

# MAC-RANSAC : reconnaissance automatique d'objets multiples

J. Rabin<sup>1</sup>

J. Delon<sup>1</sup>

Y. Gousseau<sup>1</sup>

L. Moisan<sup>2</sup>

<sup>1</sup> Télécom Paristech, LTCI CNRS  
{rabin,delon,gousseau}@telecom-paristech.fr

<sup>2</sup> Université Paris Descartes, MAP5 CNRS  
lionel.moisan@parisdescartes.fr

## Résumé

*De nombreuses applications en vision par ordinateur (reconstruction 3D, reconnaissance d'objets, construction de mosaïque) reposent sur le groupement de correspondances de points d'intérêt. Nous proposons dans cet article un algorithme de reconnaissance d'objets et d'estimation de pose qui présente la particularité d'être robuste, automatique et de permettre la détection de transformations multiples entre images. Cette méthode s'appuie sur une généralisation de l'approche a contrario de l'algorithme RANSAC proposée dans [8], et sur une approche séquentielle de la détection.*

## Mots Clés

Groupement de correspondances, reconnaissance d'objets multiples, décision a contrario, descripteurs locaux.

## Abstract

*Many applications in the field of computer vision (structure from motion, object recognition, image mosaicing) rely on the grouping of matchings between interest points. We propose an automatic and robust algorithm for object recognition and pose estimation. This method is based on the generalization of the a contrario approach of RANSAC algorithm proposed in [8]. The resulting framework permits the detection of multiple transformations between images.*

## Keywords

Geometric grouping, multiple objet recognition, a contrario decision, local features.

## 1 Problématique

L'application principale visée par cet article est la reconnaissance d'objets : il s'agit à la fois de *détecter* la présence éventuelle d'un même objet entre plusieurs images, et le cas échéant, d'*estimer* sa pose, c'est à dire la transformation géométrique entre les différentes vues. Une approche classique consiste tout d'abord à détecter des points d'intérêt et les coder par des descripteurs locaux (nous considérons dans ce papier les descripteurs SIFT [6]), puis à les mettre en correspondance par un critère de décision automatique (voir par exemple [13]). Enfin, la dernière étape

de décision et d'estimation est réalisée par la détection de groupe de correspondances cohérentes selon une transformation géométrique. Dans cet article, nous proposons une méthode de groupement robuste de telles correspondances, ne nécessitant pas de réglage de paramètres et permettant la reconnaissance de plusieurs objets. Plusieurs cas de figure requièrent en effet une méthode permettant les détections multiples.

- Détection de plusieurs objets dans des photographies ;
- Détection de plusieurs occurrences du même objet dans des photographies ;
- Segmentation du mouvement dans les séquences d'images.

**Travaux antérieurs sur le groupement de correspondances** Plusieurs approches ont été proposées pour la reconnaissance d'un ou de plusieurs objets par groupement de correspondances. Rappelons qu'en présence de nombreuses données aberrantes (*outliers*) ou de plusieurs objets, les méthodes d'estimation directe (moindres carrés, M-estimateurs par exemple) sont à proscrire.

Une première approche consiste à utiliser la transformée de Hough [5] qui définit un cadre naturel à la reconnaissance d'objets multiples. Son principe est de détecter des groupes de consensus par "vote" d'échantillons de  $n$ -uplet de correspondances ; ces votes sont comptabilisés par des accumulateurs quantifiés de l'espace des paramètres de la transformation géométrique considérée. Néanmoins, en raison de sa complexité calculatoire, cette méthode n'est jamais utilisé pour des modèles géométriques complexes comme la géométrie épipolaire ou projective.

Une alternative très populaire en vision par ordinateur est l'algorithme RANSAC (*RANdom SAmple Consensus*), que nous rappellerons dans la prochaine section. Son principe est de détecter une **transformation unique** parmi un ensemble de données par échantillonnage aléatoires de  $n$ -uplet. Pour chaque échantillon tiré, on estime la transformation correspondante et on teste son adéquation aux données restantes. Cette approche présente l'énorme avantage d'être rapide et robuste à la présence d'outliers. Cependant, son utilisation implique le réglage de paramètres sensibles. Quelques alternatives intéressantes ont été proposées pour simplifier ce réglage : MSAC et MLESAC [18],

MINPRAN [14] notamment. En particulier, nous nous intéressons dans cet article à la méthode robuste proposée dans [8]. Cette approche s’appuie sur le principe de la détection *a contrario* [2, 3]. Elle permet la détection robuste d’un groupe de points de contrôle, **sans aucun réglage de paramètre de détection**, dans le cadre de la géométrie épipolaire. Nous rappellerons en section 3 le principe de cette méthode que nous appelons AC-RANSAC (A *Contra*rio *RAN*dom *SAM*ple *CON*sensus). En vue d’une application à la reconnaissance d’objets, nous étendrons cette méthode au cas des transformations planes. Nous introduirons ensuite une nouvelle méthodologie pour adapter cette approche au groupement de mises en correspondance de descripteurs locaux de type SIFT [6].

L’inconvénient majeur de RANSAC –et des variantes précédemment citées– est la restriction à la détection d’un unique objet. Quelques approches ont été proposées pour permettre la détection de plusieurs groupes avec RANSAC, que l’on peut –pour être bref– classer suivant deux catégories. Une première stratégie [22, 16, 17, 21] consiste à détecter *simultanément* tous les objets, en fusionnant les différents groupes sélectionnés par RANSAC. Une seconde stratégie [14, 19] consiste à *utiliser séquentiellement* RANSAC, pour détecter successivement chaque groupe. Nous étudierons en section 4 les avantages de l’utilisation séquentielle de AC-RANSAC pour la détection robuste de plusieurs groupes. Nous présenterons ensuite deux nouveaux critères permettant de s’affranchir de certaines limitations propres à l’utilisation séquentielle de RANSAC. L’algorithme MAC-RANSAC étant ainsi défini, nous en illustrerons ensuite l’intérêt pour différents cas de figure dans une section expérimentale.

