MAC-RANSAC: a robust algorithm for the recognition of multiple objects

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

Lionel Moisan
MAP5 CNRS
Université Paris Descartes, MAP5 CNRS
lionel.moisan@parisdescartes.fr

Abstract

This paper addresses the problem of recognizing multiple rigid objects that are common to two images. We propose a generic algorithm that allows to simultaneously decide if one or several objects are common to the two images and to estimate the corresponding geometric transformations. The considered transformations include similarities, homographies and epipolar geometry. We first propose a generalization of an a contrario *formulation of the RANSAC algorithm proposed in [6]. We then introduce an algorithm for the detection of multiple transformations between images and show its efficiency on various experiments.*

1. Introduction

This paper tackles the problem of object recognition : the aim is both to *detect* that a given object is common to several images and to *estimate* its pose, that is, to estimate the geometric transformation between the corresponding views. A classical approach to this problem is to extract local features from images (such as SIFTs [4]) and match them using some decision criterion (see *e.g.* [10]). Then, a last step of this procedure usually consists in detecting groups of local correspondences under coherent geometric transformations. In this work, we propose a fully automatic and robust method for this last step. The method does not require any parameter tuning and allows the detection of multiple groups. These multiple groups occur in a variety of computer vision tasks, such as the detection of several objects (or several occurrences of the same object) in a photograph, or the segmentation of motion in image sequences.

Some previous works An important body of work has been dedicated to the recognition of one or several objects through the grouping of correspondences. This non-exhaustive paragraph summarizes the most common approaches to this problem. Let us first recall that, because of a potentially large number of outliers, direct estimation methods such as least squares or M-estimators are usually not considered. Such approaches also hardly deal with multiple detections.

A first classical approach is the Hough transform [3]. Groups are detected thanks to a voting procedure where samples of correspondences increment bins in a quantized transformation space. While this approach is naturally suited to the recognition of multiple objects, its complexity strongly reduces its usefulness for recognizing complex geometric models such as projective transformations or epipolar geometry.

A widely used alternative is the RANSAC (*RANdom SAmple Consensus*) algorithm [2], to be recalled in the next section. Its principle is to detect a *unique* transformation by randomly sampling *n*-uplets. For each of these *n*-uplets, the corresponding transformation(s) is estimated and the adequacy of this transformation to the remaining data is then tested. The transformation that maximizes this adequacy (yields the best consensus) is then kept. This approach has two main advantages: its robustness and its speed. Nevertheless, its practical use requires the setting of several sensitive parameters. Several interesting possibilities to tune parameters have been proposed, including MSAC and MLESAC [15], MINPRAN [11] and the *a contrario* approach presented in [6]. This last approach, that from now we call AC-RANSAC (*A Contrario RANdom SAmple Consensus*), has the advantage of allowing the automatic tuning of parameters without any *a priori* on the distribution of inliers. It is the starting point of the approach presented in the present paper. In view of object recognition in a fairly general setting, we will first generalize this approach to the case of planar transformations and adapt it to deal with matching between SIFT-like descriptors [4].

Next, we tackle the problem of multiple object recognition. Indeed, one of the strong limitation of the RANSAC algorithm is that it only allows the recognition of a single object. Several approaches have been proposed to allow the detection of multiple groups with RANSAC. These can be roughly categorized as follows. A first strategy [19, 13, 14, 18] is to detect all groups *simultaneously* by fusing the different groups found by RANSAC. A second strategy [11, 16] is to sequentially detect groups by iteratively running RANSAC. While the practical implementation of this simple idea is non-trivial, we will advocate the use of the sequential approach and detail our algorithm for the detection of multiple groups, MAC-RANSAC, in Section 4. We will see that the quality measure of a group given by AC-RANSAC is a key point of this algorithm, acting both as a way to select groups and as a stopping criterion. Eventually, we illustrate the method in an experimental section.

2. Notations and RANSAC algorithm

Let $C = \{(m_i, m'_i) \mid i = 1, \dots, N\}$ be a set of point correspondences between two images I and I', obtained for instance thanks to [10]. In order to remain completely general, consider that some of the m_i (respectively m'_i) can be identical: a point in one image can thus be matched with different points in the other image. We want to find if some of these correspondences can be explained by a common geometrical transformation. In this paper, we focus on planar transformations and on epipolar geometry between images. Recall that, in general, a planar transformation $H_{S'}$ can be entirely defined from a set S' of n correspondences, where $n = 2$ for similarities, $n = 3$ for affine transformations, and $n = 4$ for projective transformations (or homographies). As for epipolar geometry between two images, recall that a fundamental matrix $F_{S'}$ can be estimated from a set S' of $n = 7$ correspondences (m_i, m'_i) satisfying the constraints $m'_i F_{S'} m_i = 0$ in homogeneous coordinates.

