18], [20], [30]), but to the best of our knowledge, no generic procedure for the matching of local, SIFT-like features has been proposed beyond the aforementioned thresholds on distances.

In this paper, we propose in Section 3 an alternative matching criterion relying on adaptive thresholds. Matches between the query and candidate descriptors are validated by rejecting casual matches, that is matches that can be produced by chance. Similar ideas are present in works dealing with the statistical analysis of object recognition processes [31], [32]. Specifically, we resort to an *a contrario* methodology, first introduced in [33] and then applied, among other things, to shape matching [34]. This approach provides thresholds on the dissimilarity measure that adapt to the query and candidate descriptors. This matching procedure also allows multiple detections over a database, while controlling the total number of matches, in particular in cases where the structure of interest is not present.

1.2 Outline

In Section 2, we introduce the new transportation distance for comparing local descriptors, CEMD. Then in Section 3, the matching criterion relying on the *a contrario* methodology is introduced. In Section 4 the advantages of both contributions over classical approaches are demonstrated on an image database through the use of several experimental protocols.

2 Dissimilarity Measure

In this section, we introduce a dissimilarity measure designed to compare circular histograms (such as orientation histograms). This measure is a generalization of the classical Earth Mover's Distance to the circular case. It can also be seen as an application of the statistical Mallows distance to probability distributions on the unit circle. We then explain how to apply this measure to compare local, SIFT-like features.

2.1 Circular Earth Mover's Distance (CEMD)

Consider two discrete circular[1] histograms $f = (f[i])_{i=1\dots N}$ and $g = (g[i])_{i=1\dots N}$, sampled on N bins. Both histograms are supposed to be normalized, that is, $\sum_{i=1}^{N} f[i] = \sum_{i=1}^{N} g[i] = 1$.

The Earth Mover's Distance between f and g is then defined in [21] as

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

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. For circular histograms, this ground distance can naturally be chosen as

$$c(i, j) = \frac{1}{N} \min(|i - j|, N - |i - j|),\ \forall (i, j) \in \{1, \dots N\}^2.$$

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. However the computation of the Earth Mover's Distance involves heavy computations when the dimension of histograms becomes larger than two. Note that this distance is known by statisticians

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

as the Mallows distance between probability distributions [35], and is also one of the Monge-Kantorovich distances, defined in the mass transportation theory [36].

Now, for non-circular and one-dimensional histograms, when the ground distance is chosen as $c(i,j) = \frac{1}{N}|i-j|$, it is well known (see for instance chapter 2 in [36] for a proof) that $\text{EMD}(f,g)$ equals $\|F-G\|_1 = \frac{1}{N}\sum_{i=1}^{N}|F[i]-G[i]|$, where F and G are the cumulative histograms of f and g, defined as

$$F[i] = \sum_{j=1}^{i} f[j], \quad G[i] = \sum_{j=1}^{i} g[j]. \tag{2}$$

The generalization of this formula to circular histograms is not straightforward. Indeed, if f is a circular histogram, one can build as many cumulative histograms as there are bins in f, since any bin can be chosen as a starting point to cumulate the histogram. However, if f and g are circular and one-dimensional, it can be shown[2] that the (circular) Earth Mover's Distance between them equals

$$\begin{aligned} \text{CEMD}(f,g) &= \min_{\mu\in[-1,1]} \|F-G-\mu\|_1 &(3)\\ &= \frac{1}{N}\min_{\mu\in[-1,1]} \sum_i |F[i]-G[i]-\mu|, &(4)\end{aligned}$$

where F and G are defined as in Formula 2. Observe that this minimum is very easy to compute. Indeed, the function $\mu \mapsto \sum_i |F[i]-G[i]-\mu|$ reaches its minimum at a (not necessarily unique) median of the values $F[i]-G[i]$, $i = 1,\dots N$. It follows that

$$\text{CEMD}(f,g) = \frac{1}{N}\min_{k\in\{1,\dots N\}} \sum_i |F[i]-G[i]-F[k]+G[k]|. \tag{5}$$

Observe also that Formula 3 remains valid if F and G are replaced by two cumulative histograms of f and g starting from another bin. Any starting bin can be chosen and the result does not depend on this choice.

We now establish an alternative formula for $\text{CEMD}(f,g)$. For this, we define F_k and G_k, the cumulative histograms of f and g starting at the k^{th} quantization bin. For each k in $\{1,\dots,N\}$

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

The definition is similar for G_k by replacing f by g. Then, $F_k[i] = F[i] - F[k-1]$ if $i \geq k$ (with the convention $F[0]=0$) and $F_k[i] = F[i] + 1 - F[k-1]$ if $i < k$. Thus, by observing that $F[0]-G[0] = F[N]-G[N] = 0$,

$$\begin{aligned} &\text{CEMD} \quad (f,g)\\ &= \frac{1}{N}\min_{k\in\{1,\dots N\}} \sum_i |F[i]-G[i]-F[k]+G[k]|\\ &= \frac{1}{N}\min_{k\in\{1,\dots N\}} \sum_i |F[i]-G[i]-F[k-1]+G[k-1]|\\ &= \frac{1}{N}\min_{k\in\{1,\dots N\}} \sum_i |F_k[i]-G_k[i]|. \end{aligned}$$

Finally,

$$\text{CEMD}(f,g) = \min_{k\in\{1,\dots N\}} \|F_k - G_k\|_1. \tag{6}$$

This means that the distance $\text{CEMD}(f,g)$ is also 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.

2.2 Comparing SIFT-like 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 [11], [14] 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 circular histograms a_m and b_m. We suppose here that each histogram is quantized to N bins and that the whole descriptor $a = (a_1,\dots,a_M)$ is normalized to have unit weight [11].

2.2.1 Bin-to-bin distances

The most classical way to compare SIFT-like descriptors is simply to use the L^p distance as in Formula (7), usually with $p = 2$ (Euclidean distance) [11]. 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 [10] or the Jeffrey divergence. The definitions of these distances in the framework of SIFT-like descriptors are recalled in Formula (8) and (9) respectively.

$$\text{D}_{L^p}(a,b) := \left(\sum_{m=1}^{M}\sum_{i=1}^{N} |\, a_m[i]-b_m[i]\,|^p\right)^{\frac{1}{p}} \tag{7}$$

$$\text{D}_{\chi^2}(a,b) := \sum_{m=1}^{M}\sum_{i=1}^{N} \frac{(a_m[i]-b_m[i])^2}{a_m[i]+b_m[i]} \tag{8}$$

$$\begin{aligned} \text{D}_J(a,b) := \sum_{m=1}^{M}\sum_{i=1}^{N} & a_m[i]\log\left(\frac{2\,a_m[i]}{a_m[i]+b_m[i]}\right)\\ & + b_m[i]\log\left(\frac{2\,b_m[i]}{a_m[i]+b_m[i]}\right) \end{aligned} \tag{9}$$

2. The proof is provided for the reviewers as supplementary material.

2.2.2 Applying CEMD to local features

Two descriptors $a = (a_1, \ldots, a_M)$ and $b = (b_1, \ldots, b_M)$ are compared by applying CEMD to each pair of histograms a_m and b_m using Formula (3) or (6). Theoretically, this distance should be applied to normalized histograms. In practice, however, it is by far more robust to globally normalize SIFT-like descriptors to unit weight (as shown in [11]) than to normalize each histogram a_m individually. This means that we need to compute distances between histograms of different weights. Now, Formula (6) and (3) are not equivalent anymore in this case. We use preferably Formula (6) which is always independent of the choice of the origin of histograms, that is, of the choice of the orientation values stored in the first bin. This is not the case of Formula (3) when it is used to compare non-normalized histograms.