## 2 Notations et rappels sur RANSAC

Sans perdre en généralité, nous considérons dorénavant la procédure de reconnaissance d’objets pour une paire d’images notées  $I$  et  $I'$ . On désigne par  $\mathcal{P}$  et  $\mathcal{P}'$  les plans images correspondant à  $I$  et  $I'$ . On note également  $A$  et  $D$  (resp.  $A'$  et  $D'$ ) l’aire et la diagonale de l’image  $I$  (resp.  $I'$ ).

On note  $\mathcal{C} = \{(m_i, m'_i), i = 1, \dots, N\}$  l’ensemble des  $N$  mises en correspondances que l’on suppose déjà obtenues (par exemple avec [13]), où  $\{m_i\}$  et  $\{m'_i\}$  sont respectivement les points d’intérêt de  $\mathcal{P}$  et  $\mathcal{P}'$ .

**Algorithme RANSAC** Dans l’algorithme RANSAC [4], une transformation  $\mathcal{T}_{S'}$  est définie à partir d’un sous-ensemble  $S'$  de  $\mathcal{C}$  de taille  $n$ . De manière itérative, des  $n$ -uplet  $S'$  sont échantillonnés aléatoirement dans  $\mathcal{C}$ , chaque transformation  $\mathcal{T}_{S'}$  étant évaluée sur les correspondances restantes de  $\mathcal{C}$ . Le *consensus* est défini comme le groupe  $S \subset \mathcal{C}$  de correspondances dont les erreurs résiduelles pour une transformation sont plus petites qu’un seuil. La qualité de ce groupe est définie par son cardinal. L’algorithme s’arrête au terme de  $i_{max}$  itérations. Le meilleur groupe identifié (le plus grand) est validé si son cardinal est plus grand qu’un second seuil.

### Estimation d’une transformation, erreur résiduelle

Nous considérons dans cet article plusieurs transformations : les transformations planes (similitude, transformation affine et projective) et la géométrie épipolaire. Par la suite,  $\mathcal{T}_{S'}$  désignera une transformation correspondant à l’un de ces modèles.

Dans le cas de la géométrie épipolaire,  $\mathcal{T}_{S'}$  représente la matrice fondamentale  $F_{S'}$ , qui peut être estimée à partir d’un groupe  $S'$  de  $n = 7$  correspondances  $\{(m_i, m'_i)\}$ , à partir des relations  $m_i'^T F_{S'} m_i = 0$  (en coordonnées homogènes). Les contraintes épipolaires impliquent que l’erreur résiduelle pour un couple de points  $(m, m')$  est déterminée par la distance des points  $m'$  et  $m$  aux lignes épipolaires  $F_{S'} m$  dans  $\mathcal{P}'$  et  $F_{S'}^T m'$  dans  $\mathcal{P}$  respectivement. Dans le cas des transformations planes,  $\mathcal{T}_{S'}$  représente la matrice  $H_{S'}$ , estimée à partir d’un groupe  $S'$  de  $n$  correspondances, selon la relation  $m'_i = H_{S'} m_i$  (en coordonnées homogènes). Rappelons que  $n = 2$  pour une similitude,  $n = 3$  pour une transformation affine et  $n = 4$  pour une homographie. L’erreur géométrique pour un couple de points  $(m, m')$  est cette fois définie comme la distance des points  $m'$  et  $m$  aux projections  $H_{S'} m$  dans  $\mathcal{P}'$  et  $H_{S'}^{-1} m'$  dans  $\mathcal{P}$  respectivement.

Dans la suite de cet article nous notons  $d(m', \mathcal{T}_{S'} m)$ , la distance euclidienne calculée dans  $\mathcal{P}'$  entre le point  $m'$  et le terme  $\mathcal{T}_{S'} m$ , qui représente un point dans le cas d’une transformation plane, et une droite dans le cas de la géométrie épipolaire. Dans  $\mathcal{P}$ , on définit de manière analogue la distance  $d(m, \mathcal{T}_{S'}^{-1} m')$ .

## 3 Généralisation de AC-RANSAC

Dans [8] est proposée une nouvelle mesure de qualité des groupes testés lors du processus d’échantillonnage de RANSAC. Cette mesure, appelée *rigidité*, est définie pour l’évaluation et l’optimisation de la matrice fondamentale entre deux vues stéréoscopiques. Un critère de décision *a contrario* est alors utilisé pour seuiliser automatiquement cette mesure de qualité. Nous ferons référence à cet algorithme comme AC-RANSAC (A *Contra*rio *RAN*SAC).

Dans cette section nous redéfinissons ce cadre de travail de manière à prendre en compte, d’une part, les transformations planes, et d’autre part le processus de mise en correspondance automatique de descripteurs locaux qui est utilisé en amont.

### 3.1 Critère de validation d’un groupe

**Hypothèse nulle** Rappelons que  $\mathcal{C}$  désigne l’ensemble  $\{(m_i, m'_i), i = 1, \dots, N\}$  des points d’intérêts appariés entre deux images  $I$  et  $I'$ , où les points  $m_i$  (et  $m'_j$ ) ne sont pas nécessairement deux à deux distincts. Ce problème de redondance sera considéré en section 3.2.

On souhaite déterminer un sous-groupe de ces correspondances qui peut être expliqué par une unique transformation. Pour estimer une telle transformation dans le cadre de la méthodologie *a contrario* [3], on définit tout d’abord l’hypothèse nulle  $\mathcal{H}_0$  qui décrit une distribution “généri-

que” de correspondances aléatoires  $(\mathbf{m}_i, \mathbf{m}'_i)$ ,  $i = 1, \dots, N$  pour lesquelles aucun groupement ne doit être validé. Ensuite, un groupe de correspondances réelles est considéré comme significatif s’il est très improbable d’observer un tel groupe sous  $\mathcal{H}_0$ . Cette méthodologie *a contrario* permet de limiter les *erreurs de type I* (ou fausses alarmes), c’est-à-dire les validations de groupes de correspondances qui vérifient l’hypothèse nulle  $\mathcal{H}_0$ .

