A contrario matching of SIFT-like descriptors

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

Abstract

In this paper, the matching of SIFT-like features [5] between images is studied. The goal is to decide which matches between descriptors of two datasets should be selected. This matching procedure is often a preliminary step towards some computer vision applications, such as object detection and image registration for instance. The distances between the query descriptors and the database candidates being computed, the classical approach is to select for each query its nearest neighbor, depending on a global threshold on dissimilarity measure.

In this contribution, an a contrario *framework for the matching procedure is introduced, based on a threshold on a probability of false detections. This approach yields dissimilarity thresholds automatically adapted to each query descriptor and to the diversity and size of the database. We show on various experiments on a large image database, the ability of such a method to decide whether a query and its candidates should be matched.*

1. Introduction

Matching local features is a very convenient way to compare several pictures. Many applications -such as object detection, stereo correspondence, image stitching, 3D reconstruction- are based on such procedures. An exhaustive list of the applications of the matching of local descriptors is beyond the scope of this paper. Illustrating examples can be found in [5, 2].

Whereas the extraction and representation of descriptors has been thoroughly studied (see *e.g.* the references in [7], where the SIFT [5] has proven to be the most robust and invariant representation method), there are few studies about their matching. The matching process consists in comparing *query descriptors* $\{a^i\}_{i=1\dots N_Q}$ (*e.g.* extracted from a query image) with *candidate descriptors* $\{b^j\}_{j=1\dots N_C}$ from a database (*e.g.* another image or a set of images), using a dissimilarity measure (a distance) and a selection criterion. Deciding whether a query descriptor a^i matches one or several candidates from the database $\{b^1, \dots, b^{N_C}\}$ boils down to setting a threshold on distances $D(a^i, b^j)$. Ideally, this threshold should be set automatically and should depend on a^i and on the entire database.
In practice, two different criteria are used to validate matches (as detailed in [7]), both relying on user-selected thresholds. The simplest one, that we call **DT**, uses a global threshold on distances. Generally, matches are restricted to the nearest neighbor for each query descriptor, in order to avoid multiple false detections that often occur. We will refer to this criterion as **NN-DT**.

In [5], Lowe introduces another criterion to decide whether the nearest neighbor matches the query. Assuming that the query a^i occurs at best once in the database, this test only considers the comparison of distances d_1 and $d_2 > d_1$ to the first and the second nearest neighbors. A query and its closest neighbor are matched when the ratio $r = d_1/d_2$ between theses two distances is below a threshold. This popular criterion, that we call **NN-DR**, is far more robust than a simple global threshold on distances and, indeed, behaves very well when the structure to be matched is present exactly once in the candidate database. Nevertheless, it is less clear that the same global threshold r should work when the query is not present in the database, whereas computer vision systems have to deal with situations when the target is present or not. Both scenarios will be considered in the experimental section. Moreover, if the structure appears more than once (because some interest points are repeated, because of repetitive structures or because the objects of interest are present more than once in the database) this criterion may fail.

In this paper, we propose an alternative matching criterion relying on adaptive thresholds. Roughly speaking, the method rests on the rejection of matches that

occur by chance. This matching procedure allows multiple detections over a database, while controlling the total number of matches. In Section 3, we compare its performances with the classical aforementioned matching criteria.

2. *A contrario* matching criterion (AC)

The *a contrario* framework was initially proposed by Desolneux *et al.* [3] in order to group low-level visual features. The basic principle is to detect groups of features that are very unlikely under the hypothesis that features are independent. In what follows, we call such a hypothesis a *null hypothesis*. The unlikeliness is ensured by controlling the average number of false detections. This generic approach has been applied with success to, among other things, the detection of alignments [3], contrasted edges, vanishing points, and grouping [4]. Recently, this methodology has been adapted to shape matching [8]. In the next two paragraphs, we adapt this methodology to the matching of SIFT-like features.