RANSAC algorithm In this context, the aim of the Random Sample Consensus (RANSAC) algorithm is to find a geometric transformation T which explains as well as possible a maximum of correspondences in C (these correspondences are called *inliers* for T). At each step, a subset S' of n correspondences is drawn from C and used to estimate a planar transformation $H_{S'}$ or a fundamental matrix $F_{S'}$. For each correspondence (m, m') in $C \setminus S'$, a transfer error is defined to measure the adequation between (m, m') and the transformation. A symmetric transfer error can be defined for a planar transformation $H_{S'}$ as a function of the quantities $d(m', H_{S'}m)^2$ and $d(H_{S'}^{-1}m', m)^2$, where $d(.,.)$ is a distance between points. If $F_{S'}$ is a fundamental matrix, the symmetric transfer error can be defined as a function of $d(m', F_{S'}m)^2$ and $d(F_{S'}^T m', m)^2$, where $d(.,.)$ is a distance between a point and a line on the plane and where $F_{S'}^T$ is the transposed matrix of $F_{S'}$. In both cases, the inliers in C are then defined as all the correspondences (m, m') for which this error is smaller than a threshold δ. The score of a group is given by its size. This process is repeated i_{max} times and the largest consensus found during these steps is kept if its size is larger than a threshold N_C.

Observe that three thresholds must be set by the user: the distance threshold δ, the threshold on the minimal number of inliers N_C and the number of iterations i_{max}. The setting of these thresholds is critical for the output of the algorithm.

3. Generalization of AC-RANSAC

In order to avoid the shortcomings of the RANSAC algorithm (in particular the high sensibility to the different parameter settings), the authors of [6] propose an *a contrario* approach to fix these parameters automatically. This approach is presented in [6] in the context of epipolar geometry. The RANSAC algorithm is combined with a hypothesis testing framework to decide –for each sample S'– if a set of correspondences S should be considered as a valid consensus for the fundamental matrix $F_{S'}$. We present here a slightly generalized version of this work for different families of transformations, either planar transformations or epipolar geometry. We will refer to it as AC-RANSAC.

3.1. *A contrario* framework

The null hypothesis Assume that we have a set C : $\{(m_i, m'_i) \mid i = 1, \dots, N\}$ of correspondences between two images I and I'. We want to decide if a subgroup of these correspondences can be explained by a unique transformation. To answer this question, the AC-RANSAC approach relies on a null hypothesis $\mathscr{H}_0$, describing a "generic" distribution of random correspondences $(\mathbf{m_i}, \mathbf{m'_i})$, $i = 1, \dots, N$ for which it is commonly accepted that no consistent consensus should be found. A real set of correspondences will then be defined as meaningful if it is unlikely to appear under the hypothesis $\mathscr{H}_0$. This approach permits to limit the *type I errors*, *i.e.* the cases where a consensus is detected whereas the distribution of points follow $\mathscr{H}_0$. The null hypothesis $\mathscr{H}_0$ is defined in the following way, N being a fixed integer:

Definition 1 (Null Hypothesis) *A set* $\mathbf{C}$ *of* N *random correspondences* $(\mathbf{m_i}, \mathbf{m'_i})$, $i = 1, \dots, N$ *is said to follow the null hypothesis* $\mathscr{H}_0$ *if*

- *the points* $\mathbf{m_i}$ *and* $\mathbf{m'_j}$, $i, j = 1, \dots, N$ *are mutually independent random variables ;*
- *the points* $\mathbf{m_i}$, $i = 1, \dots, N$ *are uniformly distributed on the image* I *and the points* $\mathbf{m'_j}$, $j = 1, \dots, N$ *are uniformly distributed on the image* I'.

In the following, we compute the probability for a group of correspondences to follow $\mathscr{H}_0$, and we infer the definition of the rigidity of a group.