Dans [8], l’hypothèse nulle  $\mathcal{H}_0$  est définie de la manière suivante,  $N \in \mathbb{N}^+$  étant fixé :

**Définition 1 (Hypothèse Nulle)** *Un ensemble  $\mathbf{C}$  de  $N$  correspondances aléatoires  $(\mathbf{m}_i, \mathbf{m}'_i)$ ,  $i = 1, \dots, N$  suit l’hypothèse nulle  $\mathcal{H}_0$  lorsque*

- les points  $\mathbf{m}_i$  et  $\mathbf{m}'_j$ ,  $i, j = 1, \dots, N$  sont des variables aléatoires mutuellement indépendantes ;
- les points  $\mathbf{m}_i$ ,  $i = 1, \dots, N$  sont uniformément distribués sur l’image  $I$  et les points  $\mathbf{m}'_j$ ,  $j = 1, \dots, N$  sont uniformément distribués sur l’image  $I'$ .

Nous allons maintenant considérer la probabilité qu’un groupe de correspondances suive l’hypothèse nulle, et définir la mesure de qualité, appelée rigidité, qui lui est associée selon le modèle géométrique utilisé.

**Probabilité sous l’hypothèse nulle** Soit  $\mathbf{C}$  un ensemble de  $N$  correspondances aléatoires suivant le modèle de fond  $\mathcal{H}_0$ , et  $\mathbf{S}'$  un sous ensemble de  $\mathbf{C}$ , tel que  $\#\mathbf{S}' = n$ . Soit  $\mathbf{S}$  un sous-ensemble de  $\mathbf{C}$  tel que  $\mathbf{S} \cap \mathbf{S}' = \emptyset$ . Dans le cadre de la géométrie épipolaire ( $n = 7$ ), Moisan et Stival [8] définissent la **rigidité** de la matrice fondamentale  $F_{\mathbf{S}'}$  pour les correspondances de  $\mathbf{S}$  comme l’*erreur résiduelle normalisée maximale* sur tous les points de  $\mathbf{S}$  :

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

Cette rigidité mesure la cohérence entre l’ensemble  $\mathbf{S}$  et la matrice fondamentale  $F_{\mathbf{S}'}$ . Ils montrent alors que sous l’hypothèse nulle  $\mathcal{H}_0$  sur l’ensemble des correspondances aléatoires, la probabilité d’observer une rigidité  $\alpha(\mathbf{S}, F_{\mathbf{S}'})$  plus petite que  $\alpha$  est bornée par  $\alpha^{\#\mathbf{S}}$  :

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

Dans le cas d’une transformation plane  $H_{\mathbf{S}'}$ , l’inégalité (1) reste valable (en remplaçant  $F_{\mathbf{S}'}$  par  $H_{\mathbf{S}'}$ ) si l’on définit la rigidité  $H_{\mathbf{S}'}$  pour les correspondances de  $\mathbf{S}$  comme :

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

La cohérence d’un ensemble  $S$  de correspondances *réelles* selon une transformation  $\mathcal{T}$  est ainsi évaluée par la probabilité que la rigidité aléatoire  $\alpha(\mathbf{S}, \mathcal{T}_{S'})$  soit plus petite que la *rigidité observée*  $\alpha(S, \mathcal{T})$  sous l’hypothèse nulle  $\mathcal{H}_0$ .

**Critère de validation** En utilisant le cadre de travail des méthodes *a contrario*, les auteurs de [8] définissent un critère de sélection en seillant cette probabilité de manière à contrôler le nombre de fausses alarmes, *i.e.* le nombre de groupes validés qui suivent le modèle de fond. Ils introduisent pour cela une mesure de qualité, notée « NFA »,

dont on donne ici une définition générique pour les transformations planes et la géométrie épipolaire.

**Définition 2** *Soit  $\mathcal{C} = \{(m_i, m'_i), i = 1, \dots, N\}$  un ensemble de  $N$  correspondances entre les images  $I$  et  $I'$ . Soit  $S$  un sous ensemble de  $\mathcal{C}$ , constitué de  $\#S = K$  correspondances. Pour  $\varepsilon > 0$  donné, l’ensemble  $S$  est dit  $\varepsilon$ -significatif s’il existe un entier  $n \in \{2, 3, 4, 7\}$  et un sous ensemble  $S'$  de  $\mathcal{C}$ , tel que  $\#S' = n$ ,  $S' \cap S = \emptyset$  et*

$$NFA(S, S') := \gamma(N - n) \binom{N}{K} \binom{N - K}{n} (\alpha(S, \mathcal{T}_{S'}))^K \leq \varepsilon,$$

où  $\mathcal{T}_{S'}$  est la transformation associée au groupe  $S'$ <sup>1</sup>.

Pour cette définition, rappelons que la rigidité  $\alpha(S, \mathcal{T}_{S'})$  est définie différemment selon la transformation géométrique  $\mathcal{T}_{S'}$  considérée. Le coefficient  $\gamma$  désigne le nombre maximal de transformations  $\mathcal{T}_{S'}$  estimées à partir de  $n$  correspondances, soit  $\gamma = 1$  pour la géométrie plane, et  $\gamma = 3$  pour la géométrie épipolaire<sup>1</sup>.

Avec cette mesure de qualité, un groupe  $S$  dont la rigidité est mesurée selon une transformation  $\mathcal{T}_{S'}$  est d’autant plus significatif que la quantité  $NFA(S, S')$  est faible. La quantité NFA permet à la fois de définir une mesure de qualité sur les groupes (*i.e.* le groupe le plus significatif est celui minimisant le NFA) et de fournir un **critère de validation** par le choix du seuil  $\varepsilon$ . La quantité  $\varepsilon$ , qui sera toujours choisie égale à 1 dans les expériences, permet de définir des seuils de détections adaptatifs sur les rigidités  $\alpha(S, \mathcal{T}_{S'})$  des groupes testés, *via* la définition du NFA.