Recall that each SIFT-like descriptor a^i is made of M orientation histograms, $a^i = (a^i_1, \dots, a^i_M)$. In this section, we assume that the distance between two descriptors a^i and b can be written as $D(a^i, b) = \sum_{m=1}^{M} d(a^i_m, b_m)$. This is the case for classical bin-to-bin distances (Euclidean, Manhattan or χ^2 distance). The *a contrario* approach to descriptor matching then reads as follows. A candidate descriptor a^i being given, it is matched with b if $D(a^i, b)$ is small enough (as will be detailed in the next section) under the hypothesis that distances $d(a^i_m, b_m)$ are independent random variables. More precisely, we assume that b is a random descriptor, such that *the distances $d(a^i_m, b_m)$ ($m \in \{1, \dots M\}$) are mutually independent random variables.* We call this hypothesis "null hypothesis", and write it $\mathcal{H}^i_0$. Under this hypothesis, the probability density function of the random variable $D(a^i, b)$ could be written $p^i_1 * \dots * p^i_M$, where $*$ denotes the convolution product and p^i_m the pdf of the random variable $d(a^i_m, b_m)$. This enables us to compute the probability

$$\mathbb{P}\left(D(a^i, b) \leq \delta \,|\, \mathcal{H}^i_0\right) . \tag{1}$$

In order to numerically estimate these probabilities, we use empirical marginals. That is, for each $i \in \{1, \dots N_Q\}$ and each $m \in \{1, \dots M\}$, the laws p^i_m are empirically estimated over the database $\{b^1, \dots, b^{N_C}\}$. In other words, for each circular histogram a^i_m, the distribution function of the distance $d(a^i_m, b_m)$ is obtained as b_m spans the m^{th} histogram of the descriptors in the database.

A match between a^i and an element b^j in the database is considered as meaningful and validated as soon as the distance $\delta = D(a^i, b^j)$ between them is smaller than what can be expected under the hypothesis $\mathcal{H}^i_0$, *i.e.* as soon as the probability $\mathbb{P}(D(a^i, b) \leq \delta \,|\, \mathcal{H}^i_0)$ is small enough. Therefore, the probability is thresholded instead of $D(a^i, b)$ itself. In order to control the rate of false detections, it is necessary to fix δ in a way that depends on the number of queries and the number of candidate descriptors (the bigger these numbers, the more chances to observe false detections). Following previous works on *a contrario* detection or matching, we use a value of δ that is proportional to $N_Q N_C$. We define $\delta_i(\varepsilon)$ as the largest threshold δ that satisfies

$$\mathbb{P}\left(D(a^i, b) \leq \delta \,|\, \mathcal{H}^i_0\right) \leq \frac{\varepsilon}{N_Q N_C}. \tag{2}$$

A match between a^i and b is said to be ε-meaningful if $D(a^i, b) \leq \delta_i(\varepsilon)$. With this definition, it is easy to prove that *the expected number of ε-meaningful matches, when testing N_Q queries against N_C candidates satisfying the null hypotheses, is smaller than ε.*

This should be interpreted as a control over the number of false detections when comparing N_Q query descriptors against N_C candidates. In practice, ε is fixed and for each descriptor a^i we perform the following steps

1. $\delta \mapsto \mathbb{P}\left(D(a^i, b) \leq \delta \,|\, \mathcal{H}^i_0\right)$ is computed using Formula (1) ;
2. the threshold $\delta_i(\varepsilon)$ is automatically computed in function of the value ε using Formula (2) ;
3. for each descriptor b^j ($j = 1, \dots, N_C$), a^i is matched with b^j if $D(a^i, b^j) \leq \delta_i(\varepsilon)$.

We will refer from now to this matching criterion as **AC**.

Anticipating on the experimental section, let us underline the conceptual advantages of fixing ε to control the matches over other thresholds on distances. First, ε has the relatively intuitive meaning of a number of false alarms. Second, as said earlier, a single number yields thresholds that adapt to the query and the database. Note that the AC criterion permits multiple detections since the number of possible matches is not *a priori* restricted. However, it is also possible to limit the AC criterion to the matching of each query descriptor to its nearest neighbor. We call this criterion **NN-AC**. We will experimentally show in the next section the advantages of both criteria.

3. Experiments

This section presents several experiments to compare the performances of different matching criteria, making

use of a large set of $3.1\ 10^6$ descriptors, extracted from 732 images[1] modified by synthetic degradations (affine transformation and noise). Descriptors, obtained in a way similar to [5], are composed of $M = 9$ circular direction histograms, corresponding to nine disjoint regions around each oriented point. Each histograms is composed of 12 bins. We use CEMD [9], an adaptation of the Earth Mover's Distance for circular histogram, as a dissimilarity measure between features.

ROC curves are used to show the behavior of the different matching criteria introduced in Section 2.