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

$$D_{\text{CEMD}}(a, b) := \sum_{m=1}^{M} \text{CEMD}(a_m, b_m). \tag{10}$$

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 (10) is more robust.

2.2.3 Implementation and computational cost

Let $X_k[i] = F_k[i] - G_k[i]$ be the difference of the cumulative histograms computed in Formula (3). X_k can be written as a function of X_1,

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

Consequently, computing CEMD does not require to compute the k different cumulative histograms F_k and G_k in the circular case. Note that $X_1[N]$ is equal to zero when the two histograms f and g have the same weight. Compared to the classical L^1 bin-to-bin distance, the only required extra computation is the minimization over k of $\|X_k\|_1$, the L^1 norm of X_k. It follows that the complexity of the CEMD computation is 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).

Observe that in [26], Ling and Okada present an interesting variant of the Earth Mover's Distance, called EMD-L_1, designed to speed up the computation of EMD in the multidimensional case. Among their experiments, they show an application of their distance to SIFT descriptors, considered as three dimensional histograms (coding both orientation and localization). Nevertheless, this distance remains too expensive to be applied to large descriptors databases: EMD-L_1 is empirically 720 times slower than computing the Euclidean distance, according to Table VII in [26]. As an order of magnitude, performing the same evaluation as the one to be done in Section 4 with EMD-L_1 would require more than one year on a standard 2.5 GHz computer.

3 *A contrario* MATCHING CRITERION

In this section, we introduce a generic way to compute matching thresholds in the framework of local, SIFT-like descriptors. Recall that we consider N_Q query descriptors $\{a^i\}$ and N_C candidate descriptors $\{b^j\}$. The question is then: for each a^i, to which b^j (if any) should it be matched ? To answer this question, we rely on the general principle of *a contrario* methods and fix matching thresholds that ensure the rejection of casual matches.

3.1 *A contrario* methodology

The general principles of *a contrario* methods have first been proposed by Desolneux *et al.* [33] in order to detect alignments. The same principles have then been applied to a wide variety of computer vision tasks, such as the detection of contrasted edges, good continuation, vanishing points, rigid transforms or motion, see the recent monograph [37]. The main idea, presented in a generic manner in [38], is to detect groups of features that are very unlikely under the hypothesis that these features are *independent*. This hypothesis is called the *null hypothesis* in this paper. Loosely speaking, the detected groups are those that cannot result from chance. The second important point of *a contrario* methods is that to compute the degree of unlikeliness of a group, one predicts the expected number of groups under the null hypothesis, and not the (generally intractable) probability of existence of the group, see [37].

Recently, this methodology has been adapted to the problem of shape matching [34]. Again, the main idea is to reject matches that could have occurred by chance. Similar ideas are present in studies dealing with the statistical analysis of matching processes [31], [32], [39], in particular when predicting the number of false alarms. One difference is that these studies are more elaborated, but also less generic, because the analysis of the matching process relies on some shape model. When using *a contrario* approaches, one only needs a distance and an independence assumption (the null hypothesis) to validate matches. In the next two paragraphs, this methodology is adapted to the matching of SIFT-like features.

3.2 The null hypothesis

Recall that each descriptor a^i is made of M orientation histograms, $a^i = (a^i_1, \ldots, a^i_M)$. In order to define the null hypothesis, we assume that the distance between two descriptors a^i and b can be written as $D(a^i, b) = \sum_{m=1}^{M} d(a^i_m, b_m)$. Observe that this is a very mild assumption, satisfied for classical bin-to-bin distances (Euclidean, Manhattan or χ^2), as well as for the Circular Earth Mover's Distance, CEMD, introduced in

this paper. Given a random descriptor b, we then define the following null hypothesis,

$\mathcal{H}_0^i$: *"$d(a_m^i, b_m)$ ($m \in \{1, \dots M\}$) are mutually independent random variables".*

Under this hypothesis, the probability that the distance between a^i and b is smaller than a given threshold δ is

$$\mathbb{P}\left(D(a^i, b) \leq \delta \,|\, \mathcal{H}_0^i\right) = \int_{-\infty}^{\delta} \underset{m=1}{\overset{M}{*}} p_m^i(x)\, dx \,, \qquad (11)$$

where $*$ denotes the convolution product and p_m^i the probability density function of the random variable $d(a_m^i, b_m)$. The validity of a match will then be decided by thresholding this probability, as explained in the next section. This in turn yields thresholds on distances that depend on both a^i and the observed distribution of candidate descriptors.

In order to numerically compute the probability given by Equation (11), we need to estimate the probability density functions p_m^i. For this, we simply use histograms of realizations of the distances over the database. That is, for each $i \in \{1, \dots N_Q\}$ and each $m \in \{1, \dots M\}$, the law p_m^i is empirically estimated over the database $\{b^1, \dots, b^{N_C}\}$.

3.3 Meaningful matches

Let us consider two descriptors a^i and b^j at distance $\delta = D(a^i, b^j)$. We decide to match these descriptors as soon as $\mathbb{P}(D(a^i, b) \leq \delta \,|\, \mathcal{H}_0^i)$ is small enough. It therefore remains to automatically fix a threshold on this probability. Following the general approach of *a contrario* methods, we choose the threshold in order to control the average number of false detections. Since $N_Q N_C$ comparisons are performed when searching for matches between descriptors $\{a^i\}$ and $\{b^j\}$, we define the following threshold on distances, for a given $\epsilon > 0$,

$$\delta_i(\varepsilon) = \arg\max_{\delta} \left\{ \mathbb{P}\left(D(a^i, b) \leq \delta \,|\, \mathcal{H}_0^i\right) \leq \frac{\varepsilon}{N_Q N_C} \right\}. \qquad (12)$$

A match between a^i and some b^j is then said to be ε-meaningful if $D(a^i, b^j) \leq \delta_i(\varepsilon)$.

The reason behind this choice is the following: *when testing N_Q queries against N_C candidates satisfying the null hypotheses, the expected number of ε-meaningful matches is smaller than ε.*

This result is a simple consequence of the linearity of the mathematical expectation. Observe that it would have been much more difficult to bound the *probability* of false detections, since distances between different descriptors are not necessarily independent. A more in-depth analysis of this interesting aspect can be found in [37]. Let us also remark that this choice of δ is actually one of the most simple approaches to multiple testing, and is known in the statistical community as a Bonferonni correction [40]. In practice, for a fixed ϵ and for each descriptor a^i we perform the following steps

1) Probability density functions p_m^i of distances $d_m(a^i, b^j)$ are estimated by histograms of these distances when b^j spans the database;
2) $\delta \mapsto \mathbb{P}\left(D(a^i, b) \leq \delta \,|\, \mathcal{H}_0^i\right)$ is computed using Formula (11) ;
3) the threshold $\delta_i(\varepsilon)$ is automatically computed in function of the value ε using Formula (12) ;
4) for each descriptor b^j ($j = 1, \dots, N_C$), a^i is matched with b^j if $D(a^i, b^j) \leq \delta_i(\varepsilon)$.