**ORSA** Dès qu’un groupe  $S_{\text{opt}}$   $\varepsilon$ -significatif est trouvé, les auteurs de [8] propose de changer la méthode d’échantillonnage. L’échantillonnage ORSA consiste à tirer les groupes  $S'$  parmi l’ensemble  $S_{\text{opt}}$ . Cette heuristique permet d’accélérer l’optimisation de l’estimation de la transformation optimale  $\mathcal{T}_{\text{opt}}$  d’un ordre de grandeur. La supériorité de cette approche sur les algorithmes RANSAC [4] et MSAC [18], a été montrée dans [11] à la fois en termes de robustesse et de précision.

## 3.2 Hypothèse nulle et descripteurs locaux

Dans [8], l’algorithme AC-RANSAC est utilisé à partir de correspondances de points de contrôle sélectionnées sur des paires d’images stéréoscopiques. Dans [10], les auteurs proposent une extension de AC-RANSAC où le processus de mise en correspondance de descripteurs SIFT [6] est réalisé conjointement avec l’estimation de la matrice fondamentale. Dans cet article, nous considérons le cas de correspondances établies automatiquement entre descripteurs locaux par analyse de leur similarité [6, 13]. Nous présentons dans ce paragraphe quelques modifications de AC-RANSAC visant à l’adapter à ce cas de figure.

<sup>1</sup>Dans le cadre de la géométrie épipolaire, si l’on utilise l’algorithme des 7-points, jusqu’à trois matrices fondamentales  $F_{S'}$  peuvent être estimées [8]. La transformation utilisée pour la définition du NFA est celle minimisant la rigidité  $\alpha(S, F_{S'})$ .

Rappelons que le modèle de fond pour les correspondances aléatoires  $\mathbf{C}$  repose sur l'hypothèse nulle  $\mathcal{H}_0$ , selon laquelle les points appariés  $m_i \in I$  et  $m'_i \in I'$  sont des variables aléatoires mutuellement indépendantes. Or, il existe deux raisons, que nous allons décrire, pour lesquelles la construction même des correspondances entre descripteurs locaux contredit déjà cette hypothèse  $\mathcal{H}_0$ . Pour ces deux raisons, on risque donc de détecter des groupes significatifs à tort. La première est la présence de **correspondances multiples** : un point  $m$  de l'image  $I$  peut avoir plusieurs correspondants dans  $I'$ , et réciproquement. Cette propriété a l'immense avantage de permettre la reconnaissance d'un objet apparaissant plusieurs fois dans une image, ou d'objets présentant des zones répétées. Cependant, elle contredit manifestement l'hypothèse nulle  $\mathcal{H}_0$ . Le deuxième phénomène incompatible avec l'hypothèse  $\mathcal{H}_0$  est la **redondance des points d'intérêt** : tous les détecteurs de points (ou de régions) d'intérêt utilisés pour construire des descripteurs locaux (SIFT [6], MSER [7],...) ont tendance à détecter certaines structures de manière *redondante*. Une même structure, typiquement un coin, est alors représentée par plusieurs points d'intérêt, qui diffèrent légèrement en position et en échelle. Ces points redondants induisent des mises en correspondance non-indépendantes. Pour éviter de détecter de faux groupes dus à ces deux phénomènes, intrinsèques aux méthodes de mise en correspondance par descripteurs locaux, nous proposons deux principes de filtrage des correspondances.

**Définition 3 (Principe de maximalité)** *Pour une transformation donnée par  $S'$ , une seule mise en correspondance par point d'intérêt peut être sélectionnée dans un groupe. Cette correspondance est celle qui minimise l'erreur résiduelle.*

Notons que les correspondances alternatives d'un point  $m$  ne sont pas pour autant éliminées et peuvent être réutilisées dans l'optique d'une recherche itérative d'objets, comme décrit dans la section suivante.

Pour le principe suivant, on suppose qu'une mesure de qualité est associée à chaque appariement  $(m, m')$ . C'est le cas par exemple dans la méthode de mise en correspondance présentée dans [13], que nous utilisons pour les expériences de cet article. On suppose également qu'à chaque descripteur est associée une échelle caractéristique  $\Delta$ , comme c'est le cas pour les descripteurs de type SIFT.

**Définition 4 (Principe d'exclusion)** *Deux correspondances  $c_i = (m_i, m'_i)$  et  $c_j = (m_j, m'_j)$  de points d'intérêt sont considérées comme **redondantes** si l'une des deux assertions est vraie :*

$$- m_i = m_j \text{ et } \|m'_i - m'_j\|_2 < \min\{\Delta_i, \Delta_j\}$$

*ou*

$$- m'_i = m'_j \text{ et } \|m_i - m_j\|_2 < \min\{\Delta'_i, \Delta'_j\}$$

*où  $\|\cdot\|_2$  désigne la norme euclidienne, et  $\Delta_k$  représente l'échelle caractéristique du point d'intérêt  $m_k$ . Si deux correspondances sont redondantes, on ne conserve que celle dont la mesure de qualité est la meilleure.*

## 4 Reconnaissance multiple avec l'algorithme MAC-RANSAC

L'algorithme AC-RANSAC a été défini de manière à détecter une transformation unique pour un ensemble de données entachées d'erreurs. Or, nous avons vu au § 1 que plusieurs applications nécessitent la détection de plusieurs groupes de correspondances. Dans cette section, nous présentons tout d'abord les limitations d'une utilisation séquentielle de RANSAC, puis nous proposons un nouvel algorithme, MAC-RANSAC (Multiple A Contrario RANSAC), qui permet de s'en affranchir.

