



# Exploitation de vidéos aériennes multiples pour la détection de changements

---

Nicolas Bourdis  
Hichem Sahbi  
Denis Marraud

**2011D007**

septembre 2011

Département Traitement du Signal et des Images  
Groupe TII : Traitement et Interprétation des Images

# Exploitation de vidéos aériennes multiples pour la détection de changements.

**Nicolas BOURDIS**

EADS France, Innovation Works Dpt, Télécom ParisTech, Suresnes, France  
(nicolas.bourdis@eads.net)

**Hichem SAHBI**

CNRS LTCI UMR5141, Télécom ParisTech, Paris, France  
(hichem.sahbi@telecom-paristech.fr)

**Denis MARRAUD**

EADS France, Innovation Works Dpt, Suresnes, France  
(denis.marraud@eads.net)

Ce rapport présente les travaux réalisés dans le cadre d'une thèse CIFRE encadrée conjointement par Hichem SAHBI et Denis MARRAUD.

## Table des matières

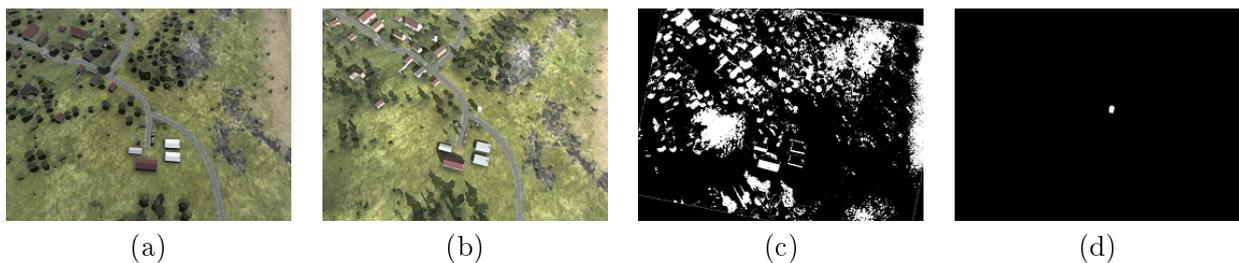
<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Aperçu de l'État de l'Art</b>	<b>4</b>
2.1	Pré-Traitement des Données . . . . .	4
2.2	Méthodes de Détection de Changements . . . . .	5
2.3	Motivations et Contributions . . . . .	6
<b>3</b>	<b>Données et Évaluations</b>	<b>6</b>
3.1	Base de Données Synthétiques . . . . .	6
3.2	Critères d'Évaluation . . . . .	7
<b>4</b>	<b>Détection de Changements entre Paires d'Images</b>	<b>8</b>
4.1	Flot Optique Brut . . . . .	8
4.2	Contrainte Épipoilaire . . . . .	9
4.3	Atténuation des Ombres . . . . .	10
<b>5</b>	<b>Exploitation de Vidéos pour la Détection de Changements</b>	<b>12</b>
5.1	Constitution de la Base de Référence . . . . .	12
5.2	Requête et Détection de Changements . . . . .	12
5.3	Exploitation de la Redondance Temporelle . . . . .	13
<b>6</b>	<b>Conclusion</b>	<b>15</b>
	<b>Références</b>	<b>16</b>

# 1 Introduction

Nous avons assisté ces dernières années à une multiplication des déploiements et de l'utilisation de systèmes d'acquisition vidéo, notamment pour la vidéo surveillance ou pour l'observation aérienne. Dans le cas de ces deux applications, le volume de données vidéos acquises est aujourd'hui tel qu'il est impossible aux opérateurs de les analyser exhaustivement. Afin de ne pas perdre de données potentiellement intéressantes, il devient donc nécessaire de recourir à une pré-analyse automatique pour filtrer rapidement les données inutilisables. Dans ce contexte, et dans le cadre de l'observation aérienne, cette thèse vise à étudier l'intérêt de la détection automatique de changements entre plusieurs données vidéos pour la focalisation de l'attention des opérateurs. On s'intéressera notamment au cas où un ensemble de vidéos de références sont disponibles et grâce auquel la détection de changements dans une vidéo de test peut-être effectuée.

La détection automatique de changements est un problème de vision artificielle ayant pour objectif l'estimation, à partir d'une *donnée de référence* et d'une nouvelle donnée communément qualifiée de *donnée de test*, du *masque de changements*, c'est-à-dire de l'étiquetage binaire (changement et non-changement) des pixels d'une des deux données. La définition de la pertinence d'un changement dépend beaucoup de l'application visée, et dans le contexte de l'observation aérienne, l'intérêt est fréquemment porté sur les déplacements, déformations, apparitions ou disparitions d'objets divers (e.g. bâtiments, voitures, piétons, ...). En revanche, il est généralement admis que les effets d'illumination (e.g. ombres, intensité lumineuse variable, ...) et les effets de point de vue (e.g. parallaxe, occultations, ...) sont des changements non-pertinents, tout comme les variabilités de capteurs (e.g. artefacts d'acquisition, gains, ...) ou d'apparence de la scène (e.g. réflexions spéculaires, mouvements d'eau, d'arbres, ...). Or, la présence de ces changements non-pertinents rendent difficile l'identification automatique des changements pertinents, comme le montre la figure 1.

**FIGURE 1** – Cette figure montre que deux images aériennes, de référence (a) et de test (b), en apparence similaires contiennent en réalité de nombreuses différences insignifiantes, mais gênant l'identification automatique des changements pertinents. Le résultat d'une détection de changements simpliste, consistant à recalculer les images par homographie puis à les soustraire, est présenté en (c). La vérité terrain est présentée en (d).



Plus généralement, lorsque les données considérées sont des images, on rencontre deux types de difficultés : le problème de la mise en correspondance, consistant à apparier les zones comparables des deux images, et celui de la détection de changements à proprement parler, qui consiste à comparer les images afin d'identifier les zones différentes. Nous verrons dans la section 2 que de nombreuses approches ont été proposées pour résoudre ces problèmes. En revanche, peu de travaux abordent le cas des données vidéos, et la majorité prennent des hypothèses simplificatrices vis-à-vis des problèmes spécifiques soulevés. Pour commencer, la détection de changements en temps réel dans une vidéo nécessite un algorithme rapide et fiable, ce qui requiert un effort d'optimisation séquentielle et influence les choix d'implémentation (pour la parallélisation éventuelle des calculs). Par ailleurs, le format de donnée vidéo étant intrinsèquement redondant, il semble évident que l'exploitation de cette redondance devrait être bénéfique, permettant par exemple l'amélioration des performances ou le soulagement des capacités de calcul. De plus, l'utilisation de données vidéo pose la question du choix de la donnée de référence à comparer avec une donnée de test fixée. Cette question n'est qu'implicitement abordée lorsque l'on considère des paires d'images, car on suppose généralement que les deux images sont

comparables a priori. Dans le cas de vidéos, étant donné une image de la vidéo de test, il n'est pas trivial de choisir l'image de la vidéo de référence avec laquelle comparer la première.

Ce rapport de mi-parcours va tout d'abord présenter, section 2, une revue de l'état de l'art dans le domaine de la détection de changements, en décrivant un certain nombre de méthodes intéressantes et leurs limitations. Les contributions de cette première moitié de thèse seront résumées à la fin de cette section. Le section 3 discutera des problématiques liées aux évaluations, et présentera la base de données de synthèse établie et utilisée pour évaluer les algorithmes développés. Les travaux réalisés pour la détection de changements, respectivement entre images et entre vidéos, seront présentés dans les sections 4 et 5. Enfin, un bilan des travaux réalisés sera établi en conclusion, qui mentionnera également les perspectives d'amélioration et les problèmes restant à résoudre.