Probabilities under the null hypothesis Let $\mathbf{C}$ be a set of N random correspondences, following the null hypothesis $\mathscr{H}_0$, and let $\mathbf{S'}$ be a given subset of $\mathbf{C}$, such that $\#\mathbf{S'} = n$. Let $\mathbf{S}$ be a subset of $\mathbf{C}$ such that $\mathbf{S} \cap \mathbf{S'} = \emptyset$. In the context of epipolar geometry ($n = 7$), Moisan and Stival [6] define the *rigidity* of the set $\mathbf{S}$ for the fundamental matrix $F_{S'}$ as

$$\alpha(\mathbf{S}, F_{\mathbf{S'}}) := \max_{(\mathbf{m},\mathbf{m'}) \in \mathbf{S}} \max\left(\frac{2D'}{A'} d(\mathbf{m'}, F_{\mathbf{S'}}\mathbf{m}), \frac{2D}{A} d(\mathbf{m'}, F_{\mathbf{S'}}^T\mathbf{m'})\right),$$

where D (resp. D') and A (resp. A') are the length of the diagonal and the area of I (resp. I'). This quantity measures the adequacy between the set $\mathbf{S}$ and the fundamental matrix $F_{S'}$. They show that under the null hypothesis $\mathscr{H}_0$, the probability to observe a rigidity $\alpha(\mathbf{S}, F_{\mathbf{S}'})$ smaller than a value α is bounded by $\alpha^{\#\mathbf{S}}$:

$$\mathbb{P}_{\mathscr{H}_0}[\alpha(\mathbf{S}, F_{\mathbf{S}'}) \leq \alpha] \leq \alpha^{\#\mathbf{S}}. \quad (1)$$

In the case of planar transformations, this inequality remains valid if one replaces $F_{\mathbf{S}'}$ by $H_{\mathbf{S}'}$ and if the rigidity is defined as

$$\alpha(\mathbf{S}, H_{\mathbf{S}'}) := \max_{(\mathbf{m},\mathbf{m}')\in\mathbf{S}} \max\left(\frac{\pi}{A'} d(H_{\mathbf{S}'}\mathbf{m}, \mathbf{m}')^2, \frac{\pi}{A} d(\mathbf{m}, H_{\mathbf{S}'}^{-1}\mathbf{m}')^2\right).$$

Validation criterion A set S is considered as a valid consensus for $T_{S'}$, the transformation associated to the group S', as soon as $\mathbb{P}_{\mathscr{H}_0}[\alpha(\mathbf{S}, T_{\mathbf{S}'}) \leq \alpha(S, T_{S'})]$ is small enough. To put it more simply, S is considered as a good consensus for $T_{S'}$ if it is highly unlikely to observe such a small rigidity under the hypothesis $\mathscr{H}_0$. The following definition precises what we mean by "small".

Definition 2 *Let $C = \{(m_i, m'_i) \mid i = 1, \ldots, N\}$ be a set of N correspondences between I and I'. Let S be a subset of C, constituted of $\#S = K$ correspondences. For a given $\varepsilon > 0$, the set S is said to be ε-meaningful if there exists $n \in \{2,3,4,7\}$ and a subset S' of C, such that $\#S' = n$, $S' \cap S = \emptyset$ and*

$$NFA(\mathbf{S}, \mathbf{S}') := \gamma(N-n)\binom{N}{K}\binom{N-K}{n}\left(\alpha(S, T_{S'})\right)^K \leq \varepsilon.$$

In this definition, the coefficient γ denotes the largest number of transformations $T_{S'}$ which can be estimated from n correspondences, that is $\gamma = 1$ for planar transformations and $\gamma = 3$ for epipolar geometry [1]. The quantity $NFA(S, S')$ defines a *consensus quality* between the set S and the transformation $T_{S'}$. Observe that this quantity combines in the same term the $T_{S'}$-rigidity of S and its size. By using a unique threshold ε, which will always be set to 1 in the experiments, this formula yields adaptive thresholds on the rigidities $\alpha(S, T_{S'})$.

3.2. AC-RANSAC algorithm

At each iteration i, a set S' of size n is drawn from C and the transformation $T_{S'}$ is estimated. All the correspondences $c_i = (m_i, m'_i)$ in $C \setminus S'$ are then sorted according to the value of the residuals $\alpha_i = \alpha(c_i, T_{\mathbf{S}'})$. At this point, finding the set S which minimizes $\mathrm{NFA}(S, S')$ is easy: for each value of K smaller than $N-n$, the set of size K which minimizes $\mathrm{NFA}(S, S')$ is composed of the K correspondences with the smallest α_i. The algorithm continues until i_{max} iterations are realized or until a set S satisfying $\mathrm{NFA}(S, S') < 1$ is found. In this last case, a refinement step, called ORSA (Optimal Random Sampling Algorithm) focuses on inliers to optimize the consensus (for more details the reader is referred to [6]).

[1] In the case of epipolar geometry, the 7-points algorithm can yield 3 fundamental matrices [6]. The matrix used for the definition is the one minimizing the rigidity $\alpha(S, F_{S'})$.