3.1 Experimental protocols

The experiments follow two protocols. The first one, called $A \rightarrow A'$, consists in matching interest points between an image A and an image A' obtained by applying an affine transform and adding Gaussian noise to A (with $\sigma = 5$ for 8-bit coded images). A match is declared false (*i.e.* false positive) or correct (*i.e.* true positive) depending on some spatial tolerance (we follow exactly the protocol of [7]). This classical protocol checks out very simply the behavior of a matching procedure when two images containing exactly the same "objects" (before and after some transformations) are compared.

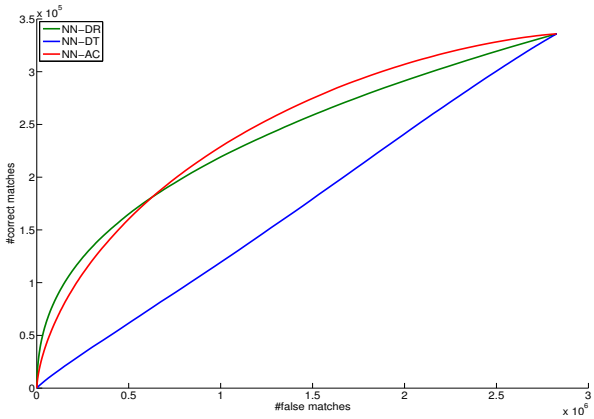
Now, because a great number of computer vision systems are confronted to cases when the target is not always present (the search of an object in an image database for instance), we propose an additional protocol called $A\begin{smallmatrix}\nearrow A' \\ \searrow B\end{smallmatrix}$. In this protocol, both comparisons of the image A with the modified image A' and with an image B, independent of A, are performed using the same thresholds. Correct and false matches between A and A' are defined in the same way as in the protocol $A \rightarrow A'$. All matches between the N_A descriptors from A and the N_B descriptors from B are considered as false matches.

Recall that for the NN-AC and AC criteria, a threshold on distances is obtained by thresholding a probability of false detections (see Equation (2)). For the $A \rightarrow A'$ protocol, Equation (2) is applied with $N_Q = N_C = N_A$, and for the $A\begin{smallmatrix}\nearrow A' \\ \searrow B\end{smallmatrix}$ protocol with $N_Q = N_A$ and $N_C = N_A + N_B$.

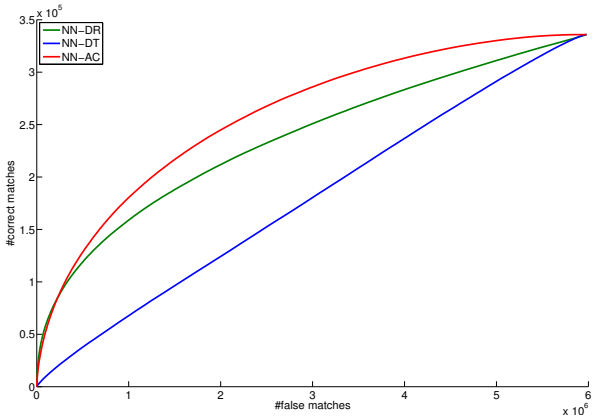
3.2 Performance evaluation

The evaluation, based on a set of 732 images, is performed with approximately 25 billion comparisons of descriptors. Each image A is compared to a modified image A' (protocol $A \rightarrow A'$) and to the next image B in the image set (protocol $A\begin{smallmatrix}\nearrow A' \\ \searrow B\end{smallmatrix}$), yielding 732 ROC curves for both protocols.

In order to compare the performances of the different matching procedures on the whole database, we draw ***global ROC curves***: for each threshold value, we plot the total number of correct matches on the whole database versus the total number of false matches.



(a) $A \rightarrow A'$ protocol.



(b) $A\begin{smallmatrix}\nearrow A' \\ \searrow B\end{smallmatrix}$ protocol

Figure 1. Global ROC curves (on 732 images and $3.1\ 10^6$ descriptors) for nearest neighbor criteria: NN-AC (red), NN-DT (blue) and NN-DR (green).
(a) $A \rightarrow A'$ protocol (an image A is matched against its transformed version A').
(b) $A\begin{smallmatrix}\nearrow A' \\ \searrow B\end{smallmatrix}$ protocol (an image A is matched separately against A' and an independent image B).