## 2 Aperçu de l'État de l'Art

Le domaine de la détection de changements pertinents dispose d'un remarquable article de revue [20] des méthodes publiées avant 2005. Cet article qui catégorise les approches en plusieurs catégories, insiste sur l'intérêt que peut avoir un certain nombre d'étapes de pré-traitements, en particulier concernant les ajustements géométriques et radiométriques. De la même façon, nous commencerons par présenter diverses techniques de pré-traitement, et nous discuterons ensuite des méthodes de détection de changement.

### 2.1 Pré-Traitement des Données

L'étape de pré-traitements a pour but d'atténuer le plus possible la variabilité due aux changements non-pertinents avant de procéder à la détection des changements pertinents. Cette section décrit quelques techniques permettant d'atténuer les changements dus aux modifications de point de vue et d'illumination.

Lorsqu'il y a modification du point de vue entre la donnée de test et celle de référence, les changements résultants sont généralement compensés grâce à un recalage préliminaire des données. Le recalage d'images est une problématique classique en vision artificielle [25], et se fait généralement par une transformation rigide, le plus souvent affine ou projective, ou non rigide [7, 9]. En revanche, le recalage d'une image, issue d'une séquence, avec un ensemble de vidéos est plus rarement abordé. Une hypothèse fréquemment employée est de restreindre l'application au cas d'une vidéo de test et d'une vidéo de référence, toutes deux acquises selon des trajectoires proches [19, 22]. Dans ce cas simplifié mais correspondant à certains scénarios opérationnels, le recalage peut-être réalisé grâce à la synchronisation temporelle des vidéos suivie d'un recalage simple des images synchronisées.

Par ailleurs, les variations d'illumination, et en particulier les ombres, doivent être atténuées avant de pouvoir procéder à la comparaison d'images. Pour cela, on citera deux approches intéressantes, permettant de s'abstraire de l'illumination d'une scène et d'obtenir une image approximativement invariante par modification d'illumination. Sous l'hypothèse que la scène est constituée de surfaces Lambertiennes (i.e. réfléchissant la lumière de manière diffuse), [23] modélise les intensités dans l'image comme le produit d'une composante décrivant l'illumination et d'une composante décrivant les propriétés de la scène (réflectance), qui sont mutuellement indépendantes. Une hypothèse supplémentaire, selon laquelle la variation spatiale de la première composante est faible par rapport à celle de la seconde, permet de s'abstraire de l'illumination grâce à un filtrage passe-haut (dit homomorphique). Cependant, cette dernière hypothèse est très restrictive, car elle ne permet pas d'aborder correctement le problème des ombres. Pour cela, sous l'hypothèse d'un éclairage modélisable par la théorie des corps noir, [12] propose une approche plus générale permettant d'atténuer fortement toute variabilité due à l'illumination et notamment les ombres. Ces travaux montrent que l'intensité couleur  $(R; G; B)$  d'un point de la scène, projetées dans l'espace 2D des chromaticités  $(R/G; B/G)$  varie linéairement avec la température de corps noir de la source de lumière. Cela signifie qu'une projection adaptée permet de s'abstraire du type et de l'intensité d'illumination, et d'obtenir une image en niveau de gris invariante par variation d'éclairage.

## 2.2 Méthodes de Détection de Changements

Comme mentionné en introduction, le problème de la détection de changement peut-être décomposé en deux sous-problèmes : le problème de mise en correspondance et celui de la comparaison des images. Les différentes approches rencontrées sont discutées ci-dessous.

Dans un certain nombre de cas, l'effort des auteurs porte sur la détection de changements à proprement parler, en supposant que les données sont recalées préalablement [8, 17]. En conséquence, les méthodes de détection de changements ne sont souvent pas conçues pour être robustes aux erreurs de recalage, et elles ne sont donc pas adaptées aux applications impliquant des modifications de points de vue et des scènes contenant du relief.

Une autre approche plus récente consiste à reconstruire un modèle 3D de la scène grâce à la vidéo de référence, puis à générer une image de ce modèle selon le point de vue de l'image de test. Suivant cette approche, [18, 10] génèrent un modèle volumétrique de la scène, grâce à une structure hiérarchique de voxels contenant une information d'occlusion et d'apparence. Du fait de la lourdeur du modèle 3D, cette approche présente un problème de passage à l'échelle, car elle est adaptée à des scènes restreintes mais inapplicable dans le cas de scènes étendues. [5] propose une autre approche plus légère, mais plus sensible aux erreurs de reconstruction, basée sur un algorithme de "Structure From Motion". Cependant, dans les deux cas, la reconstruction 3D nécessite une bonne exploration de la scène dans la vidéo de référence, ce qui peut constituer une contrainte importante sur la trajectoire des engins aériens.

Enfin, une approche intermédiaire entre les deux précédentes, prenant en compte le relief et la possibilité de modification du point de vue mais sans reconstruire de modèle 3D, consiste à exploiter la géométrie épipolaire [14] pour la recherche des correspondances entre images. [24] implémente cette approche en cherchant un appariement des formes de toits de bâtiments dans des images aériennes avec axe optique proche de la verticale (vue nadir). Cette étude propose également d'utiliser le ratio des intensités dans les images pour distinguer les changements réels des variations d'ombres. [6] exploite également la géométrie épipolaire pour la recherche de correspondance. Un point remarquable de cette étude est l'utilisation d'une technique de programmation dynamique pour rendre la méthode robuste aux occlusions de faces d'objets, qui ne sont généralement pas abordées dans les autres méthodes. Cependant dans les deux cas, les temps de calcul sont prohibitifs pour le traitement temps réel de vidéos.

De nombreuses approches permettant la comparaison de deux images ont été proposées [20]. Une première catégorie d'approches consiste à comparer en chaque pixel un seuil avec un critère très simple combinant les deux images. Les premières techniques utilisaient simplement la différence des intensités en chaque pixel. Un certain nombre de travaux ont étudié la manière de déterminer rigoureusement les valeurs de seuil, notamment via des tests statistiques [23]. Diverses variantes ou extensions ont été développées pour améliorer ce type d'approche. Par exemple, pour renforcer la cohérence spatiale du masque de changements, certains auteurs ont proposé une modélisation impliquant un champ de Markov aléatoire [17]. Un autre exemple de variante est la transformation des intensités des images avant la comparaison, comme l'utilisation des réponses des opérateurs de Sobel [19]. Par ailleurs, plutôt que la différence des intensités, certains auteurs [24] proposent d'utiliser le ratio d'intensité, qui permet en outre de filtrer certains effets d'illumination.

Une seconde catégorie d'approches consiste à analyser l'adéquation de la donnée de test avec un modèle obtenu de diverses manières. [3] décrit une méthode permettant d'estimer, grâce à un algorithme Expectation Maximization (EM), les paramètres d'un mélange de modèles a priori décrivant les sources de variabilités dans une donnée de référence (e.g. mouvement, illumination, réflexions spéculaires, ...). Les changements peuvent alors être identifiés comme les zones de la donnée de test où les intensités ne sont pas explicables par ce mélange de modèles. [8] suggère que l'analyse de l'ensemble des intensités d'une paire d'images peut suffire pour construire un modèle de manière non supervisée, et propose d'entraîner un réseau de neurone à estimer l'évolution d'intensité de l'une vers l'autre. Les changements sont alors identifiés comme les zones excessivement différentes de la prédiction. Dans le cas de caméras fixes, [21] propose une méthode permettant de construire un modèle d'apparence en chaque pixel à

partir d'une séquence de données de référence. Un pixel de la donnée de test est alors considéré comme changement lorsque que son intensité n'est pas explicable par ce modèle. Deux thèses [18, 10] ont étendu cette méthode au cas de caméras mobiles, par la construction d'un modèle volumétrique de la scène.

### 2.3 Motivations et Contributions

Nous l'avons vu, l'état de l'art en détection de changements est très abondant dans le cas de paires d'images, mais nettement moins dans le cas de séquences vidéos, et c'est ce contexte d'application qui motivera les travaux de cette thèse.