3.3. Null hypothesis and local features

In [6], the AC-RANSAC algorithm is used to estimate the epipolar geometry with control point correspondences selected from stereoscopic image pairs. In [8], the authors propose an extension of AC-RANSAC where the matching procedure of SIFT features is performed simultaneously with the estimation of fundamental matrix. In this paper, we suppose that the correspondence set C is obtained by matching local features between images I and I', *e.g.* in a similar way to [4, 10]. However, the *null hypothesis* on which AC-RANSAC is based, is not necessarily appropriate in such a case. Indeed, let us recall that the null hypothesis $\mathscr{H}_0$ for random correspondences $\mathbf{C}$ rely on the assumption that the matched points $\mathbf{m}_i \in I$ and $\mathbf{m}'_i \in I'$ are mutually independent random variables. As a consequence, there are two structural reasons for which some matches obtained from local descriptors cannot follow this assumption.

The first one is the presence of **multiple correspondences**: a point m of image I can have several correspondents in I', and reciprocally. This property has the great advantage of allowing the detection of an object appearing several times in an image, or the detection of objects presenting some repeated structures (such as texture). However, it contradicts obviously the null hypothesis $\mathscr{H}_0$. In order to deal with such multiple correspondences, we introduce the simple following maximality principle.

Definition 3 (Maximality principle) *Only one correspondence per interest point could be selected into a consensus set S. For a given interest point, the chosen correspondence is the one which minimizes the residual error α according to the transformation $T_{S'}$ evaluated.*

The second phenomenon incompatible with the null hypothesis $\mathscr{H}_0$ is the **redundancy of interest points**: detectors of interest points (or regions) used to build local descriptors (SIFT [4], MSER [5], …) tend to detect certain structures in a *redundant way*. The same structure, typically a corner, are then represented by several interest points, which only differs slightly in position and scale. These redundant points induce not-independent matches.

Before sequentially running the AC-RANSAC algorithm, we propose to detect and discard redundant correspondences using the scale information of interest points.

Redundant correspondences are those that share the same interest point in either I or I' and whose corresponding points in the other image are close enough.

Definition 4 (Redundant correspondences) *Two correspondences $c_i = (m_i, m'_i)$ and $c_j = (m_j, m'_j)$ between interest points are redundant if* ***one of the following statements is true***

- $m_i = m_j$ *and* $\|m'_i - m'_j\|_2 < \min\{\sigma_i, \sigma_j\}$

or

- $m'_i = m'_j$ *and* $\|m_i - m_j\|_2 < \min\{\sigma'_i, \sigma'_j\}$

where $\|.\|_2$ *stands for the Euclidean norm, and* σ_k *represents the characteristic scale of the interest point* m_k.

Next, a criterion is needed to choose among redundant matches. We propose to discard the less meaningful redundant matches c, making use of the quality measure $q(c)$ from the matching step. Note that other RANSAC-like methods also exploit the quality measure of correspondences, see *e.g.* [1, 8]. Using this quality measure, we define the following exclusion principle for redundant matches. A match c_i is discarded as soon as it is redundant with a match c_j such that $q(c_j) < q(c_i)$.

4. MAC-RANSAC for multiple group detection

The RANSAC algorithm (and consequently its a contrario extension AC-RANSAC) enables one to detect only one transformation from a set of correspondences. Now, as it has been previously stressed, it is often necessary to be able to detect several groups of correspondences. This multiple detection may be achieved through the sequential use of RANSAC, although various authors have pointed out several difficulties raised by this approach[11, 12, 16].

4.1. Sequential AC-RANSAC

The so-called "sequential RANSAC" algorithm consists in iteratively applying RANSAC on the set of correspondences, from which detected inlier groups are withdrawn after each iteration. The algorithm stops only if no new groups are detected. This simple approach suffers from strong limitations:

- **Detection of false transformations**. This refers to the validation of groups that are composed of outliers.
- **Fusion of nearby transformations** [12]. In the case of multiple transformations, it is quite likely that two or more of these are detected as the same consensus set.
- **Segmentation of a single transformation** into smaller ones. This phenomenon, opposite to the previous one, will occur for instance when the spatial tolerance is too small.
- **Ghost transformations** [7] This term refers to the detection of several spurious transformations echoing a single real transformation. This happens when the object of interest is composed of repeated structures, such as a building facade.

In practice, combining the *a contrario* framework with sequential RANSAC offers significative advantages. First of all, the validation criterion 2 also defines a very robust stopping criterion for the sequential detection. In addition, the detection thresholds are automaticaly set on the residual errors of inliers, without requirering any prior knowledge on the data, which avoids the problem of over-segmentation.