### 4.1 Avantages et limitations de l'utilisation séquentielle de AC-RANSAC

L'algorithme RANSAC peut être utilisé de manière itérative, en éliminant à chaque étape les correspondances du dernier groupe détecté. La procédure s'arrête lorsqu'aucun nouveau groupe n'est détecté. Cependant, cette approche présente de nombreux défauts :

- **Détections de fausses transformations.** Ceci désigne le fait de valider des groupes de correspondances incorrects, constitués exclusivement de faux appariements.
- **Fusion de transformations proches.** Plusieurs groupes ayant *des transformations distinctes peuvent être fusionnés*, la transformation obtenue étant très approximative. Ce phénomène, étudié par [15] est du à l'aspect « glouton » de RANSAC, et en particulier de son critère d'optimisation.
- **Sur-segmentation en de multiples groupes.** Ce phénomène, opposé au précédent, se traduit par le fait de détecter pour un unique objet de multiples groupes au lieu d'un unique ensemble cohérent. Cela se produit lorsque la tolérance sur l'erreur résiduelle est sous-estimée.
- **Détections d'« échos » liés à l'autosimilarité.** Ce terme désigne la détection de plusieurs transformations artificielles qui viennent faire écho à une unique transformation réelle. Cela se produit lorsqu'un objet est autosimilaire : un mur de brique, la façade d'un bâtiment, etc.

Le fait de remplacer RANSAC par AC-RANSAC permet d'obtenir un critère d'arrêt très robuste, ce qui nous affranchit du problème de la détection de faux groupes. Par ailleurs, le critère de validation permet de définir automatiquement les seuils de validation sur l'erreur résiduelle (rigidité), ce qui évite le problème de sur-segmentation lié au choix d'une tolérance spatiale trop faible.

Les deux autres problèmes de détections (fusion de groupes et autosimilarité) sont cependant toujours présents, et nous présentons deux nouveaux critères permettant de les éviter dans les paragraphes § 4.3 et § 4.4.

### 4.2 Vue d'ensemble de MAC-RANSAC

Avant d'introduire ces deux nouveaux critères, nous donnons en table 1 une vue d'ensemble de l'algorithme séquentiel MAC-RANSAC. Rappelons que les seuils de validation utilisés sont toujours fixés à  $\varepsilon = 1$ .

TAB. 1 – Algorithme de reconnaissance d’objets multiples

MAC-RANSAC	
<b>Entrées :</b> Ensemble $\mathcal{C}$ de correspondances non redondantes (critère 4) et nombre d’itérations $i_{max}$ . Initialisation : $i := 0$ et liste $\mathcal{S} := \{\emptyset\}$ .	
1)	<b>Détection :</b> Tant que $i < i_{max}$ , tirage de $S' \subset \mathcal{C}$ puis recherche de $S \subset \mathcal{C} \setminus S'$ minimisant $NFA(S, S')$ . – Si $NFA(S, S') < 1$ , $(S_{opt}, S'_{opt}) := (S, S')$ puis étape 2). – Sinon $i := i + 1$ . Si $i = i_{max}$ , arrêt de l’algorithme.
2)	<b>Optimisation (ORSA) :</b> Répéter $i_{max}/10$ fois le tirage de $S' \subset S_{opt}$ et la recherche de $S \subset \mathcal{C} \setminus S'$ minimisant $NFA(S, S')$ . Si $NFA(S, S') < NFA(S_{opt}, S'_{opt})$ , $(S_{opt}, S'_{opt}) := (S, S')$ .
3)	<b>Détection de fusion de groupes :</b> Recherche d’un sous-groupe optimal de $S_{opt}$ (critère 6). – Si détection de fusion, obtention de 2 couples 1-significatifs $(S_1, S'_1)$ et $(S_2, S'_2)$ . – Sinon, $S_1 := S_{opt}$ et $S_2 := \emptyset$ .
4)	<b>Filtrage :</b> Élimination des correspondances autosimilaires avec $S_1$ dans $\mathcal{C}$ (critère 5). puis définition de $\mathcal{C} := \mathcal{C} \setminus S_1$ .
5)	<b>Itération :</b> Ajout de $S_1$ à la liste $\mathcal{S}$ , $i := 0$ . – Si $S_2 = \emptyset$ , retour à l’étape 1) – Sinon, $(S_{opt}, S'_{opt}) := (S_2, S'_2)$ puis étape 2).
<b>Sortie :</b> Liste de groupes disjoints $\mathcal{S}$ .	

### 4.3 Filtrage des groupes autosimilaires

À notre connaissance, peu de publications [20, 19] s’intéressent au problème de l’autosimilarité pour la reconnaissance d’objets. Monasse montre dans [9] que les correspondances multiples liées à l’autosimilarité d’un objet sont responsables de détections artificielles faisant écho à l’unique “vraie” transformation. Un même objet est donc détecté suivant différentes poses.

Le premier groupe détecté, *i.e.* celui ayant le plus grand nombre de correspondances, correspond théoriquement à la transformation réelle qui nous intéresse. Les correspondances que l’on qualifie d’*autosimilaires* sont alors identifiées à l’aide du critère suivant :

**Définition 5 (Correspondances autosimilaires)** Soit  $\mathcal{G}$  un groupe validé, et  $\mathcal{C}$  l’ensemble des correspondances restantes, de telle sorte que  $\mathcal{C} \cap \mathcal{G} = \emptyset$ . Une correspondance  $c_i = (m_i, m'_i) \in \mathcal{C}$  entre un point  $m_i$  de  $I$  et un point  $m'_i$  de  $I'$  est dite *autosimilaire vis-à-vis du groupe  $\mathcal{G}$*  si les deux conditions suivantes sont simultanément vérifiées :

$$\cdot \exists m \in \mathcal{G} \text{ tel que } \|m - m_i\|_2 < \min\{\Delta, \Delta_i\},$$

et

$$\cdot \exists m' \in \mathcal{G} \text{ tel que } \|m' - m'_i\|_2 < \min\{\Delta', \Delta'_i\}.$$

où  $\|\cdot\|_2$  désigne la norme euclidienne, et  $\Delta_k$  représente l’échelle caractéristique du point d’intérêt  $m_k$ .