Dans le cas de paires d'images, le problème de la détection de changements est vu comme un cas particulier du problème de mise en correspondance, dans lequel l'objectif est d'identifier les zones sans correspondance. Ce cadre permet de résoudre les deux sous-problèmes simultanément, comme peu d'approches l'envisagent [3], et de manière rapide et flexible. Grâce à l'utilisation d'une technique d'atténuation d'ombres et de la géométrie épipolaire, un effort particulier est porté sur la gestion adéquate des effets d'illumination et de point de vue. Enfin, une base d'évaluation constituée d'images aériennes synthétiques a été créée afin de permettre l'évaluation d'algorithmes de détection de changements. Ces contributions ont donné lieu à la publication d'un article à la conférence IGARSS 2011 [4].

D'autre part, nous étudions un système se basant sur l'algorithme précédent, et permettant le traitement efficace de vidéos de test via l'exploitation des données de référence disponibles. Une indexation spatiale permet de répondre au problème du choix de la donnée de référence optimale, et la redondance des données vidéos est exploitée pour améliorer les résultats.

## 3 Données et Évaluations

L'obtention de données réelles permettant l'évaluation d'algorithmes de détection de changements est délicate pour plusieurs raisons. D'une part, organiser une campagne d'acquisition dédiée demande à la fois d'importants moyens aériens pour l'acquisition et une lourde logistique au sol pour la mise en place de changements statistiquement pertinents. D'autre part, dans ce domaine, les données obtenues sur terrain opérationnel sont pour la plupart confidentielles. Pour finir, l'évaluation d'algorithmes nécessite la connaissance de la vérité-terrain, dont l'obtention constitue un travail d'expert qui peut varier selon les personnes [20].

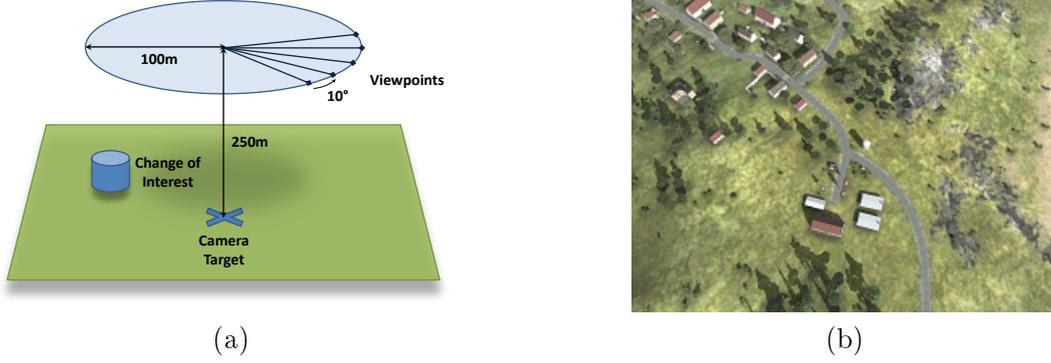
Pour ces raisons, nous avons choisi d'utiliser dans un premier temps un ensemble de données synthétiques. La possibilité d'acquisition de données réelles est cependant activement étudiée, et permettra une analyse plus pertinente des algorithmes développés d'ici la fin de la thèse.

### 3.1 Base de Données Synthétiques

La base de données synthétique a été générée grâce au moteur de rendu du "serious game" Virtual Battle Station 2, développé par Bohemia Interactive Simulations, dont le réalisme et la diversité des modèles permet de se rapprocher au mieux de données réelles, tout en bénéficiant de fonctionnalités avancées telles que l'extraction objective de vérité-terrain.

Cette base d'évaluation contient 1000 paires d'images, de résolution  $800 \times 600$  pixels, issues de 100 scènes distinctes, vues depuis 5 points de vue différents et générées avec ou sans ombres. Comme l'illustre le schéma de la figure 2, les différents points de vue sont écartés d'un angle de 10 degrés autour d'un cercle de rayon 100 mètres et à une altitude de 250 mètres. Cette distribution des points de vue permet l'évaluation de l'influence de l'écart entre points de vue sur les performances. D'autre part, la possibilité d'utilisation des images avec ou sans ombres portées permet d'analyser les performances en détection de changements indépendamment d'une éventuelle technique d'atténuation d'ombres (N.B. dans la version sans ombre portée, il reste malgré tout des variations d'illumination).

**FIGURE 2** – Cette figure illustre (a) le mode d’acquisition des images de la base d’évaluation et (b) le réalisme des images générées par le moteur de rendu utilisé.



### 3.2 Critères d’Évaluation

L’évaluation d’un masque de changements estimé par rapport au masque de changements réel correspondant peut être réalisée de diverses manières selon les applications. Dans le contexte de l’imagerie satellite (e.g. pour la monitoring de ressources naturelles), il peut être important de quantifier avec précision les changements survenus, et au vu de la résolution au sol, il peut être intéressant de comparer les deux masques en chaque pixel. En revanche, dans un contexte de focalisation d’attention en vidéo surveillance, il peut être plus intéressant de comparer ces masques après une décision sur l’image globale, afin de minimiser la fréquence d’erreur. Dans un contexte de focalisation d’attention en imagerie aérienne, le critère image semble manquer de localisation tandis que le critère pixelique peut paraître trop strict. Un nouveau critère, faisant un compromis entre les deux critères mentionnés, a donc été défini pour plus de flexibilité.

Étant donné une paire d’images  $I_r$  et  $I_t$ , on dénote le masque de changements réel par  $Mr_I$  et le masque de changements estimé par  $Me_I$ . Les pixels non-nuls de ces deux masques peuvent être séparés en composantes connexes, notées  $\{r_I^i\}_i$  pour le masque réel et  $\{e_I^j\}_j$  pour le masque estimé. Dénotons enfin le nombre de pixels dans une composante connexe  $c$  par  $|c|$ . Considérons maintenant les définitions des fonctions (1) à (4), qui dépendent du paramètre  $\alpha \in [1; +\infty]$  relaxant la contrainte de localisation (en pratique, nous utilisons  $\alpha = 5$ ).

$$tp_r : \begin{cases} \{r_I^i\}_i & \longrightarrow \mathbb{N}^+ \\ r & \longmapsto \begin{cases} |r| & \text{si } \frac{|r \cap \bigcup_j e_I^j|}{|r|} \geq \frac{1}{\alpha} \\ \alpha \cdot |r \cap \bigcup_j e_I^j| & \text{sinon} \end{cases} \end{cases} \quad (1)$$

$$fn_r : \begin{cases} \{r_I^i\}_i & \longrightarrow \mathbb{N}^+ \\ r & \longmapsto \begin{cases} 0 & \text{si } \frac{|r \cap \bigcup_j e_I^j|}{|r|} \geq \frac{1}{\alpha} \\ |r| - \alpha \cdot |r \cap \bigcup_j e_I^j| & \text{sinon} \end{cases} \end{cases} \quad (2)$$

$$tp_e : \begin{cases} \{e_I^j\}_j & \longrightarrow \mathbb{N}^+ \\ e & \longmapsto \begin{cases} |e| - |e \cap \bigcup_i r_I^i| & \text{si } \frac{|e \cap \bigcup_i r_I^i|}{|e|} \geq \frac{1}{\alpha} \\ (\alpha - 1) \cdot |e \cap \bigcup_i r_I^i| & \text{sinon} \end{cases} \end{cases} \quad (3)$$

$$fp_e : \begin{cases} \{e_I^j\}_j & \longrightarrow \mathbb{N}^+ \\ e & \longmapsto \begin{cases} 0 & \text{si } \frac{|e \cap \bigcup_i r_I^i|}{|e|} \geq \frac{1}{\alpha} \\ |e| - \alpha \cdot |e \cap \bigcup_i r_I^i| & \text{sinon} \end{cases} \end{cases} \quad (4)$$