Nevertheless, the two other problems in the above taxonomy (Fusion of transformations and ghost transformations) remain. We address these two limitations in the next paragraphs, first by proposing a spatial filtering of correspondences in § 4.3, then by defining a splitting criterion in § 4.4.

4.2. MAC-RANSAC algorithm overview

The final and complete MAC-RANSAC algorithm – including the principles to be defined in the next two paragraphs– is summarized in Table 1. Observe that the only parameter is the maximal number of iterations i_{max}, since, as it is classical with *a contrario* methods, the NFA is simply thresholded at the value $\varepsilon =$ **one**, automatically yielding both a spatial tolerance threshold and a minimum number of inliers to validate a group. All experiments to be displayed later on will be obtained with this value for ε.

4.3. Spatial correspondence filtering

A remaining source of error when sequentially applying AC-RANSAC is what we have coined the *ghost transforms*. To the best of our knowledge, few studies [7, 17, 16] described the phenomenon of self-similarity in object recognition. Indeed, repetitive matches due to self-similarity create artificial detections "echoing" to the unique true transformation between the two images. The same object is then detected several times with different poses.

The correct transformation usually has the best score, so that it is detected first. In order to then discard subsequent detections of echoing transformations, the following definition of repetitive correspondences is used.

Definition 5 (Repetitive correspondences) *Let S be a given consensus set, and $\mathscr{C}$ the remaining correspondences (so that $\mathscr{C} \cap S = \emptyset$). A match $c_i = (m_i, m'_i) \in \mathscr{C}$ is repetitive with respect to S if* ***both the following conditions are satisfied****:*

- $\exists\, m \in S$ *s.t.* $\|m - m_i\|_2 < \min\{\sigma, \sigma_i\}$

and

- $\exists\, m' \in S$ *s.t.* $\|m' - m'_i\|_2 < \min\{\sigma', \sigma'_i\}$

Table 1. MAC-RANSAC algorithm for multiple group detection

MAC-RANSAC

Inputs: Set of non-redundant correspondences $\mathscr{C}$ (obtained using definition 4) and maximal number of iterations i_{max}. *Initialization*: $i := 0$ and $\mathscr{S} := \{\emptyset\}$.

1) **Detection:** *While* $i < i_{max}$, uniform sampling of $S' \subset C$ and search of $S \subset C \setminus S'$ minimizing $\text{NFA}(S,S')$.
   ◦ *If* $\text{NFA}(S,S') < 1$, $(S_{\text{opt}}, S'_{\text{opt}}) := (S,S')$ and go to step 2).
   ◦ *Else* $i := i+1$.
2) **ORSA optimization:** Repeat $i_{max}/10$ times : sampling of S' among S_{opt} and search of $S \subset C \setminus S'$ minimizing $\text{NFA}(S,S')$.
   ◦ *If* $\text{NFA}(S,S') < \text{NFA}(S_{\text{opt}}, S'_{\text{opt}})$, $(S_{\text{opt}}, S'_{\text{opt}}) := (S,S')$.
3) **Fusion detection:** Search of an optimal subset pairs included in S_{opt} with criterion 1.
   ◦ *If* detection of fusion, two 1-meaningful subsets (S_1,S'_1) and (S_2,S'_2) are identified.
   ◦ *Else*, $S_1 := S_{\text{opt}}$ and $S_2 := \emptyset$.
4) **Spatial filtering:** Discarding of correspondences that are repetitions of S_1 (identified with definition 5). Definition of $C := C \setminus S_1$.
5) **Iteration:** Addition of S_1 to the list $\mathscr{S}$, $i := 0$.
   ◦ *If* $S_2 = \emptyset$, go to step 1)
   ◦ *Else*, $(S_{\text{opt}}, S'_{\text{opt}}) := (S_2,S'_2)$, then go to step 2).

Output: List of disjoint groups $\mathscr{S}$.

where $\|.\|_2$ *stands for the Euclidean norm, and* σ_k *represents the characteristic scale of the interest point* m_k.

Now, each time a new group of correspondences is validated, the repetitive correspondences according to this group are discarded from the remaining correspondences set C (*i.e.* step 4 of the MAC-RANSAC algorithm in Table 1).

4.4. Detection of the fusion of several groups

We will see in the experimental section that, when more than one transformation is concerned, sequential AC-RANSAC could validate groups of correspondences which result from the *fusion of several transformations*. To avoid this phenomenon we propose an original algorithm for the detection of such fusions.