From now on, we will refer to this matching criterion as AC. Let us now comment on this criterion. First, one needs to fix the value of ε, that in turn yields a threshold on distances. Since this value corresponds to an expected number of false detections, we claim that it is much simpler to set than a threshold on distances. Indeed, it is well known that distances between descriptors vary very much from one descriptor to another or one image to another, as will be illustrated in the experimental section. Now, the threshold on distances computed thanks to step 3) above depends on both the particular descriptor at hand, a^i, and the database (e.g. an image, or a set of images) against which it is matched. This is due both to the learning of marginals p_m^i and to the fact that the number of descriptors is taken into account by Formula (12). In particular, one can hope that the proposed matching criterion works well over a relatively large image database and in the presence of distractors, as will be confirmed by the experimental section. Last, observe also that the number of matches is not restricted to the nearest neighbor, even though one has the possibility to add such a restriction depending on the application.

4 Experimental results

In this section, several experiments are performed on an image database to illustrate the performances of both the dissimilarity measure and the matching criterion introduced in this paper. These experiments are performed on images modified by synthetic degradations (affine transformation and noise). We introduce several experimental protocols to illustrate the behavior of the proposed matching method in cases of single or multiple occurences of the structure of interest, as well as in the presence of distractors.

4.1 Experimental setup

4.1.1 Local features

In this paragraph, we briefly describe the local features that are used for the experiments. These are obtained in a very similar way to the original SIFTs [11]. We first use a combined Laplace and Harris keypoint detector, which provides a set of interest points together with their corresponding scales. We then build a histogram of gradient orientations in a neighborhood of each point and segment it to obtain reference directions. A set of $M = 9$ circular histograms of gradient orientations with respect to the reference direction is then built. These histograms

correspond to 9 disjoint regions of the neighborhood of each interest point. We use a polar localization grid as in [14] (a central region, 4 regions on a first ring and 4 more regions on a second ring).

4.1.2 Experimental protocols

We use several protocols to illustrate the versatility of the proposed matching criterion: ability to detect a structure when we know it is present exactly once, ability to decide whether the structure is present or not and ability to detect multiple occurrences.

The first protocol, called $A \rightarrow A'$, consists in matching keypoints between an image A and an image A' obtained by applying an affine transform and adding Gaussian noise to A (with a standard deviation $\sigma = 5$ for 8-bit images). A match is declared false (*i.e.* a false positive) or correct (*i.e.* a true positive) depending on some spatial tolerance. More precisely, and following the protocol of [14], a match between a and b is considered as correct if the overlap error is below 50 percent. The overlap error between a and b is defined from the ratio between the intersection and the union of the corresponding SIFT regions in the image A, respectively R_a and R_b:

$$1 - (R_a \cap R_b)/(R_a \cup R_b) \, .$$

This classical protocol, $A \rightarrow A'$, measures very simply the behavior of a matching procedure when two images containing exactly the same "objects" (before and after some transformations) are compared.

Now, many real computer vision systems involving a matching step have to deal with situations in which the target is not always present (*e.g.* the search of an object in an image database). In order to estimate matching procedures in such situations, we introduce another protocol called $A \rightarrow \{^{A'}_{B}$. In this protocol, the image A is first compared with the modified image A' and then with an image B, independent of A (the next image in the database to be presented in the next section). Of course, both comparisons are made using the same thresholds. Correct and false matches between A and A' are defined in the same way as in the protocol $A \rightarrow A'$. Meanwhile, all matches between A and B are considered as false matches. The total number of false matches is the addition of false matches in A' and B. A matching procedure should be able to match A and A' without finding too many correspondences between A and B.

In Section 4.3.2, protocol $A \rightarrow \{^{A'}_{B}$ will be extended by replacing B by the entire database to be introduced in the next section. In Section 4.4, another protocol will be introduced to test the ability to detect multiple occurrences.

4.1.3 Performance evaluation

Performances of both the dissimilarity measure and the matching criterion introduced in this paper are evaluated on approximately 3.10^6 descriptors, extracted from a set of 732 generic images[3]. The size of this database is in the same order of magnitude as the one used in the evaluation paper [15], containing 100 query objects and 535 irrelevant images which constitute a 10^5 feature set. In this paper as in ours, an exhaustive feature comparison is performed. The use of such a dataset is of great importance, because performances can vary very much from an image to another. Observe also that using much bigger datasets to perform *exhaustive comparisons* would require quite heavy computing facilities.

As is usually done, for each experiment, a ROC curve shows the ratio of correct matches as a function of the ratio of false matches for different values of the matching threshold. More precisely, for a given threshold, the ratios of correct matches and false matches are defined as

$$\begin{cases} \text{correct matches ratio} = \dfrac{\#\text{correct matches}}{\#\text{possible matches}}, \\ \text{false matches ratio} = \dfrac{\#\text{false matches}}{\#\text{total number of matches}}. \end{cases}$$

Such a curve can be obtained for each image of the database. In order to evaluate the performances of different matching procedures (distances and criteria) on the whole database, thorough comparisons and analyses are made in the next sections, relying on these ROC curves.

4.2 Evaluation of the dissimilarity measures

We compare here the performances of the usual L^1 (Manhattan) distance, L^2 (Euclidean) distance, Jeffrey divergence, and χ^2 distance with the performances of the proposed Circular Earth Mover's Distance (CEMD). Since our purpose in this paragraph is not to evaluate matching criteria, we choose to use a simple threshold on distances restricted to the nearest neighbor (that is, criterion NN-DT) with the $A \rightarrow A'$ protocol. The comparison is performed for two quantization levels ($N = 8$ and $N = 12$) of the circular histograms.

Some images from the database and their associated ROC curves are shown in Fig. 1. For the sake of clarity, only CEMD, L^1 and L^2 distances are represented on these examples, respectively in red, blue and green continuous lines, for the value $N = 12$. We see on these curves that results can be quite different from one image to the other, even though CEMD shows better results than other distances.

Performances of the various distances are thus evaluated on the complete database. We follow the classical protocol used for image retrieval evaluation, see e.g. [21], and draw average performance curves to evaluate the ability of a given distance to retrieve correct information first. Average ROC curves show the average ratio of correct matches as a function of the ratio of false matches. The average correct matches ratio is defined (see (13)) as the average of correct matches ratio for the same given

3. Images available at: http://www.tsi.enst.fr/~rabin/matching/

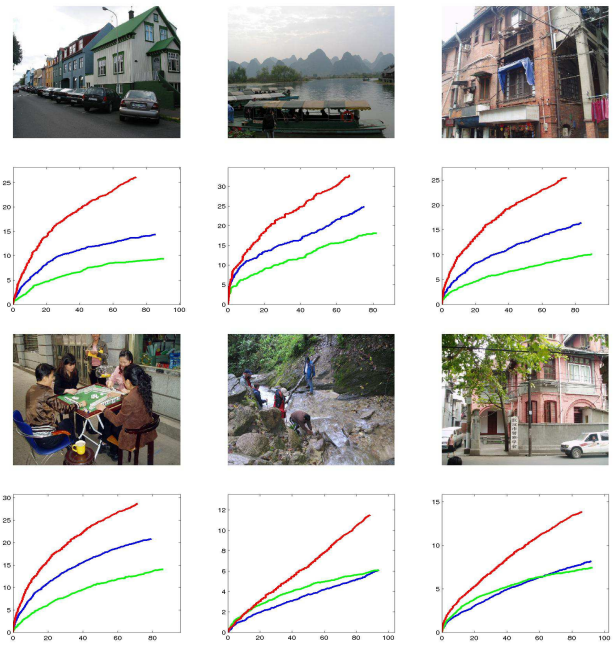


Fig. 1. Six sample images from the database and the corresponding ROC curves. The red curve corresponds to CEMD, the blue one to the L^1 distance and the green one to the L^2 distance.

false matches ratio with each query image A_i, weighted by its number of descriptors $N_{Q,i}$, so that the larger the number of descriptors in an image, the greater its weight in the final average ROC curve.

$$\text{average correct matches ratio} = \frac{1}{\sum_{i=1}^{732} N_{Q,i}} \sum_{i=1}^{732} \left(N_{Q,i} \frac{\#\text{correct matches}(A_i)}{\#\text{possible matches}(A_i)} \right) \quad (13)$$

Consequently, for each distance defined in Section 2.2, performances are evaluated on the database (involving approximately 25.10^9 descriptor comparisons). Observe that curves are quite smooth because of this large number of comparisons.