On peut noter la présence du terme  $-|e \cap \bigcup_i r_I^i|$  dans la fonction  $tp_e$ , qui permet de ne pas compter deux fois, dans  $tp_e$  et dans  $tp_r$ , la contribution de l’intersection. L’évaluation du masque estimé par

rappport au masque réel se fait alors en comptant comme suit les vrais positifs (TP), faux négatifs (FN), faux positifs (FP) et vrais négatifs (TN) :

- $TP_I = \sum_i tp_r(r_I^i) + \sum_j tp_e(e_I^j)$
- $FN_I = \sum_i fn_r(r_I^i)$
- $FP_I = \sum_j fp_e(e_I^j)$
- $TN_I = |I_t| - (TP_I + FN_I + FP_I)$

Les définitions des fonctions (1) à (4) permettent de déduire que  $\forall r \in \{r_I^i\}_i$ ,  $tp_r(r) + fn_r(r) = |r|$  et que  $\forall e \in \{e_I^j\}_j$ ,  $tp_e(e) + fn_e(e) = |e| - |e \cap \bigcup_i r_I^i|$ . Comme par ailleurs, les  $r_I^i$  (respectivement, les  $e_I^j$ ) sont deux à deux d'intersection vide, on peut en déduire que  $\forall I, TP_I + FN_I + FP_I = |\bigcup_i r_I^i \cup \bigcup_j e_I^j|$ . Nous déduisons alors le nombre de vrais négatifs en considérant le complémentaire de l'ensemble  $\bigcup_i r_I^i \cup \bigcup_j e_I^j$  dans l'image  $I_t$ . On obtient alors l'expression de  $TN_i$  ci-dessus.

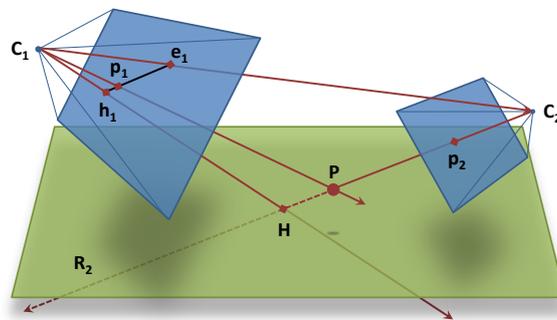
## 4 Détection de Changements entre Paires d'Images

Cette section décrit les travaux réalisés sur la détection de changements entre paires d'images. L'algorithme de détection de changements actuel, c'est-à-dire prenant en compte les diverses étapes présentées ci-dessous, traite une image  $800 \times 600$  en 4 secondes, ce qui permet de penser qu'une accélération GPU permettrait d'atteindre un framerate d'environ 2 frames par seconde.

### 4.1 Flot Optique Brut

La figure 1 présentée en introduction montre bien que l'utilisation de la différence d'images ne permet pas fondamentalement de différencier les vrais changements des diverses variations non-pertinentes, et en particulier des effets de point de vue. Pour faire la distinction, nous nous sommes orienté vers l'utilisation du champ vectoriel résiduel<sup>1</sup> de parallaxe après recalage, dont il a été démontré dans [15] qu'il suit l'orientation du champ épipolaire (voir illustration de la figure 3). En effet, considérons deux caméras de centres optiques  $C_1$  et  $C_2$  observant un point  $P$  dans la scène, ainsi qu'une surface arbitraire  $S$ , ne contenant pas  $P$ , par rapport à laquelle les deux images sont recalées (typiquement, un plan). Le rayon lumineux  $R_2$  joignant les points  $C_2$  et  $P$  intersecte la surface  $S$  au point  $H$ . La projection du rayon  $R_2$  dans le plan image de  $C_1$  reste une ligne joignant les projections de  $C_2$ ,  $P$  et  $H$ , respectivement notées  $e_1$ ,  $p_1$  et  $h_1$ . Puisque  $e_1$  est l'épipoles dans l'image 1, on en conclut que le vecteur  $\overrightarrow{h_1 p_1}$ , c'est-à-dire le vecteur résiduel de parallaxe associé à  $P$  après recalage, est orienté selon la droite épipolaire.

**FIGURE 3** – Cette figure illustre le fait que les vecteurs résiduels de parallaxe, après recalage, sont orientés selon les droites épipolaires.

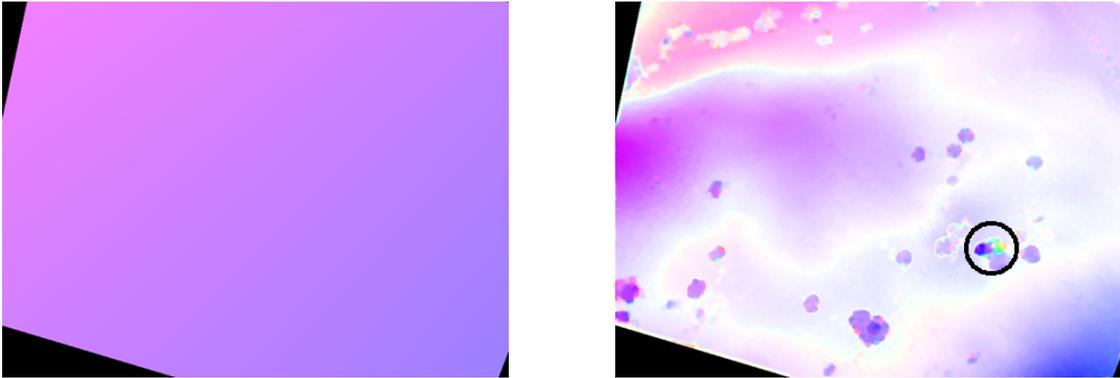


En considérant l'orientation des vecteurs résiduels, qui est épipolaire dans le cas des effets de parallaxe et arbitraire dans le cas des vrais changements, il est alors possible de distinguer les premiers des seconds. L'estimation pratique de ces vecteurs résiduels est un problème classique de mise

1. Le champ vectoriel résiduel représente le déplacement résiduel nécessaire pour obtenir un recalage parfait après un recalage préalable. Ce déplacement résiduel peut être dû aux effets de parallaxe, aux mouvements dans la scène, ...

en correspondance d’images, consistant à raffiner un recalage grossier obtenu préalablement. Plusieurs techniques existantes visent à résoudre ce problème, notamment les techniques de flot optique ou d’appariement stéréo. Ces dernières techniques exigent souvent une rectification des images, consistant à les transformer de manière à rendre les droites épipolaires parallèles, ce qui peut poser problème lorsque l’épipole est situé à l’intérieur de l’image. Nous avons donc choisi d’utiliser une technique de flot optique pour estimer les vecteurs résiduels, et plus particulièrement la technique introduite dans [11] qui présente de nombreux avantages, dont la disponibilité du code source dans OpenCV, le caractère dense du flot optique estimé, et la pertinence et l’efficacité du critère de comparaison employé. La figure 4 illustre ce procédé de détection de changements, en comparant les visualisations du champ épipolaire et du champ de vecteurs résiduels. Dans le premier cas, les vecteurs sont de norme constante, dans le second, leur norme est proportionnelle à la distance du point par rapport au plan dominant de la scène, permettant de visualiser le relief (collines) ainsi que les objets (arbres, batiments ...).

**FIGURE 4** – Cette figure présente une visualisation du champ épipolaire (a) et du champ vectoriel résultant du flot optique (b), selon un encodage des vecteurs dans l’espace de couleurs Teinte-Saturation-Valeur (TSV) : la Teinte représente l’orientation du vecteur (entre 0 et 360 degrés) et la Saturation représente la norme du vecteur (normalisée par la norme maximale dans l’image). La Valeur n’est pas utilisée est est constante égale à 255. La figure montre que les Teintes sont très similaires entre les deux images, excepté autour du changement (cercle) et de quelques fausses alarmes.