Figure 1(a) shows the global ROC curve for the nearest neighbor criteria NN-AC, NN-DT and NN-DR, with the $A \rightarrow A'$ protocol. Let us remark that such a curve permits to evaluate how stable an optimal threshold is from one experiment to the other. In particular, it shows that NN-DT is not stable in this regard, and is clearly outperformed by other criteria. We observe that both NN-AC and NN-DR, which have very similar global ROC

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

curves, are more stable. In this case, the proposed NN-AC criterion does not offer significant advantages in comparison with NN-DR. Indeed, as explained in Section 1, the NN-DR criterion is well adapted to the case where the target is present.

Figure 1(b) shows the global ROC curve (as before, computed on the whole image database) of the same three criteria, this time on the $A \nwarrow^{A'}_{B} \swarrow$ protocol. As explained earlier, this protocol mixes the matching results obtained by comparing separately A with A' and A with B. We can see that the performances of NN-DR clearly decrease in comparison to the proposed NN-AC criterion. On average, for a given number of correct correspondences between A and A', NN-AC yields fewer false correspondences than NN-DR. This means that NN-AC is better at deciding if the target is present or not, which is crucial in the context of object recognition.

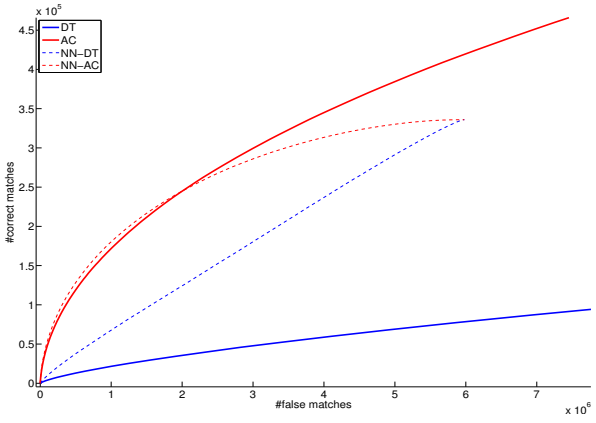


Figure 2. Global ROC curves (on 732 images) with $A \nwarrow^{A'}_{B} \swarrow$ protocol for both AC (red) and DT (blue) criteria in continuous lines. In comparison, the same criteria with nearest neighbor restriction from Fig. (1) are represented in dashed lines.

Nearest neighbor restriction. In section 2, the AC matching criterion is defined without nearest neighbor restriction. Indeed, the thresholds $\delta_i(\varepsilon)$ on dissimilarity measure, computed from Formula (2), could also define for each query descriptor the number of nearest neighbors which have to be matched. On Figure 2, we show the global ROC curves obtained for the two matching criteria AC and DT with the $A \nwarrow^{A'}_{B} \swarrow$ protocol. The previous results obtained with nearest neighbor restriction with the same protocol are displayed in dashed line. 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. This quite surprising 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.

4. Conclusion

A new criterion for the matching of SIFT-like features has been proposed. Our approach, based on the *a contrario* framework [4], first estimates the probability of false detections to define adaptive thresholds on a dissimilarity measure. Experimental results, obtained on a large image database, show that this *a contrario* matching criterion behaves well in situations where the target is present or not, without requiring any nearest neighbor restriction.

An extension of the $A \nwarrow^{A'}_{B} \swarrow$ protocol, where A' contains several occurrences of the target A and where B is a large image set, is currently being studied to show the advantage of the AC criterion for multiple object detection.

Moreover, the matching criterion (introduced in this paper) is generic and could be applied to other local descriptors, such as affine invariant descriptors [6] or shape context [1]. Next, we plan to group matches under geometrical constraints, by using RANSAC-type algorithms. Here again, the same *a contrario* methodology can be used to decide if an object is present in a database.

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] A. Desolneux, L. Moisan, and J.-M. Morel. Meaningful alignments. *IJCV*, 40(1):7–23, 2000.

[4] A. Desolneux, L. Moisan, and J.-M. Morel. A grouping principle and four applications. *IEEE Trans. PAMI*, 25(4):508–513, 2003.

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

[6] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

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

[8] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel. An a contrario decision method for shape element recognition. *IJCV*, 69(3):295–315, 2006.

[9] J. Rabin, J. Delon, and Y. Gousseau. Circular Earth Mover's Distance for the comparison of local features. In *Proc. ICPR*, 2008.