Fig. 2 clearly shows the advantage of CEMD for all quantization choices. As one could expect, this measure deals well with the geometric deformations applied to each image which induce slight shifts in orientation histograms. Moreover, one observes that increasing N increases the quality of the matching when using CEMD. The number of bins is therefore only driven by computational complexity. In contrast, this is not the case for classical bin-to-bin distances, for which using too many bins yields inefficient comparisons between histograms. The average ROC curve in the case of the $A \rightarrow \{^{A'}_{B}$ protocol shows a similar behavior and is omitted for brevity. We will see in the next paragraphs that, in contrast, matching criteria behave differently depending on the matching protocol.

As previously mentioned in paragraph 2.2.2, EMD could be used to compare descriptors considered as three

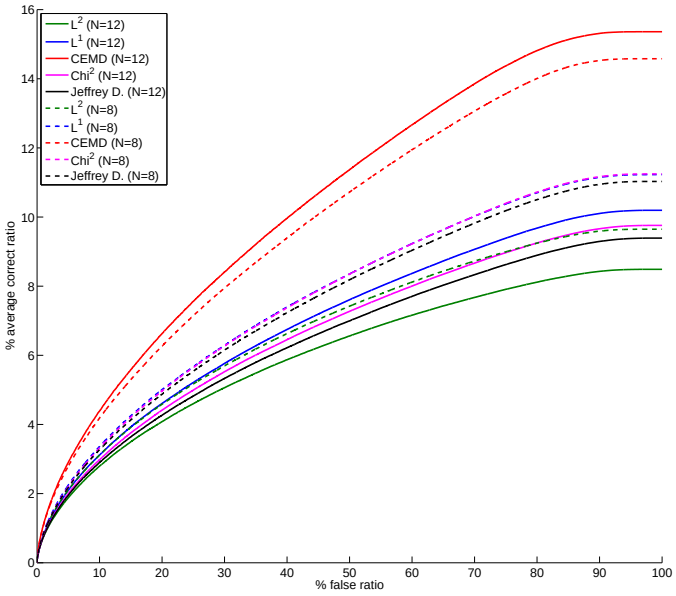


Fig. 2. Average ROC curves (on 732 images) and 3.1 million descriptors for CEMD (red), L^1 (blue), L^2 (green), χ^2 distance (magenta) and Jeffrey divergence (black), with two different quantization levels ($N = 8$ for dashed lines and $N = 12$ for continuous lines).

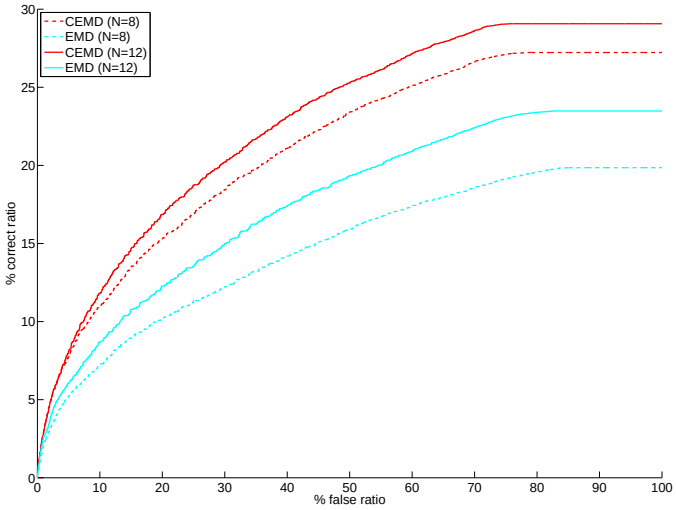


Fig. 3. Average ROC curves (on 10 images) for CEMD (red) and 3-dimensional EMD (cyan), with two different quantization levels ($N = 8$ for dashed lines and $N = 12$ for continuous lines).

dimensional histograms (one dimension for the gradient orientations and two for the location of the region on the polar localization grid). We also saw that such a method involves intractable computation times, even when using the efficient implementation proposed by Ling et al. [26]. Nevertheless, we performed a small scale experiment comparing such a use of EMD and the proposed CEMD on ten images from the database. The 3-dimensional EMD makes use of a ground distance obtained from a circular L^1 distance between orientations histograms and

a L^1 ground distance on the position of regions of the descriptor. We chose to simply add these two ground distances without trying to optimize their combination. We used the EMD code kindly provided by Y. Rubner [21]. Firstly, this implied computation times approximately 1000 times slower than when using CEMD. Secondly, one observes that 3-dimensional EMD, with this choice of ground distance, is less efficient than CEMD.

4.3 Comparison of matching criteria - single match

Three matching criteria are compared in this section. All three criteria *limit matches to the nearest neighbor*, but make use of different thresholds. The first one is a threshold on distances, called NN-DT. The second threshold acts on the ratio between the distance to the nearest neighbor and the distance to the second nearest neighbor, as explained in Section 1.1. This criterion will be called NN-DR. The third criterion, called NN-AC, is the restriction to the nearest neighbor of the new matching criterion introduced in Section 3. Recall that a threshold on distances is obtained by thresholding a probability of false detections (see (12)). For the $A \to A'$ protocol, (12) is applied with $N_Q = N_C = N_A$, and for the $A \to \{^{A'}_{B}$ protocol with $N_Q = N_A$ and $N_C = N_A + N_B$. We use CEMD for all three matching methods.

Some images and associated ROC curves are shown in Fig. 4, both using the $A \to A'$ protocol (second and fifth rows) and the $A \to \{^{A'}_{B}$ protocol (third and sixth rows). In these curves, the NN-AC, NN-DT and NN-DR matching criteria are represented respectively in red, blue and green continuous lines. As in the previous paragraph, we can see that results can be quite different from one image pair to the other.

In order to compare the relative performances of different matching criteria, the same decision thresholds should be used for different query images, as is done in [15]. A *global ROC curve* is thus obtained by plotting the total number of correct matches on the whole database versus the total number of false matches, for different threshold values. Such a curve permits to evaluate how stable a given threshold is from one experiment to the other. The next two paragraphs present and interpret results on the whole database, relying on such curves, respectively for the $A \to A'$ and $A \to \{^{A'}_{B}$ protocols.