La méthode de détection de changements envisagée ici consiste donc en une succession d’étapes. Pour commencer, des primitives SURF [1] sont extraites de chaque image et appariées selon la distance entre leurs descripteurs. Ces appariements sont ensuite utilisés pour l’estimation de la matrice fondamentale, par une approche de type Random Sample Consensus (RANSAC) [13]. Ces appariements sont également exploités pour estimer une homographie de recalage préliminaire, permettant de faciliter l’estimation des vecteurs résiduels. Après le calcul du flot optique, le masque de changements est finalement estimé à partir de la comparaison entre l’orientation du vecteur résiduel en chaque pixel et de la direction de l’épipole, selon le critère défini comme suit. Considérant un pixel  $\mathbf{x}$  du masque de changements, soit  $\vec{v}(\mathbf{x})$  le vecteur résiduel estimé par flot optique et  $\vec{e}(\mathbf{x})$  la direction (unitaire) de l’épipole au pixel  $\mathbf{x}$ . On décide alors qu’un changement est survenu en  $\mathbf{x}$  si l’orientation du vecteur résiduel est trop différente de la direction de l’épipole, c’est-à-dire si  $\epsilon(\mathbf{x}) = \left\| \vec{v}(\mathbf{x}) - \left( \vec{e}(\mathbf{x}) \cdot \vec{v}(\mathbf{x}) \right) \vec{e}(\mathbf{x}) \right\| \geq \tau$ , où  $\tau$  est un seuil. Ce seuil peut être déterminé a priori par une formulation sous forme de test statistique (e.g. un ratio de vraisemblances). La courbe Précision/Rappel obtenue en faisant varier ce seuil est présentée à la figure 5 (N.B. cette méthode correspond à la légende “Unconstrained Optical Flow”). Les performances associées ne sont pas satisfaisantes, dû à un grand nombre de fausses alarmes que la méthode présentée dans la section suivante permet de réduire.

## 4.2 Contrainte Épipoilaire

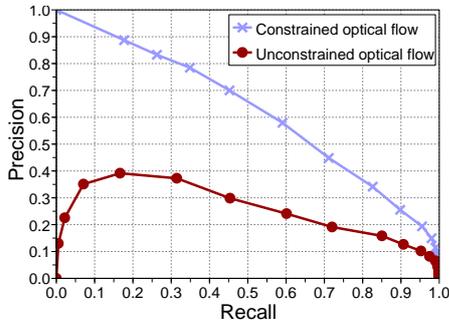
Comme mentionné en commentaire de la figure 4, un certain nombre de fausses alarmes apparaissent autour des objets sujets à la parallaxe. Cela peut s’expliquer par le fait que l’algorithme de flot

optique cherche à maximiser la similarité de patches locaux, or le patch de similarité maximale n'est pas forcément celui correspondant réellement à la position recherchée.

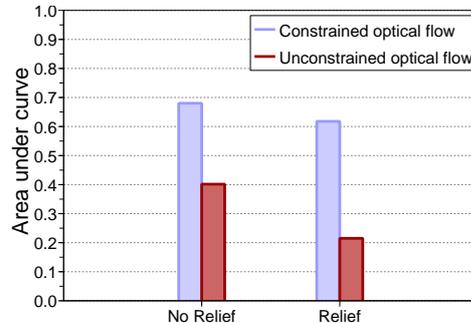
Pour contourner ce problème, et puisque nous nous intéressons principalement ici à être robuste aux effets de parallaxe, une solution naturelle consiste à contraindre le flot optique à faire la recherche directement le long de la droite épipolaire. Dans ce cas, le critère de décision de la méthode précédente n'est plus adapté. Par simplicité, nous choisissons donc d'utiliser comme nouveau critère l'erreur d'appariement local résultant du flot optique comparée à un seuil. Il est à noter qu'une telle approche exploitant une contrainte épipolaire n'est pas pertinente lorsque les vecteurs résiduels ne sont pas dus à la parallaxe, mais par exemple au déplacement d'objets. Cela aura pour conséquence que les déplacements d'objets seront détectés comme changements, ce qui est acceptable dans les applications visées.

La figure 5 compare les performances obtenues avec cette méthode, avec contrainte épipolaire, à celles de la méthode précédente, sans contrainte épipolaire, pour des points de vue différenciés de 10 degrés. La convergence de la seconde courbe vers le coin inférieur gauche, qui est inhabituelle, signifie que pour les plus hautes valeurs de seuil les détections correspondent exclusivement à des fausses alarmes, la Précision tendant alors vers zéro. Outre la meilleure gestion des fausses alarmes, l'utilisation de la contrainte épipolaire permet d'orienter le flot optique vers la solution souhaitée, rendant la méthode plus robuste aux cas difficiles, tels que les scènes contenant beaucoup de relief ou les paires d'images avec fort écartement des points de vue, comme le montrent les figures 6 et 7. La figure 6 compare les aires sous la courbe Précision/Rappel sur une partition binaire de la base de données en scènes contenant ou non du relief, et avec ou sans contrainte épipolaire. Nous pouvons voir qu'avec présence de relief, les performances de la méthode sans contrainte chutent d'environ 50%, alors qu'elles diminuent de moins de 15% pour la méthode contrainte. Cela démontre une meilleure robustesse au relief, et plus généralement, à un mauvais recalage préalable des images.

**FIGURE 5** – Cette figure compare les performances obtenues sur la base d'évaluation par la méthode avec et sans contrainte épipolaire.



**FIGURE 6** – Cette figure évalue l'influence de la présence de relief, dans le cas de la méthode avec et sans contrainte.

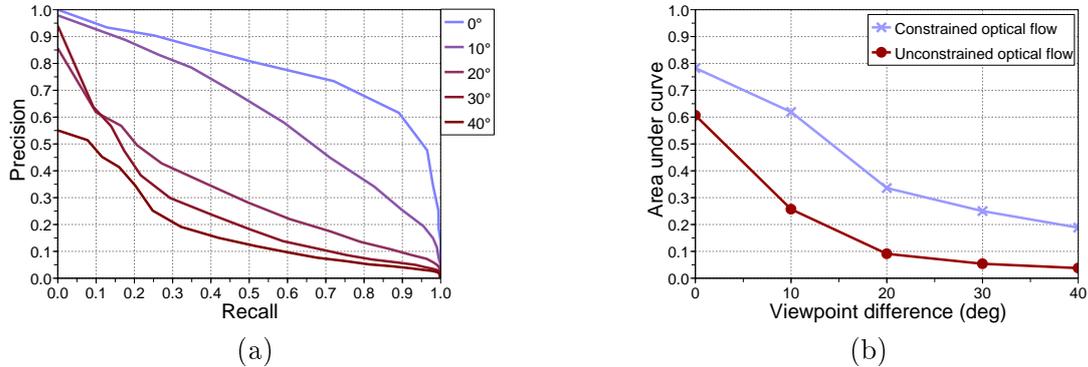


Par ailleurs, la figure 7 compare l'évolution de l'aire sous la courbe Précision/Rappel en fonction de l'écart des points de vue (on retrouve la courbe obtenue pour un écart de 10 degrés). Si l'on compare l'évolution entre 0 et 10 degrés, on remarque que la encore les performances de la méthode sans contrainte chute de plus de 50%, alors qu'elles diminuent de moins de 25% dans le cas contraint. On peut remarquer qu'au delà d'un écart de 20 degrés entre les points de vue, l'évolution des courbes est similaire avec ou sans contrainte. Cela peut être expliqué par le fait qu'il est plus probable que des facettes occultées (e.g. de bâtiments) apparaissent lorsque les écarts de points de vue sont grands, or dans les deux cas ces apparitions sont faussement interprétées comme des changements.