Désormais en itérant MAC-RANSAC, dès qu’un nouveau groupe est validé, les correspondances autosimilaires de ce groupe sont éliminées des correspondances restantes.

### 4.4 Détection de la fusion de groupes

Nous verrons dans la partie expérimentale que lorsque plus d’une transformation est en jeu, il existe des situations où le critère de validation entraîne la détection d’un groupe résultant de la fusion de plusieurs transformations distinctes mais néanmoins similaires. Pour éviter ce phénomène, nous proposons un nouvel algorithme de détection des fusions.

**Critère de détection de fusion de deux groupes** Pour simplifier l’étude du problème, nous nous intéressons dans un premier temps au cas de la fusion de deux transformations. Soient  $S_0$  un groupe de correspondances,  $S_1$  et  $S_2$  deux sous-groupes de  $S_0$  distincts, tels que  $S_1 \cap S_2 = \emptyset$  et  $S_0 \supset S_1 \cup S_2$ . On parle de fusion de groupes lorsque  $S_0$  est validé alors que les transformations de  $S_1$  et  $S_2$ , respectivement notées  $T_1$  et  $T_2$ , sont *suffisamment* différentes. La transformation  $T_0$  du groupe  $S_0$  est alors une transformation “moyenne” de  $T_1$  et  $T_2$ .

La raison pour laquelle le groupe  $S_0$  a été validé est que sa significativité (NFA) est plus faible que celles de chacun des deux sous-groupes  $S_1$  et  $S_2$ , bien qu’ils aient une plus faible rigidité.

$$NFA(S_0, S'_0) < \min\{NFA(S_1, S'_1), NFA(S_2, S'_2)\}.$$

Ce résultat est inhérent à l’algorithme AC-RANSAC qui cherche à minimiser le NFA pour chaque groupe *via* une procédure de minimisation gloutonne. Afin de pouvoir détecter la fusion de deux groupes, nous devons donc définir un critère comparant le groupe  $S_0 = S_1 \cup S_2$  avec les deux groupes  $S_1$  et  $S_2$  simultanément.

En nous inspirant de l’approche de [1], où un tel critère est défini, nous proposons le critère de découpage suivant :

**Définition 6 (Critère de découpage)** Soit  $S_0$  un groupe  $\varepsilon$ -significatif de  $\mathcal{C}$ , de transformation  $T_{S_0}$ . S’il existe  $S_1$  et  $S_2$ , deux sous-ensembles disjoints de  $S_0$ , ainsi que deux *n-uplets* disjoints  $S'_1$  et  $S'_2$  de  $S_0 \setminus \{S_1 \cup S_2\}$ , tels que les conditions suivantes sont vérifiées :

$$\cdot NFA(S_1, S'_1) \leq \varepsilon \text{ et } NFA(S_2, S'_2) \leq \varepsilon$$

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

alors, les 2 groupes  $S_1$  et  $S_2$  sont validés en tant que deux objets distincts à la place de  $S_0$ .

Afin de pouvoir exploiter le critère de découpage 6, nous allons maintenant définir une stratégie de découpage récursif en sous-groupes.

**Recherche récursive de découpage dyadique en sous-groupes** Nous proposons un algorithme dont le but est d’identifier un seul des sous-groupes impliqués dans la fusion. Pour cela, nous définissons une stratégie de découpage en deux sous-groupes qui est ensuite appliquée récursivement sur le groupe  $S_0$ .

Nous définissons  $S_1$  comme le plus petit des deux sous-groupes  $S_1$  et  $S_2$  appartenant à  $S_0$ . Comme le remarque Stewart dans [14], on vérifie alors la relation suivante :  $\#S_1 \leq \#S_0/2 \leq \#S_2$ . Le sous-groupe  $S_1$  est obtenu

en cherchant (par échantillonnage aléatoire) un  $n$ -uplet  $S'_1 \subset S_0$  qui minimise la quantité  $NFA(S_1, S'_1)$  sous la contrainte que  $\#S_1 \leq \frac{\#S_0}{2}$ . Le sous-groupe  $S_2$  est ensuite obtenu en cherchant un  $n$ -uplet  $S'_2$  parmi les correspondances restantes de  $S_0$  qui minimise  $NFA(S_2, S'_2)$ . Le critère 6 nous permet alors de savoir si ce découpage est viable.

Tant que le critère valide le découpage, on applique **récurivement** cette stratégie de découpage en deux sous-groupes au groupe  $S_1^{(k)}$  qui vient d'être identifié à l'itération  $k$ . Ce principe est illustré en figure 1 dans le cas de la fusion de 5 groupes.

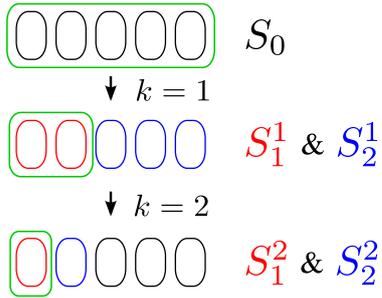


FIG. 1 – Détermination d'un sous-groupe optimal par découpage récursif en 2 sous-groupes.

Le processus s'arrête lorsque le dernier sous-ensemble  $S_1^{(k)}$  trouvé correspond à une transformation unique (*i.e.* lorsque l'on n'a pas trouvé 2 nouveaux sous-groupes  $\varepsilon$ -significatifs, ou lorsque le critère de découpage rejette son découpage en  $S_1^{(k+1)}$  et  $S_2^{(k+1)}$ ).