4.3.1 Single match. The target is present.

Global ROC curves are displayed on Fig. 5 for the nearest neighbor criteria (namely NN-AC, NN-DT and NN-DR), using the $A \to A'$ protocol. We observe that both NN-AC and NN-DR have very similar global ROC curves, and that the NN-DT criterion is especially unstable. In this case, the NN-AC criterion proposed in this paper does not offer significant advantages in comparison with the classical NN-DR criterion. Indeed, as explained in Section 1.1, the NN-DR criterion is well adapted to the case where the target is present and yields excellent results in the special case of two images A and A' of the same scene,

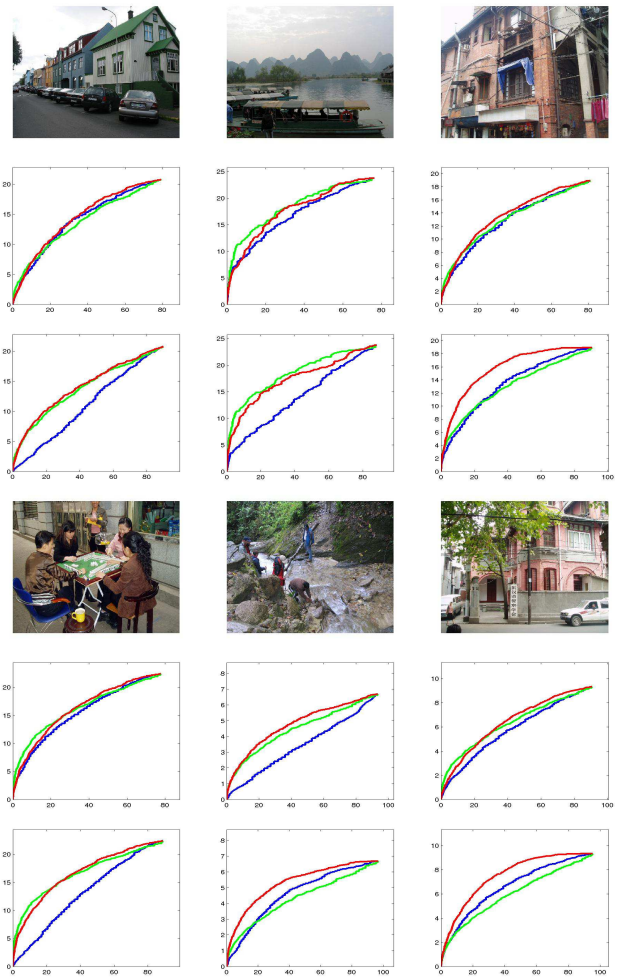


Fig. 4. Six sample images from the database and the corresponding ROC curves. The red curves correspond to NN-AC, the blue ones to NN-DT and the green ones to NN-DR. The second and fifth rows show the curves obtained with the $A \to A'$ protocol, and the third and sixth rows show the results of the $A \to \{^{A'}_{B}$ protocol. Note that the relative performances of the three criteria depend strongly on the experiment.

containing no distractors. Let us remark that we obtain results that are extremely close to the one shown in [15], where the authors obtain a flat global ROC curve for the NN-DT criterion and significant improvement with the NN-DR criterion. This analogy between our results and the ones in [15] also confirms the interest of using relatively large databases.

4.3.2 Single match. Is the target present ?

This section investigates the performances of the matching criteria on the whole database when using the $A \to \{^{A'}_{B}$ protocol. Fig. 6 shows the global ROC curve for this protocol. We can see that the performances of NN-DR clearly decrease in comparison to the ones of the proposed NN-AC criterion. For a given number of correct correspondences between A and A', NN-AC yields fewer

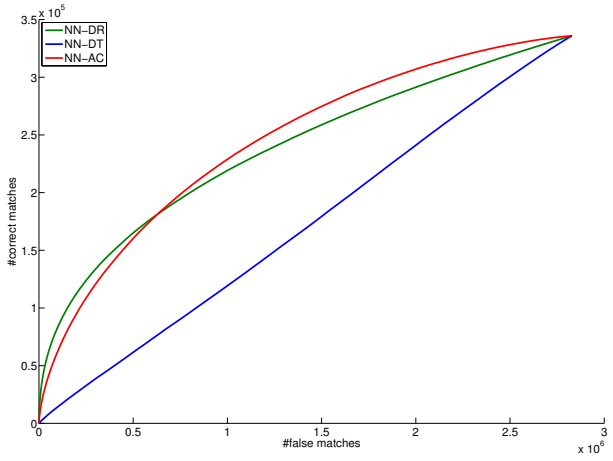


Fig. 5. Global ROC curves (on the whole database) for different matching criteria: NN-AC (red), NN-DT (blue) and NN-DR (green). Experimental protocol is $A \rightarrow A'$ (an image A is matched against its transformed version A').

false correspondences than NN-DR. As explained earlier, this shows the ability of the NN-AC criterion to discriminate between cases where the target is present and cases where it is not, which can be crucial for practical applications.

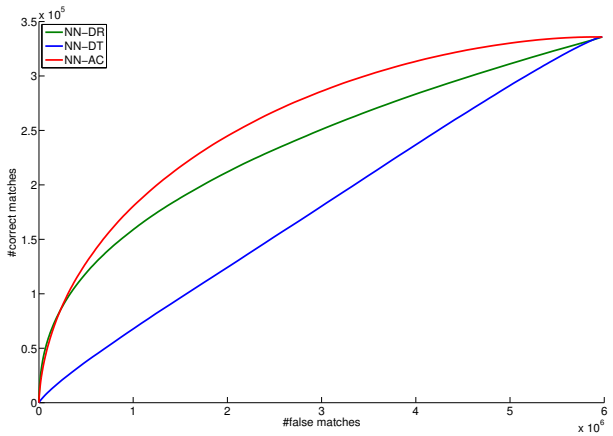


Fig. 6. Global ROC curves (on the whole database) for different matching criteria: NN-AC (red), NN-DT (blue) and NN-DR (green). Experimental protocol is $A \rightarrow \{^{A'}_{B}$ (an image A is matched separately against A' and an independent image B).

Next, we propose an extension of this last protocol where, for each query image A, the distractor image B is replaced by the entire database (deprived of A). Since this test involves much more computations than the previous one, it has been performed for only 100 images from the database (representing approximately $1.5\,10^{12}$ descriptor comparisons). Fig. 7 shows the corresponding global ROC curve. Again, one observes the substantial improvement provided by the NN-AC criterion. In fact, the improvement is greater than when only one image is used as a distractor, which suggests that the NN-AC criterion behaves well when the object of interest is seldom encountered.

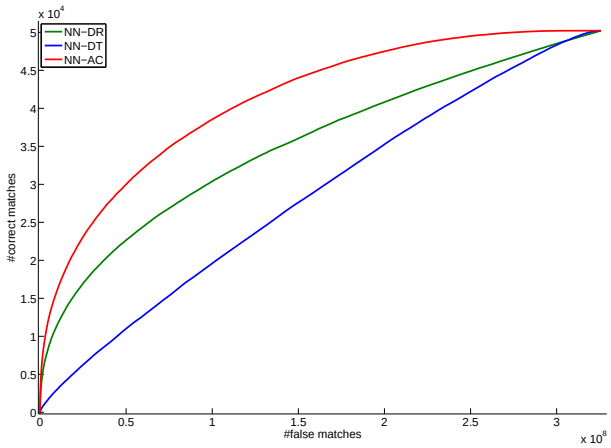


Fig. 7. Global ROC curves (on 100 images) for different matching criteria: NN-AC (red), NN-DT (blue) and NN-DR (green). Experimental protocol is the same as $A \rightarrow \{^{A'}_{B}$, except that B is replaced by the complete database. That is, an image A is matched separately against A' and against each other image.