Detection of fusions involving two groups For the sake of clarity, we first study the case of a fusion involving two transformations. Let S_0 be a group of correspondences, S_1 and S_2 two disjoint sub-groups of S_0, such that $S_1 \cap S_2 = \emptyset$ and $S_0 \supset S_1 \cup S_2$. The group S_0 is said to result from the fusion of these distinct groups when it is validated whereas the transformations of S_1 and S_2, respectively noted T_1 and T_2, are *sufficiently* different.

The reason why the fusion group S_0 can be validated is that its meaningfulness (NFA) is lower than those of each of the two sub-groups S_1 and S_2, whereas these groups have a lower rigidity:

$$\begin{cases} \text{NFA}(S_0,S'_0) & < \quad \min\left\{\text{NFA}(S_1,S'_1)\,,\text{NFA}(S_2,S'_2)\right\} \\ \alpha\left(S_0,\ T_{S'_0}\right) & \geq \quad \max\left\{\alpha\left(S_1,T_{S'_1}\right),\alpha\left(S_2,T_{S'_2}\right)\right\} \end{cases} .$$

This result is due to the "greedy" behavior of AC-RANSAC which looks for the group of correspondences minimizing the NFA. In order to be able to detect the fusion of two groups, we must define a criterion which compares the group $S_0 = S_1 \cup S_2$ with both groups S_1 and S_2 simultaneously.

We propose the following splitting criterion:

Criterion 1 (Splitting criterion) *Let S_0 be an ε-meaningful group of correspondences from the set C, with transformation $T_{S'_0}$. If there exist two disjoint subsets S_1 and S_2 from S_0, and also two disjoint subsets S'_1 and S'_2 in $S_0 \setminus \{S_1 \cup S_2\}$, such that both following statements are satisfied:*

◦ *$NFA(S_1,S'_1) \leq \varepsilon$ and $NFA(S_2,S'_2) \leq \varepsilon$*

◦ *$NFA(S_1,S'_1) \times NFA(S_2,S'_2) < NFA(S_0,S'_0)$,*

then, the two groups S_1 and S_2 are validated as two distinct groups of correspondences instead of a unique group S_0.

In order to be able to exploit this splitting criterion, we now define a strategy to find optimal subsets S_1 and S_2, as well as a recursive splitting scheme to deal with the fusion of more than two groups.

Recursive dyadic splitting into sub-groups. We choose to identify *only one of the sub-groups* involved in the previously described fusion. For that purpose, we first define a strategy to test for the splitting of a group into two sub-groups. This strategy is then applied recursively to the initial group S_0.

First observe that if S_1 and S_2 are two subsets of S_0, and if S_1 is the smallest of the two, then $\#S_1 \leq \#S_0/2 \leq \#S_2$. This straightforward observation suggests (as also noticed by Stewart in [11]) the following. The sub-group S_1 is obtained by searching (through random sampling) a subset $S'_1 \subset S_0$ which *minimizes the quantity $NFA(S_1,S'_1)$* under the constraint that $\#S_1 \leq \frac{\#S_0}{2}$. The sub-group S_2 is then obtained by searching a set S'_2 among the remaining correspondences of S_0 which minimizes $\text{NFA}(S_2,S'_2)$. The splitting criterion (1) then enables one to know if this splitting should be kept or not.

As long as the criterion validates the splitting, this strategy of dyadic splitting is **recursively applied** to the group $S_1^{(k)}$ which has been identified at the previous iteration k.

This principle is illustrated in Figure 1 in a case involving the fusion of 5 groups.

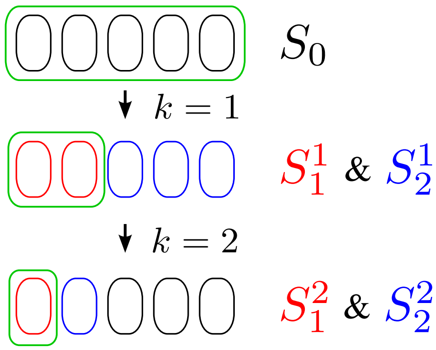


Figure 1. *Illustration of the recursive splitting approach used to identify one optimal subset.*

This process stops when the last subset found $S_1^{(k)}$ corresponds to a single transformation, *i.e.* when the splitting criterion does not validate its split into $S_1^{(k+1)}$ and $S_2^{(k+1)}$, or when such ε-meaningful sub-groups do not have been found.

<u>*Remark:*</u> When S_0 is split into S_1 and S_2, the only group validated is S_1. The second group S_2 being ε-meaningful, it is considered as a detection and the MAC-RANSAC algorithm go directly to the optimization stage (step 2 in table 1).