### 4.3 Atténuation des Ombres

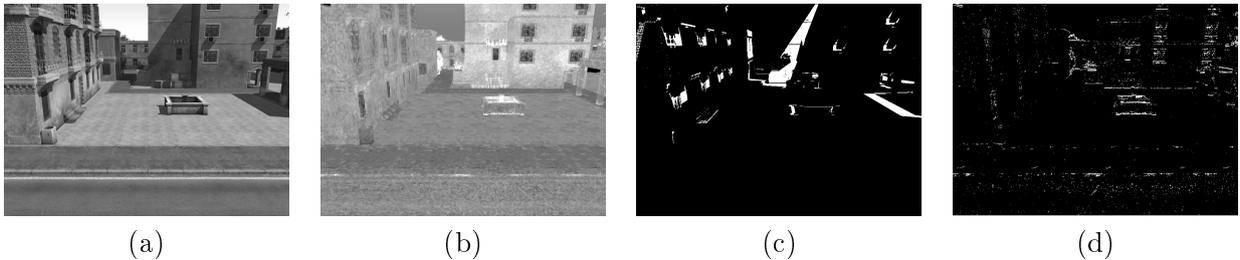
Les évaluations effectuées jusqu'ici étaient réalisées sur la base d'images sans ombres portées, mais même ainsi et bien que la technique de flot optique utilisée soit invariante par transformation affine des intensités entre paires d'images, un certain nombre des fausses alarmes produites par l'algorithme

**FIGURE 7** – Cette figure évalue l’influence de l’écart des points de vue sur les performances de la méthode avec contrainte (a), et compare les cas des méthodes avec et sans contrainte du point de vue de l’aire sous la courbe Précision/Rappel (b).



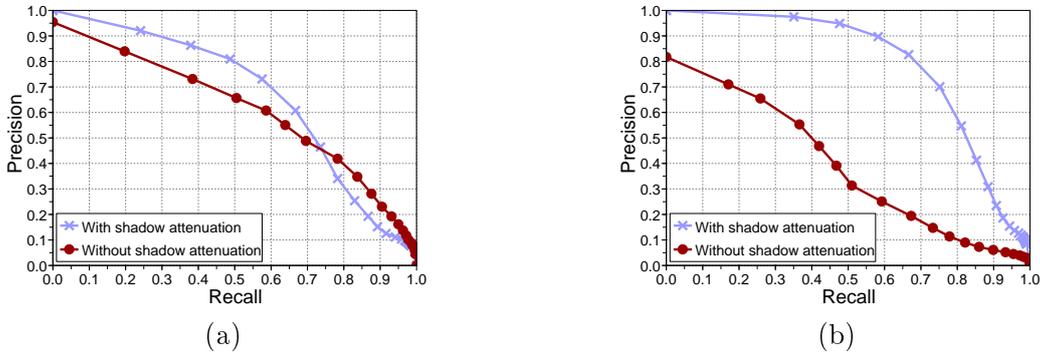
correspondent à des zones ombragées différemment (e.g. les différents cotés des bâtiments). Pour rendre la méthode robuste à la présence d’ombres portées ou de fortes différences d’ombrage, nous avons utilisé la méthode introduite dans [12]. Par une projection appropriée de l’espace des couleurs RGB vers un espace en niveaux de gris, cette méthode permet d’atténuer très fortement toute variation causée par l’illumination, à condition d’avoir préalablement calibré la caméra. La figure 8 illustre les résultats de cette approche, ainsi que le problème de perte d’information liée à la projection utilisée.

**FIGURE 8** – Cette figure illustre les résultats de la méthode d’atténuation d’ombres [12]. Une image est présentée avant (a) et après (b) atténuation. On remarque que le bâtiment au premier plan à gauche a perdu toute texture, illustrant bien la perte d’information due à la projection d’un espace 3D vers un espace 1D. En revanche les ombres sont fortement atténuées, comme le montre la différence de l’image avec une seconde (non présentée) éclairée différemment, avant (c) et après (d) atténuation des ombres.



La figure 9 présente les performances obtenues avec ou sans atténuation des ombres, sur la base avec ou sans ombres portées. Naturellement, on peut voir que dans le cas où les ombres portées sont présentes, la méthode avec atténuation des ombres réalise de bien meilleures performances. En revanche, dans le cas sans ombres portées la conclusion est moins triviale. Pour les valeurs de Rappel inférieures à 0.75, la méthode avec atténuation d’ombres permet d’obtenir des valeurs de Précision supérieures à celles de la méthode sans atténuation d’ombres, ce qui semble naturel puisqu’on filtre mieux les fausses alarmes dues aux fortes différences d’ombrage. En revanche, pour les valeurs de Rappel supérieures à 0.75, les valeurs de Précision obtenues avec atténuation d’ombres sont inférieures à celles de la méthode sans atténuation d’ombres. Ce résultat peut être expliqué par la perte d’information, qui rend parfois les vrais changements indiscernables de leur voisinage. Dans de tels cas, il est très difficile de parvenir à détecter les changements sans réduire considérablement le seuil de décision, ce qui résulte en une forte augmentation des fausses alarmes. Ainsi, lorsque l’on utilise la méthode d’atténuation d’ombres, la perte d’information engendrée résulte en une plus grande difficulté d’obtenir des valeurs de Rappel élevées sans générer un surplus de fausses alarmes.

**FIGURE 9** – Cette figure évalue l’impact de la méthode d’atténuation des ombres sur les performances. Le graphique (a) présente les courbes Précision/Rappel sur la base sans ombres portées, et le graphique (b) présente les courbes sur la base avec ombres portées.



## 5 Exploitation de Vidéos pour la Détection de Changements

Cette section présente les solutions apportées aux problèmes liée à la détection de changements entre une vidéo de test et un ensemble de vidéos de référence.

### 5.1 Constitution de la Base de Référence

Disposant possiblement de plusieurs vidéos de référence, il semble évident qu’une approche brutale consistant à comparer une image de la vidéo de test avec chaque image de chaque vidéo de référence serait totalement inefficace. Afin de ne traiter qu’une fraction des données de références disponible, il est donc nécessaire de les organiser de manière à permettre un accès efficace. [18] introduit pour cela une intéressante structure 3D probabiliste, mais présentant l’inconvénient d’être lourde d’utilisation et seulement adapté aux scène restreintes.

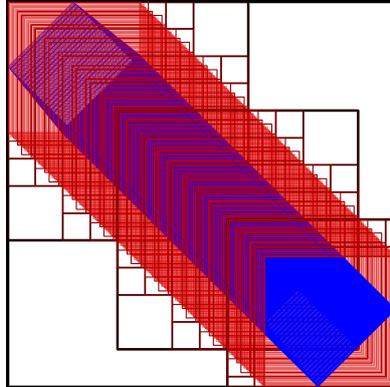
Nous utilisons donc une structure d’organisation plus simple, permettant d’indexer spatialement les images des vidéos de référence en fonction de leur trace au sol. Une comparaison des traces au sol permet d’estimer très rapidement la similarité des contenus des images, et une organisation arborescente de ces traces permet un accès efficace aux images pertinentes vis-à-vis d’une image de test donnée.