4.4 Comparison of matching criteria - multiple matches

This section is a first attempt to compare matching criteria allowing multiple matching, thus *not restricted to the nearest neighbor*. First, there is no obvious way to define such an extension for the NN-DR criterion. We therefore compare the following two criteria: a simple threshold on distances, that we call DT and the criterion introduced in this paper (without restricting matches to nearest neighbors), that we called AC. Both criteria allow multiple correspondences for each query descriptor.

For this comparison, we propose a protocol similar to $A \rightarrow \{^{A'}_{B}$, except that A' is replaced by a single image, called $A'+A''$, which is the concatenation of two different transformations of A. In this experiment, each structure of A appears twice in $A'+A''$. Correct and false matches are counted exactly in the same way as in the protocol $A \rightarrow \{^{A'}_{B}$. This protocol is called $A \rightarrow \{^{A'+A''}_{B}$. Fig. 8 shows how the AC criterion clearly outperforms the DT criterion on the image database in this case of multiple matches.

4.5 Is the nearest neighbor restriction necessary ?

Following the previous section, it is quite natural to wonder whether not reducing the matches to the nearest neighbor yields a loss of performance *in the case where the target is present at most once*.

On Figure 9 we show, in continuous lines, global ROC curves for the two matching criteria AC and DT using the $A \rightarrow \{^{A'}_{B}$ protocol. Results for the matching criteria NN-AC and NN-DT, previously shown in Fig. 6,

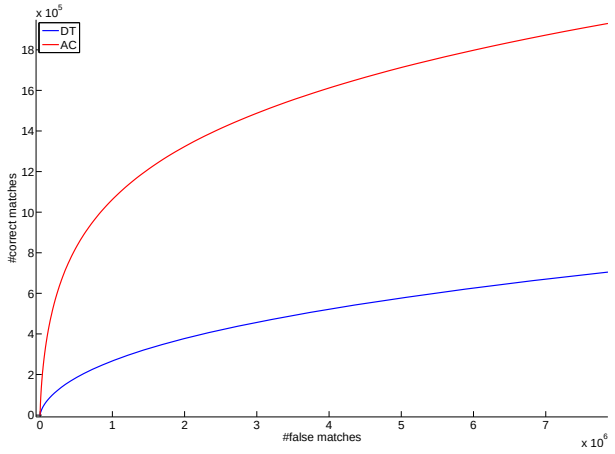


Fig. 8. Global ROC curves (on the whole database) for the $A \rightarrow \{^{A'+A''}_{B}$ protocol (the target is present twice, see Section (4.4)). Criterion AC is shown in red and criterion DT is shown in blue.

are represented in dashed lines. As could be expected, the performance of DT decreases significantly in comparison to NN-DT. Yet, we observe that AC and NN-AC criteria have similar results, even though AC does not have any restriction on the number of matches per query descriptor. This quite remarkable result indicates that the adaptive matching criterion introduced in this paper permits the rejection of false matches without any restriction on the number of possible matches.

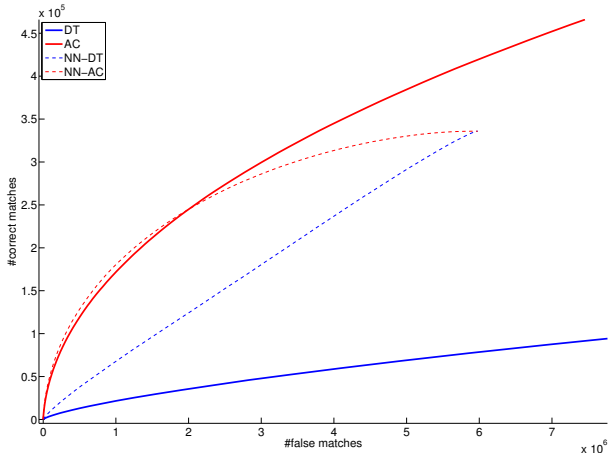


Fig. 9. Global ROC curves (on the whole database) for different matching criteria and for the $A \rightarrow \{^{A'}_{B}$ protocol. Dashed lines: matches are restricted to the nearest neighbor (NN-AC in red and NN-DT in blue). Continuous lines: the number of matches per query is not restricted (AC in red and DT in blue).

4.6 Some more experiments

In order to visually illustrate the behavior of the proposed matching procedure, this section presents some additional examples of matching between images.

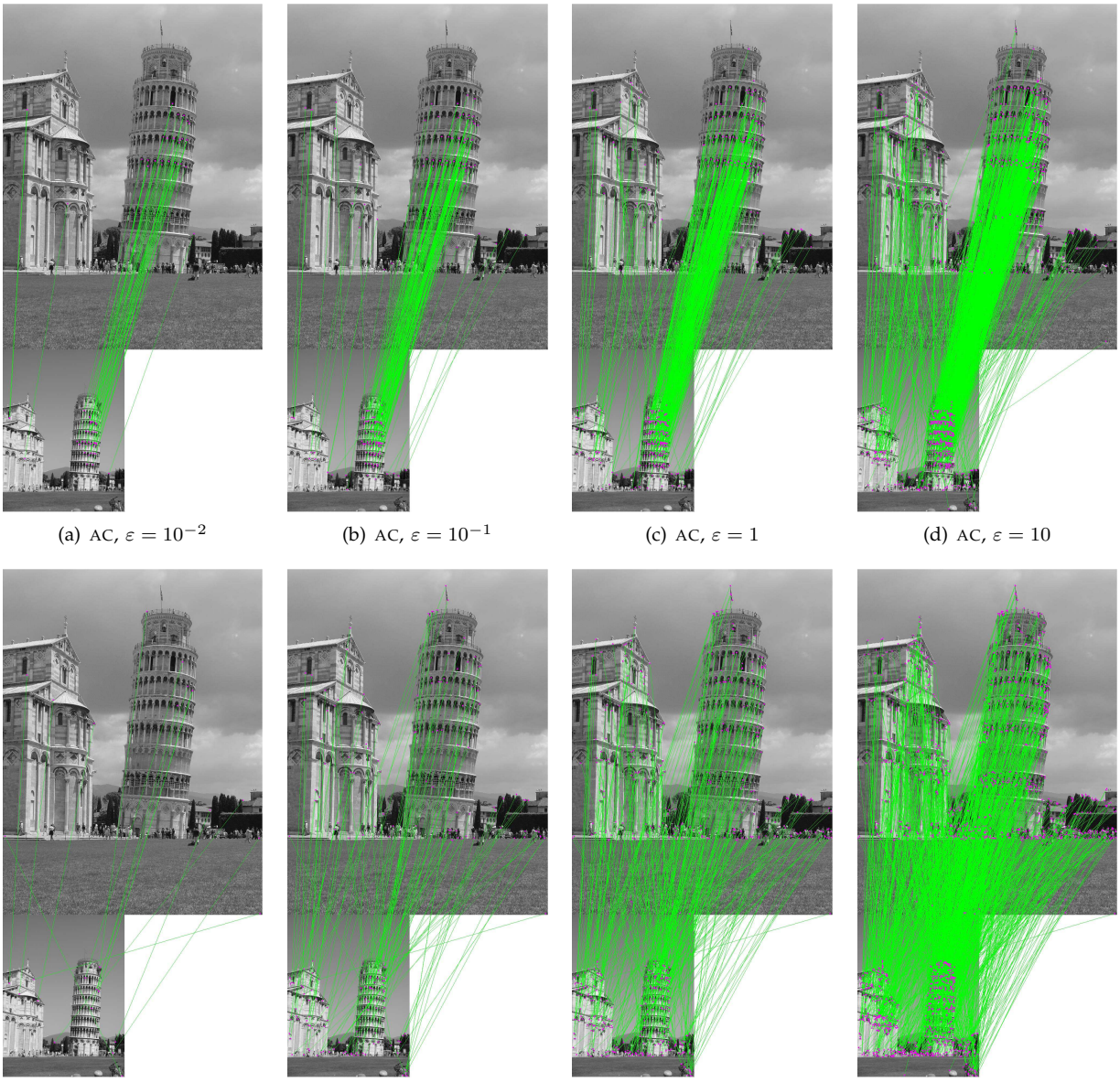
Firstly, we show the behavior of the proposed matching procedure using different thresholds in the case of a scene with repetitive structures. Such a situation is common in the case of, e.g., images of buildings. As pointed out in [20], these are difficult correspondence problems. Classical approaches could fail to provide enough relevant correspondences between images of the same scene. We compare two different views of the tower of Pisa shown in Fig. 10. Criterion NN-DR (used in Figs. 10(e), 10(f), 10(g) and 10(h) with CEMD and respectively $r = 0.7$, $r = 0.8$, $r = 0.85$ and $r = 0.9$) can only correctly match a relatively low number of points while controlling the number of false matches. Indeed, the presence of repetitive structures can foul the NN-DR criterion because of several candidate descriptors at a similar distance to the query. On the contrary, using the AC matching criterion -which is not restricted to the nearest neighbor-, results in multiple matches between columns and arches (Figs. 10(a), 10(b), 10(c) and 10(d) with CEMD and respectively $\varepsilon = 10^{-2}$, $\varepsilon = 10^{-1}$, $\varepsilon = 1$, and $\varepsilon = 10$).