5. Experimental results

In this experimental section, we illustrate the efficiency of the proposed MAC-RANSAC algorithm in various situations. Preliminary correspondences between points are obtained following the matching procedure described in [10], with a detection threshold fixed to $\varepsilon = 1$. Observe that this matching procedure allows multiple matchings between points, a desirable property in view of multiple detections.

In the following experiments, the MAC-RANSAC algorithm is used with various geometrical models (planar transformations and epipolar geometry), with a detection threshold always fixed to $\varepsilon = 1$ (a group is validated as soon as its NFA is smaller than 1). We first consider the picture pair shown in Figure 2. This experiment illustrate the interest of the recursive splitting procedure presented in the previous section. A poster has been photographed before and after its folding into three parts. On this pair, MAC-RANSAC is performed using the homographic model. We thus expect to detect the tree planes corresponding to the folded parts. Without using the splitting criterion (see Figure 2(c)), two groups are detected, of which the largest (in red) results from the fusion of three distinct transformations. The use of the splitting criterion allows us to detect the three correct groups (see Figure 2(b)).

In order to illustrate the robustness of our approach for multiple object recognition runs, we now examine in the

(a) Pair of photographs analyzed

(b) Grouping of correspondences **with** the splitting criterion

(c) Grouping of correspondences **without** the splitting criterion

Figure 2. *Illustration of the interest of the recursive splitting procedure.*

four following experiments various scenarios with different geometrical models.

Figure 3(a) displays two frames from the sequence *Leuven castle* [9], where the camera moves and the scene is fixed. This scene represents an 'L'-shaped building with self-similarity. The result of the MAC-RANSAC algorithm with epipolar constraints is shown in Figure 3(b). Correctly, only one group is detected in this case. Results using homographies is displayed in Figure 3(c)). Three different groups corresponding to the different planes of the building are detected, as well as an additional group related to the trees in background (group of blue dots). Observe that there are no detections of ghost transformations, thanks to the spatial filtering of repetitive correspondences (criterion 5).

In Figure 4(a) are displayed two different points of view on the same scene in which only one object (the phone) that has been displaced. In the case of epipolar geometry (Figure 4(b)), two groups are correctly identified: a large group (in red) corresponding to the static part of the scene, and a second group (in blue) corresponding to the phone which has been moved. When considering homographic transformations (Figure 4(b)), MAC-RANSAC provides the same group for the telephone (in yellow), the rest of the scene being segmented in several planes.

Figure 5(a) shows an extreme case for the recognition of multiple objects. A first picture represents a soda can, and a second one a scene with 28 cans having the same logo. The result of the grouping with MAC-RANSAC using planar transformations is given in Figure 5(b) (similarity, affine transform and homography give very similar results) : the 28 occurrences of the can are correctly detected. Let us insist on the fact that each detected group corresponds to a small proportion of correspondences according to the initial

(a) Two views of a 3D scene from the image sequence of [9]

(b) A unique group is detected when using epipolar geometry

(c) 4 groups are detected with homography

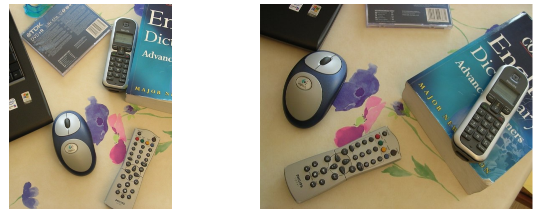
Figure 3. *Application of the MAC-RANSAC algorithm for a static scene with a camera motion.*

set of correspondences. In particular, the two groups corresponding to partially occluded cans (in dark blue) represent each only 1% of the total of correspondences. After these 28 cans have been identified, the algorithm automatically stops because no meaningful group (i.e. with $NFA \leq 1$) is found.

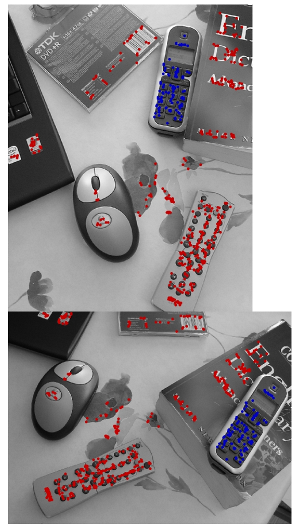
In the last example, two photographs are used (Figure 6(a)), representing some identical objects in different context: a cereal box and three similar cans. Nothing distinguishes these three cans, so that 9 corresponding transformations are theoretically possible between the two pictures. By using MAC-RANSAC with a planar transformation (similarity, affine or projective transform), a single group corresponding to the cereal box is correctly detected, and 9 groups of correspondences are found between the cans (Figure 6(b)).