**Remarque 1 :** Si  $S_0$  est découpé en  $S_1^{(k)}$  et  $S_2^{(k)}$ , seul le groupe  $S_1^{(k)}$  est validé. Le groupe  $S_2^{(k)}$  étant  $\varepsilon$ -significatif, il est quant à lui considéré comme une détection : l'algorithme MAC-RANSAC passe donc en phase d'optimisation (étape 2 en table 1).

**Remarque 2 :** Nous avons vu que le groupe  $S_0$ , s'il résulte d'une fusion de plusieurs groupes, est estimé à partir d'une transformation très approximative  $\mathcal{T}_{S_0}$ . Le groupe  $S_0$  ne capture donc pas nécessairement l'ensemble des correspondances des différents objets regroupés. Le groupe  $S_1^{(k)}$  validé risque alors d'être incomplet (risque de sursegmentation). Pour éviter cela, on lui ajoute toutes les correspondances  $\tilde{S}_1$  de l'ensemble  $\mathcal{C} \setminus S_0$  dont la rigidité est telle que  $\alpha(\tilde{S}_1, S'_1) \leq \alpha(S, S'_1)$ .

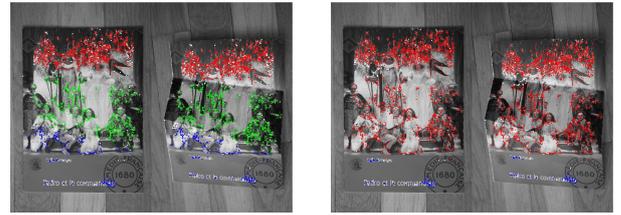
## 5 Évaluation expérimentale

Dans cette partie expérimentale, nous utilisons la procédure de mise en correspondance de descripteurs locaux décrite dans [13], avec un seuil de détection  $\varepsilon = 1$ . L'algorithme MAC-RANSAC est utilisé pour différentes transformations, avec un seuil de détection également fixé à  $\varepsilon = 1$ . Nous illustrons tout d'abord par la figure 2 l'intérêt du critère de découpage en sous-groupes. Une affiche est photographiée avant et après pliage en trois parties. On applique

MAC-RANSAC pour la recherche de transformation projective. Sans utilisation du critère de sélection de sous-groupes (figure 2(c)), deux groupes sont détectés, dont le plus grand (en rouge) résulte de la fusion de trois transformations. L'utilisation de notre procédure de découpage récursif en sous-groupes permet au contraire de grouper avec précision les correspondances selon trois ensembles corrects (figure 2(b)).



(a) Paire d'images analysée



(b) Groupement de correspondances avec critère de découpage

(c) Groupement de correspondances sans critère de découpage

FIG. 2 – Illustration de l'intérêt de la procédure de découpage récursif en sous-groupes.

Pour illustrer la robustesse de notre approche pour la multi-détection, nous examinons maintenant dans les quatre expériences suivantes les différents cas de figures évoqués en introduction, pour différentes transformations géométriques.

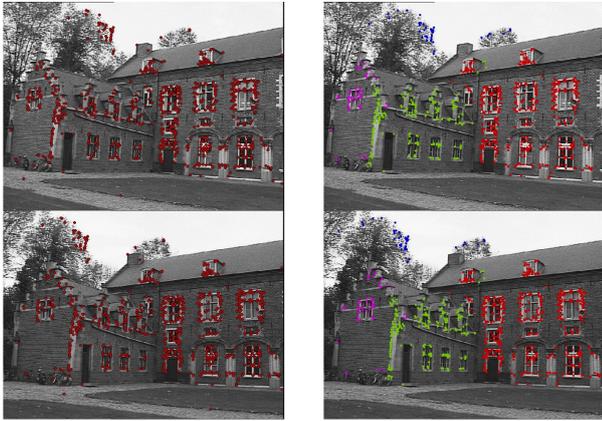
La figure 3(a) montre deux images tirées d'un film [12], où la caméra est mobile et la scène fixe. Cette scène représente un bâtiment en forme de 'L' ayant plusieurs autosimilarités. Nous utilisons MAC-RANSAC avec la transformation épipolaire (figure 3(b)) et avec l'homographie (figure 3(c)). On vérifie qu'un seul groupe est détecté avec la géométrie épipolaire, tandis qu'en considérant la géométrie projective, on détecte trois groupes différents correspondant aux différents plans du bâtiment, ainsi qu'un groupe supplémentaire lié aux arbres en arrière plan (groupe en bleu). Notons par ailleurs qu'il n'y a pas de détection d'échos.

Dans la figure 4(a), deux photographies sont prises d'une scène où un seul objet a bougé entre les deux prises de vue. Avec la géométrie épipolaire (figure 4(b)), deux groupes sont alors identifiés : un groupe dominant (en rouge) correspondant à la partie statique de la scène, et un groupe (en bleu) correspondant au téléphone qui a été déplacé entre les deux prises de vues. Si l'on considère la transformation homographique (figure 4(b)), on retrouve le même groupe pour le téléphone (en jaune), tandis que le reste de la scène est fragmenté en différents plans.

La figure 5(a) montre une paire de photographies, dont la première représente une canette de soda, et la seconde une scène avec 28 canettes ayant le même logo. Le résultat du



(a) Deux vues d'une scène 3D (tirées de [12])



(b) Un unique groupe est détecté avec la géométrie épipolaire (c) 4 groupes sont obtenus avec l'homographie

FIG. 3 – Application de MAC-RANSAC pour une scène statique, avec mouvement de la caméra.

groupement de correspondances est donné en figure 5(b) selon une transformation plane (la similitude, la transformation affine et l'homographie donnent un résultat identique). On vérifie que l'algorithme permet de détecter les 28 occurrences de la canettes. Insistons sur le fait que chacun des groupes détectés correspond à une faible proportion de l'ensemble des correspondances. En particulier, les deux groupes correspondant à deux canettes fortement occultées (en bleu foncé) ne représentent chacun que 1% du total des correspondances.