La construction de cette base de référence peut être réalisée en pré-traitement, et n’est donc pas soumise aux éventuelles contraintes temps-réel. Par ailleurs, nous supposons connues les paramètres d’acquisition de chaque image de référence, qui en pratique peuvent être déduites des données auxiliaires (position GPS, données gyroscopes, ...) de manière plus ou moins approchée. Pour chaque vidéo de référence, nous calculons alors les traces au sol de chaque image en supposant que le sol est un plan à une altitude connue. Il est à noter qu’en pratique cette hypothèse est généralement suffisante, mais qu’il peut être envisagé par la suite d’exploiter la connaissance d’un Modèle Numérique de Terrain (MNT). Les traces au sol sont ensuite organisées dans un R-Tree, structure arborescente adaptée à l’indexation spatiale d’objets non-ponctuels, selon l’algorithme du Sort-Tile-Recursive [16]. La figure 10 montre une visualisation de l’organisation arborescente d’une donnée vidéo. On peut y voir deux possibilités d’amélioration du temps de requête : d’une part, l’important recouvrement existant entre certains nœuds de l’arbre, et d’autre part, l’inadéquation entre les traces au sol contenues dans les feuilles et leurs rectangles englobants contenus dans les nœuds. Par ailleurs, nous envisageons par la suite d’inclure dans cette base de référence une notion de modèle d’apparence, également réalisé au sol, qui pourrait permettre de parvenir à une meilleure robustesse aux variations d’apparence (e.g. vent dans les arbres, vagues, ...).

### 5.2 Requête et Détection de Changements

Étant donné une image de la vidéo de test, il convient maintenant de récupérer les données de référence pertinentes par une requête dans la base et de procéder à la détection de changement. Du fait

**FIGURE 10** – Cette figure illustre l’organisation des données de référence dans un R-Tree. La vidéo est acquise selon une trajectoire rectiligne depuis le coin supérieur gauche vers le coin inférieur droit. Les quadrilatères bleus sont les traces au sol indexées dans la base, et les rectangles rouges sont les rectangles englobants des différents nœuds de l’arbre, dont le trait est plus épais et plus sombre à mesure qu’on se rapproche du nœud racine.



du fort recouvrement des données, chaque requête renvoie en général plusieurs centaines d’images de référence dont la trace au sol intersecte celle de l’image de test. Comme il serait trop coûteux de faire autant de détection de changements, nous avons choisi de fusionner ces différentes images de référence en une unique mosaïque avant de procéder à la détection. Cette approche introduit deux nouveaux problèmes spécifiques : d’une part, une décision doit être faite sur la manière d’assembler la mosaïque, et d’autre part, le caractère composite de la mosaïque perturbe la géométrie épipolaire qui peut être différente pour chaque fragment d’image utilisée. En effet, les centres optiques des images intégrées dans la mosaïque peuvent être différents, ce qui implique que chaque fragment de la mosaïque finale a un épipôle particulier. Ce point, sensible du fait de l’utilisation de la contrainte épipolaire dans le calcul de flot optique, a été résolu par la construction d’une carte d’épipôles indiquant pour chaque pixel de la mosaïque la position de l’épipôle correspondant.

Concernant le mode d’assemblage de la mosaïque, il a été choisi de favoriser la robustesse aux effets de parallaxe, ce qui requiert que les points de vue des images utilisées pour la détection de changements soient les plus proches possibles. Pour cela, on trie les résultats de la requête par ordre de distance décroissante du centre optique avec celui de la image de test, et on insère les intensités dans la mosaïque selon cet ordre de priorité. On peut alors effectuer la détection de changements avec une composition des images de référence la plus proche possible de l’image de test.

### 5.3 Exploitation de la Redondance Temporelle

Comme mentionné en introduction, la redondance due au format vidéo peut faire l’objet d’un traitement particulier visant à améliorer les performances. Pour cela, nous avons envisagé et comparé deux méthodes : la moyenne temporelle des scores de détection et le paradigme de Propagation de Croissance [2].

L’approche par moyenne temporelle se veut simple et rapide. L’idée consiste, pour chaque pixel, à moyenner temporellement les erreurs d’appariement de flot optique avant la comparaison au seuil (c.f. section 4.2), et à comparer cette moyenne à un seuil pour décider si le pixel correspond à un changement. Cette moyenne temporelle permet d’homogénéiser la décision sur différentes observations successives du même point physique, permettant ainsi de fiabiliser la décision. Comme la caméra peut s’être déplacée entre la capture de deux images successives, il convient cependant de savoir retrouver les correspondances entre pixels afin de comptabiliser correctement les moyennes, et deux approches sont envisageables pour cela. Premièrement, le compte des moyennes peut se faire dans un référentiel absolu, par exemple au sol, mais cela impose des limites a priori sur l’étendue de la scène observée. L’alternative consiste à comptabiliser les moyennes dans un référentiel relatif, le plus naturel étant les coordonnées de

l'image courante, et de recalculer la carte des moyennes à chaque image. C'est cette dernière approche qui a été implémentée pour la comparaison, afin d'éviter d'imposer des limites a priori à la scène observée. Cette solution présente néanmoins un inconvénient de poids : les ré-échantillonnages successifs (dus au recalage de la carte des moyennes à chaque image) détériorent rapidement les données, et cette approche n'est donc utilisable que sur des tronçons assez courts.

La seconde approche, utilisant la Propagation de Croyance, est mieux justifiée théoriquement mais est aussi plus lourde à mettre en œuvre. La suite détaille la modélisation utilisée. Soit  $D_t^{test}$  l'image courante de la vidéo de test, réduite aux dimensions  $(W, H)$  pour accélérer le traitement, dont nous voulons fiabiliser la détection grâce aux  $T - 1$  images précédentes  $\{D_{t-k}^{test}\}_{k \in \llbracket 1, T-1 \rrbracket}$ . Soit  $M_{t-k}^{ref}$ ,  $k \in \llbracket 0, T - 1 \rrbracket$  la mosaïque de référence, initialement de la même taille que l'image de test mais également retaillée aux dimensions  $(W, H)$ , utilisée pour la comparaison avec l'image de test  $D_{t-k}^{test}$ . Enfin, soit  $E_{t-k}$ ,  $k \in \llbracket 0, T - 1 \rrbracket$  la carte des erreurs de flot optique entre  $D_{t-k}^{test}$  et  $M_{t-k}^{ref}$ , également retaillée aux dimensions  $(W, H)$ , et  $\Delta_{i,j,k} = \left| \left( M_{t-k}^{ref}(i, j) - D_{t-k}^{test}(i, j) \right) \right|$  la différence en  $(i, j, k)$  entre l'image de test et la mosaïque de référence. Définissons les intervalles  $\mathcal{W} = \llbracket 0, W - 1 \rrbracket$ ,  $\mathcal{H} = \llbracket 0, H - 1 \rrbracket$  et  $\mathcal{T} = \llbracket 0, T - 1 \rrbracket$ . Pour  $t$  fixé, le graphe utilisé est un champ de Markov aléatoire en 3D, non-orienté, tel que  $G_t = \{V_t, E_t\}$  où l'ensemble des nœuds est  $V_t = \{X_{i,j,k}\}_{(i,j,k) \in \mathcal{W} \times \mathcal{H} \times \mathcal{T}}$ ,  $X_{i,j,k}$  étant une variable aléatoire binaire décrivant l'état (changement ou non) du nœud, et où l'ensemble des arêtes est  $E_t = \{([i, j, k] \rightarrow [i', j', k']), |i - i'| + |j - j'| + |k - k'| = 1\}_{(i,i',j,j',k,k') \in \mathcal{W}^2 \times \mathcal{H}^2 \times \mathcal{T}^2}$ . Chaque variable aléatoire a donc six voisins : quatre spatialement et deux temporellement. Nous définissons le vecteur des observations par  $\mathbf{Y}_{i,j,k} = \begin{pmatrix} \mathbf{Y}_{i,j,k}^1 \\ \mathbf{Y}_{i,j,k}^2 \end{pmatrix} = \begin{pmatrix} E_{t-k}(i, j) \\ \Delta_{i,j,k} \end{pmatrix}$ .