6. Summary and future work

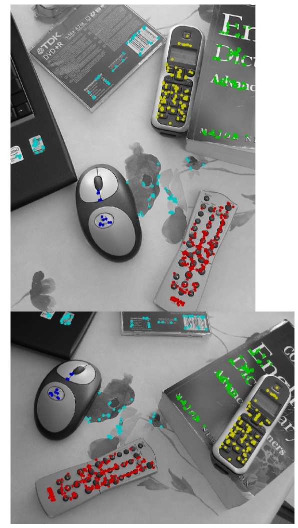
In this paper we have proposed a robust algorithm for multiple object recognition and 3D transformation estimation. The MAC-RANSAC algorithm avoids strong limitations and drawbacks of the sequential RANSAC approach; it does not require the setting of detection parameters, except the maximal number of iterations i_{max} which is generally fixed under time complexity considerations. An interesting aspect of this approach, which has not been presented



(a) Pair of photographs analysed



(b) Grouping with epipolar geometry constraints



(c) Grouping with projective geometry constraints

Figure 4. *Application of the MAC-RANSAC algorithm for a non-static scene with a camera motion.*

here, is the automatic selection of geometrical models for object pose estimation. Indeed, this may be achieved by comparing the NFA's corresponding to different geometrical models and provide an interesting alternative to model selection criteria relying on information theory, a direction that we are currently investigating.

Acknowledgements The authors acknowledge the support of the French *Agence Nationale de la Recherche* (ANR), under grant ANR-09-CORD-003: "Calibration en vision stéréo par méthodes statistiques" (Callisto).

References

[1] O. Chum and J. Matas. Matching with PROSAC - progressive sample consensus. In *Proc. CVPR*, pages 220–226, 2005.

[2] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[3] P. Hough. Machine analysis of bubble chamber pictures. In *International Conference on High Energy Accelerators and Instrumentation*, pages 554–556, 1959.

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

(a) The target object (left picture) is present several times in the second picture.

(b) Grouping with MAC-RANSAC: the 28 occurrences of the target objet are recognized.

Figure 5. *Application of the MAC-RANSAC algorithm for multiple detection of the same objet.*

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

[6] 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*, 57(3):201–218, 2004.

[7] P. Monasse. *Morphological representation of digital images and application to registration*. PhD thesis, CMLA, ENS de Cahan, 2000.

[8] N. Noury, F. Sur, and M.-O. Berger. Fundamental matrix estimation without prior match. In *ICIP*, 2007.

[9] M. Pollefeys. Séquence vidéo *Leuven castel*, disponible à : www.cs.unc.edu/~marc/data/castlejpg.zip.

[10] J. Rabin, J. Delon, and Y. Gousseau. A statistical approach to the matching of local features. *SIAM Journal on Imaging Sciences*, 2(3):931–958, 2009.

[11] C. V. Stewart. MINPRAN: A New Robust Estimator for Computer Vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(10):925–938, 1995.

[12] C. V. Stewart. Bias in robust estimation caused by discontinuities and multiple structures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(8):818–833, 1997.

(a) Two different scenes sharing some objets.

(b) The MAC-RANSAC algorithm is able to detect the 10 different groups of correspondences.

Figure 6. *Mutiple detection of several objects in presence of clutter.*

[13] R. Toldo and A. Fusiello. Robust multiple structures estimation with j-linkage. In *ECCV (1)*, pages 537–547, 2008.

[14] P. H. S. Torr and D. W. Murray. Stochastic motion clustering. In *ECCV (2)*, pages 328–337, 1994.

[15] P. H. S. Torr and A. Zisserman. MLESAC: a new robust estimator with application to estimating image geometry. *Comput. Vis. Image Underst.*, 78(1):138–156, 2000.

[16] E. Vincent and R. Laganiere. Detecting planar homographies in an image pair. *Image and Signal Processing and Analysis*, pages 182–187, 2001.

[17] W. Zhang and J. Kosecka. Generalized RANSAC Framework for Relaxed Correspondence Problems. In *Proceedings of 3DPVT'06*.

[18] W. Zhang and J. Kosecká. Nonparametric estimation of multiple structures with outliers. In *WDV*, pages 60–74, 2006.

[19] M. Zuliani, C. S. Kenney, and B. S. Manjunath. The multiRANSAC algorithm and its application to detect planar homographies. In *ICIP*, 2005.