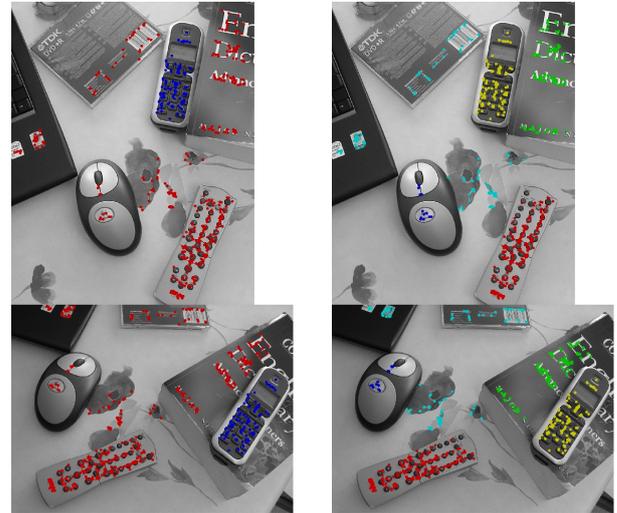
Dans le dernier exemple sont utilisées deux photographies contenant certains objets en commun dans des poses différentes (figure 6(a)) : une boîte de céréales, ainsi que trois canettes identiques. Rien ne distingue les trois canettes, si bien que 9 transformations sont théoriquement possibles entre les deux vues pour ces 3 objets. En utilisant MAC-RANSAC avec une transformation plane (similitude, transformation affine ou projective), un unique groupe correspondant à la boîte de céréales est correctement détecté, ainsi que 9 groupes de correspondances entre les canettes (figure 6(b)).

## 6 Conclusion

Nous avons présenté un algorithme robuste de groupement multiple de mises en correspondances de points d'intérêt pour la reconnaissance d'objets. Cette approche utilisant le cadre méthodologique *a contrario* ne requiert le réglage d'aucun paramètre, hormis que le nombre d'itérations  $i_{max}$  qui est généralement fixé selon des considérations computationnelles.



(a) Paire d'images utilisée



(b) Groupement sous contrainte épipolaire (c) Groupement sous contrainte projective

FIG. 4 – Application de MAC-RANSAC pour une scène non-statique, avec un mouvement de la caméra.

Nous avons proposé une généralisation aux transformations planes du cadre théorique introduit dans [8] pour la géométrie épipolaire, et dont l'intérêt a été démontré pour la reconnaissance d'objets dans la partie expérimentale. Un autre aspect intéressant de cette approche, mais non présenté ici, est la sélection de modèles géométriques pour l'estimation de la pose de l'objet.

## Références

- [1] F. Cao, J. Delon, A. Desolneux, P. Musé, and F. Sur. A unified framework for detecting groups and application to shape recognition. *J. Math. Imaging Vision*, 27(2), 2007.
- [2] A. Desolneux, L. Moisan, and J.-M. Morel. Meaningful alignments. *Int. J. Comput. Vision*, 40(1) :7–23, 2000.
- [3] A. Desolneux, L. Moisan, and J.-M. Morel. *From Gestalt Theory to Image Analysis : A Probabilistic Approach*. Springer Verlag, 2008.
- [4] 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.
- [5] P. Hough. Machine analysis of bubble chamber pictures. In *International Conference on High Energy Accelerators and Instrumentation*, pages 554–556, 1959.
- [6] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2) :91–110, 2004.



(a) L'objet recherché à gauche est présent de nombreuses fois dans la scène de droite.



(b) Résultat du groupement avec MAC-RANSAC : les 28 occurrences de la cible sont détectées.

FIG. 5 – Application de MAC-RANSAC pour la reconnaissance d'occurrences multiples d'un même objet.

- [7] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, pages 384–393, 2002.
- [8] 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.
- [9] P. Monasse. *Représentation morphologique d'images numériques et application au recalage*. PhD thesis, CMLA, ENS de Cahan, 2000.
- [10] N. Noury, F. Sur, and M.-O. Berger. Fundamental matrix estimation without prior match. In *ICIP*, 2007.
- [11] N. Noury, F. Sur, and M.-O. Berger. Modèles statistiques pour l'estimation de la matrice fondamentale. In *ORASIS*, 2007.
- [12] M. Pollefeys. Séquence vidéo *Leuven castel*, disponible à : [www.cs.unc.edu/~marc/data/castlejpg.zip](http://www.cs.unc.edu/~marc/data/castlejpg.zip).
- [13] 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.
- [14] C. V. Stewart. Minpran : A new robust estimator for computer vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(10) :925–938, 1995.
- [15] C. V. Stewart. Bias in robust estimation caused by discontinuities and multiple structures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(8) :818–833, 1997.
- [16] R. Toldo and A. Fusiello. Robust multiple structures estimation with j-linkage. In *ECCV (1)*, pages 537–547, 2008.



(a) Deux scènes différentes partageant des objets en commun.



(b) L'algorithme MAC-RANSAC détecte les 10 groupes recherchés.

FIG. 6 – Reconnaissance de plusieurs objets dans un contexte différent.

- [17] P. H. S. Torr and D. W. Murray. Stochastic motion clustering. In *ECCV (2)*, pages 328–337, 1994.
- [18] 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.
- [19] E. Vincent and R. Laganriere. Detecting planar homographies in an image pair. *Image and Signal Processing and Analysis*, pages 182–187, 2001.
- [20] W. Zhang and J. Kosecka. Generalized ransac framework for relaxed correspondence problems. In *Proceedings of 3DPVT'06*.
- [21] W. Zhang and J. Kosecká. Nonparametric estimation of multiple structures with outliers. In *WDV*, pages 60–74, 2006.
- [22] M. Zuliani, C. S. Kenney, and B. S. Manjunath. The multi-ransac algorithm and its application to detect planar homographies. In *ICIP*, 2005.