Next, a single image (blue-framed) is matched separately with 8 different images (Fig. 11(a)). Four of them contain (one or several times) a common object with the query image (a can). The four other images do not contain the can. The complete matching procedure presented in this paper (CEMD for the distance and the AC criterion) is shown in Fig. 11(b)). It is compared to two classical matching procedures: Euclidean distance and NN-DR criterion in Fig. 11(c) or NN-DT criterion in Fig. 11(d). For each method, all images are matched with the same threshold ($\varepsilon = 10^{-2}$ for AC, $r = 0.8$ for NN-DR, and $t < 0.45$ for NN-DT). These thresholds are set in such a way as to obtain roughly the same number of correct matches between the query image and the image at the center of the leftmost column.

This matching experiment leads us to the same conclusions as the previous ROC curves. The AC criterion yields much fewer false matches on images where the object is not present and better detection of multiple occurrences. It is also interesting to notice that there are less false matches even in images where the object is present. This is not contradictory with the results of Section 4.3.1 (concluding to the equivalence of NN-DR and NN-AC when using the $A \rightarrow A'$ protocol), since many descriptors of either the query or the candidate image do not correspond to the object shared by the two images. This experiment shows (on a specific example) the versatility and adaptivity (all images are matched using the same threshold) of the proposed matching procedure.

5 Conclusion

In this paper, a new procedure for the matching of local, SIFT-like features has been proposed. First, a robust distance between circular histograms has been introduced and its advantages have been experimentally



(a) AC, $\varepsilon = 10^{-2}$ (b) AC, $\varepsilon = 10^{-1}$ (c) AC, $\varepsilon = 1$ (d) AC, $\varepsilon = 10$

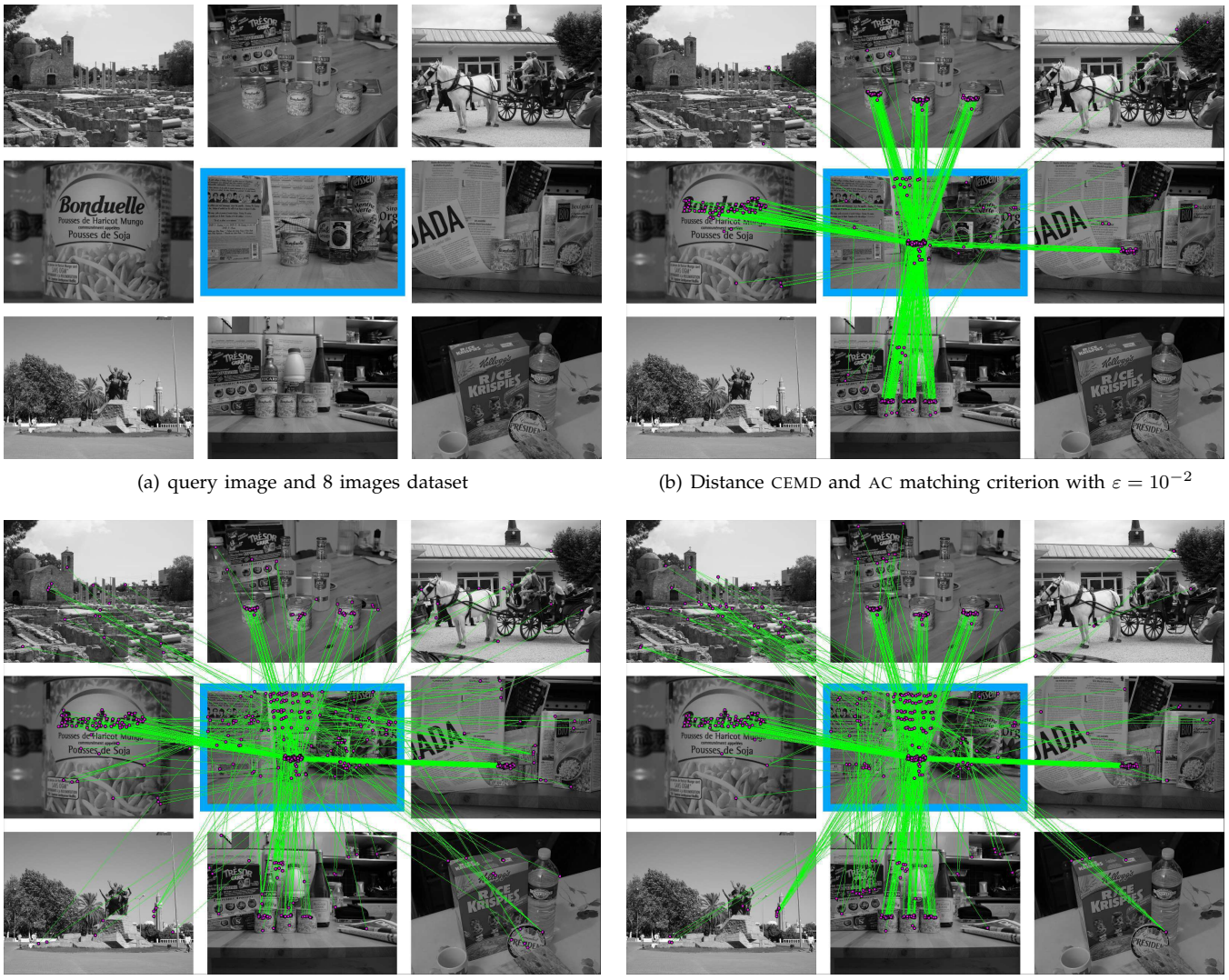
(e) NN-DR, $r = 0.7$ (f) NN-DR, $r = 0.8$ (g) NN-DR, $r = 0.85$ (h) NN-DR, $r = 0.9$

Fig. 10. Matching an object with repetitive structures: the tower of Pisa. Two different matching procedures are used with different thresholds: the first row corresponds to the AC criterion and the second row corresponds to the NN-DR criterion. The first criterion permits to match the repeated elements of the tower.

demonstrated on an image database. Second, a statistical matching criterion has been defined, relying on a threshold on a probability of false detections. The ability of this criterion to deal with situations where we do not know if the target is present has been demonstrated, as well as its ability to deal with multiple matches.

Several extensions of this work are foreseen. First, even though the computation of the proposed matching thresholds is not computationally demanding (it only requires to compute M convolutions for each query descriptor a^i), it cannot benefit in a straightforward way from fast nearest neighbor search schemes [11], [27]. It is of interest to investigate the possibility to approximate the probability of false detections using only a small subset of candidate descriptors.

The distance introduced in Section 2 can also be applied to other descriptors made of circular histograms, such as color (hue) histograms. Observe also that the matching methodology presented in Section 3 is completely generic and could be applied to other local descriptors, such as the affine invariant descriptors described in [41]. This matching methodology also enables us to simultaneously use different local features, by adapting the independence assumptions made in Section 3. Preliminary experiments on the joint use of color and direction histograms show promising results.



(a) query image and 8 images dataset

(b) Distance CEMD and AC matching criterion with $\varepsilon = 10^{-2}$

(c) Euclidean distance and NN-DR matching criterion with $r = 0.8$

(d) Euclidean Distance and NN-DT matching criterion with $t < .45$

Fig. 11. Comparison of different matching procedures (Distance + Matching Criterion). One query image (blue framed) containing a can is matched separately against 8 images (Fig. 11(a)). Only half of these images contain the can, present one or several times. For each matching procedure, the query image is compared with all 8 images using the same threshold. These thresholds are chosen such that the number of correct matches with the image at the center of the left column is the same for all procedures. Observe that for a given number of correct matches with this left-centered image, the matching procedure introduced in this paper (CEMD + AC criterion) yields more correct matches in other images while providing a better control of the number of false detections than classical procedures.

ACKNOWLEDGMENTS

The authors thank Henri Maître for his useful comments. J. Delon acknowledges 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] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision – 2nd Edition*. Cambridge University Press, 2004.

[2] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 712–727, 2008.

[3] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," *To appear in Int. J. Comput. Vision*, 2008.

[4] J. Sivic and A. Zisserman, "Video Google: Efficient visual search of videos," in *Toward Category-Level Object Recognition*, ser. LNCS. Springer, 2006, vol. 4170, pp. 127–144.

[5] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, "Segmenting, modeling, and matching video clips containing multiple moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 477–491, 2007.

[6] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vision*, vol. 74, no. 1, pp. 59–73, 2007.

[7] J. Jia and C.-K. Tang, "Image stitching using structure deformation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 617–631, 2008.

[8] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *BMVC*, 2002, pp. 384–393.

[9] R. Deriche, Z. Zhang, Q. Luong, and O. Faugeras, "Robust recovery of the epipolar geometry for an uncalibrated stereo rig," in *Proc. ECCV*, 1994, pp. 567–576.

[10] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vision*, vol. 73, no. 2, pp. 213–238, 2007.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[12] V. Ferrari, T. Tuytelaars, and L. Gool, "Simultaneous object recognition and segmentation from single or multiple model views," *Int. J. Comput. Vision*, vol. 67, no. 2, pp. 159–188, 2006.

[13] A. Kushal and J. Ponce, "Modeling 3D objects from stereo view and recognizing them in photographs," in *Proc. ECCV*, 2006.

[14] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, 2005.

[15] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," *Int. J. Comput. Vision*, vol. 73, no. 3, pp. 263–284, 2007.

[16] O. Chum and J. Matas, "Matching with PROSAC - progressive sample consensus," in *Proc. CVPR*, 2005, pp. 220–226.

[17] A. Baumberg, "Reliable feature matching across widely separated views," in *Proc. CVPR*, 2000.

[18] L. Moisan and B. Stival, "A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix," *International Journal of Computer Vision*, vol. 57, no. 3, pp. 201–218, 2004.

[19] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, 2002.

[20] W. Zhang and J. Kosecka, "Generalized ransac framework for relaxed correspondence problems," in *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 854–860.

[21] Y. Rubner, C. Tomasi, and L. J. Guibas, "The Earth Mover's distance as a metric for image retrieval," *Int. J. Comput. Vision*, vol. 40, no. 2, pp. 99–121, 2000.

[22] J. Rabin, J. Delon, and Y. Gousseau, "Circular Earth Mover's Distance for the comparison of local features," in *Proc. ICPR*. IEEE Computer Society, 2008.

[23] ——, "A contrario matching of SIFT-like descriptors," in *Proc. ICPR*. IEEE Computer Society, 2008.

[24] C. W. Niblack, R. Barber, W. Equitz, M. D. Flickner, E. H. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, "Qbic project: querying images by content, using color, texture, and shape," C. W. Niblack, Ed., vol. 1908, no. 1. SPIE, 1993, pp. 173–187.

[25] H. Ling and K. Okada, "Diffusion distance for histogram comparison," in *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 246–253.

[26] ——, "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, may 2007.

[27] J. Beis and D. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Proc. CVPR*, 1997, pp. 1000–1006.

[28] F. Destrempes, M. Mignotte, and J.-F. Angers, "Localization of shapes using statistical models and stochastic optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1603–1615, 2007.

[29] M. Brown, R. Szeliski, and S. Windner, "Multi-image matching using multi-scale oriented patches," in *Proc. CVPR*, 2005, pp. 510–517.

[30] G. Carneiro and A. D. Jepson, "Flexible spatial configuration of local image features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2089–2104, 2007.

[31] M. Lindenbaum, "An integrated model for evaluating the amount of data required for reliable recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 11, pp. 1251–1264, 1997.

[32] ——, "Bounds on shape recognition performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 7, pp. 666–680, 1995.

[33] A. Desolneux, L. Moisan, and J.-M. Morel, "Meaningful alignments," *Int. J. Comput. Vision*, vol. 40, no. 1, pp. 7–23, 2000.

[34] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel, "An a contrario decision method for shape element recognition," *Int. J. Comput. Vision*, vol. 69, no. 3, pp. 295–315, 2006.

[35] E. Levina and P. Bickel, "The Earth Mover's Distance is the Mallows distance: some insights from statistics," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2, 2001, pp. 251–256 vol.2.

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

[37] A. Desolneux, L. Moisan, and J.-M. Morel, *From Gestalt Theory to Image Analysis: A Probabilistic Approach*. Springer Verlag, 2008.

[38] ——, "A grouping principle and four applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 4, pp. 508–513, 2003.

[39] C. Olson and D. Huttenlocher, "Automatic target recognition by matching oriented edge pixels," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 103–113, 1997.

[40] R. G. Miller, *Simultaneous Statistical Inference*. Springer-Verlag, New York, 1991.

[41] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vision*, vol. 60, no. 1, pp. 63–86, 2004.