L'attache aux données, liant l'estimation des variables cachées aux observations, est définie par la fonction unaire d'observation  $\phi(X_{i,j,k}, \mathbf{Y}_{i,j,k})$  à l'équation 5. Le second terme de  $f$  signifie que les hautes valeurs d'erreurs  $E_{t-k}(i, j)$  favorisent l'étiquetage comme changement, et vice versa. Par ailleurs, un terme dépendant de  $\Delta_{i,j,k}$  permet de légèrement favoriser l'étiquetage comme changement dans le cas où la valeur de  $\Delta_{i,j,k}$  est importante. Enfin,  $\tau$  est un paramètre utilisé par la suite comme un seuil afin de faire varier Précision et Rappel. En pratique, nous utilisons  $c_0 = 0.6$  et  $c_1 = \frac{\log(2)}{30}$  dont les valeurs ont été obtenues par l'expérience.

$$\phi(X_{i,j,k}, \mathbf{Y}_{i,j,k}) = \begin{cases} f(\mathbf{Y}_{i,j,k}, \tau) & \text{si } X_{i,j,k} = 0 \text{ (pas de changement)} \\ 1 - f(\mathbf{Y}_{i,j,k}, \tau) & \text{si } X_{i,j,k} = 1 \text{ (changement)} \end{cases} \quad (5)$$

avec :

$$f(\mathbf{Y}_{i,j,k}, \tau) = \left[ 1 - c_0 \cdot (1 - \exp(-c_1 \cdot \mathbf{Y}_{i,j,k}^2)) \right] \cdot \left[ \exp\left(-\frac{\mathbf{Y}_{i,j,k}^1}{\tau}\right) \right]$$

Par ailleurs, la fonction de compatibilité des états voisins  $\psi(X_{i,j,k}, X_{i',j',k'})$  permet d'exploiter les relations de voisinage pour renforcer la cohérence spatio-temporelle des étiquetages. L'expression de cette fonction est donnée à l'équation 6. L'allure de la famille de fonctions  $\lambda$  est présentée à la figure 11. Les fonctions  $\lambda(\Delta_{i,j,k}, \Delta_{i',j',k'})$  sont du type Radial Basis Function (RBF) et permettent de favoriser ou non les transitions d'états en fonction du gradient  $|\Delta_{i,j,k} - \Delta_{i',j',k'}|$  de la différence des images  $\Delta_{i,j,k}$ . Lorsque la valeur du gradient est importante, cela signifie que les nœuds  $(i, j, k)$  et  $(i', j', k')$  correspondent probablement à différents objets de la scène. Dans un tel cas, il n'est pas nécessaire de favoriser des états identiques pour les deux nœuds, d'où la pondération du second terme par 0.5. Quand le gradient est faible, les deux nœuds correspondent probablement à un même objet, et on favorise fortement les états similaires (pondération par 0.95) par rapport aux états différents (pondération par 0.05). En pratique, nous utilisons pour la fonction  $\lambda$ , les valeurs de paramètre  $b = 0.175$  et  $a = 25$  qui donnent de bons résultats.

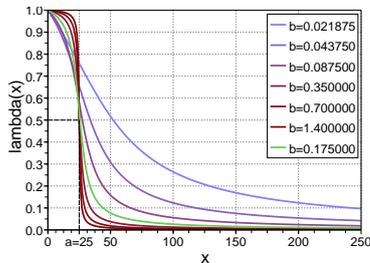
$$\psi(X_{i,j,k}, X_{i',j',k'}) = \begin{cases} 0.95 \cdot \lambda(\Delta_{i,j,k}, \Delta_{i',j',k'}) + 0.5 \cdot (1 - \lambda(\Delta_{i,j,k}, \Delta_{i',j',k'})) & \text{si } X_{i,j,k} = X_{i',j',k'} \\ 0.05 \cdot \lambda(\Delta_{i,j,k}, \Delta_{i',j',k'}) + 0.5 \cdot (1 - \lambda(\Delta_{i,j,k}, \Delta_{i',j',k'})) & \text{si } X_{i,j,k} \neq X_{i',j',k'} \end{cases} \quad (6)$$

avec :

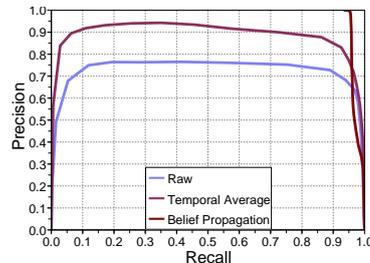
$$\lambda(\Delta_{i,j,k}, \Delta_{i',j',k'}) = \frac{\frac{\pi}{2} - \arctan(b \cdot (|\Delta_{i,j,k} - \Delta_{i',j',k'}| - a))}{\frac{\pi}{2} - \arctan(-b \cdot a)}$$

L'algorithme de Propagation de Croyance permet alors d'estimer, en chaque nœud du graphe, la probabilité qu'il corresponde à un changement. Cet algorithme consiste à propager, de chaque nœud vers ses voisins, des messages portant une information de vraisemblance relative à chaque état possible (voir [2] pour plus de détails). Il a été démontré que sous certaines conditions, cet échange de message converge. Dans notre cas, la convergence n'est pas garantie du fait des boucles présentes dans le graphe, mais l'expérience montre qu'il y a bien convergence, vers une solution intéressante. La figure 12 présente les performances comparées des deux approches détaillées ci-dessus, ainsi que les performances sans exploitation de la redondance temporelle. On peut noter que la courbe correspondant à l'approche par Propagation de Croyance n'est pas complète, car la plage de variation du seuil n'a pas permis d'atteindre de faibles valeurs de Rappel. Cependant, le fragment de courbe montre qu'il existe un point de fonctionnement donnant d'excellents résultats. Par ailleurs sur ce graphique, on peut globalement voir que l'exploitation de la redondance améliore les performances. On peut aussi noter que l'approche par Propagation de Croyance, permettant un traitement plus informé, résulte en de meilleures performances que l'approche par moyenne temporelle. En revanche, cette complexité accrue se traduit par des temps de traitements plus long pour l'approche par Propagation de Croyance.

**FIGURE 11** – Cette figure illustre la forme de la fonction RBF lambda utilisée dans la matrice de compatibilité des états. La courbe utilisée est tracée en vert.



**FIGURE 12** – Cette figure présente les performances obtenues avec les différentes méthodes d'exploitation de la redondance temporelle des vidéos.



## 6 Conclusion

En conclusion, cette première moitié de thèse a permis une réflexion constructive sur les méthodes de détection de changements existantes, et sur les problèmes rencontrés lorsque l'on travaille sur des paires d'images ou sur de multiples vidéos. Cette réflexion a débouché sur la conception d'un algorithme innovant [4] de détection de changements dans des paires d'images, à la fois rapide et performant, ainsi que sur la proposition d'un système de détection de changements entre plusieurs vidéos.

Nous l'avons vu, il reste cependant un certain nombre d'axes d'amélioration et de problèmes à résoudre. On peut par exemple mentionner le problème de la gestion des faces d'objets occultées, ou l'intégration d'un modèle d'apparence dans la base de référence afin d'éviter une certaine proportion de fausses alarmes dues aux variations d'apparence. Plus particulièrement, le problème de l'exploitation de la redondance temporelle des vidéos est peu étudié dans la littérature, et dispose d'un potentiel intéressant vis-à-vis de l'amélioration des performances. Le développement d'une solution à la fois rapide et fournissant de bonnes performances semble donc être un objectif à privilégier.

## Références

- [1] H. Bay, T. Tuytelaars, and L. VanGool. Surf : Speeded up robust features. *Proceedings of the European Conference on Computer Vision*, pages 404–417, 2006.
- [2] C. Bishop. Graphical models. In *Pattern Recognition and Machine Learning*, pages 359–422. Springer, 2006.
- [3] M. Black, D. Fleet, and Y. Yacoob. Robustly estimating changes in image appearance. *Computer Vision and Image Understanding*, 78 :8–31, 2000.
- [4] N. Bourdis, D. Marraud, and H. Sahbi. Constrained optical flow for aerial image change detection. In *2011 IEEE International Geoscience and Remote Sensing Symposium*. IEEE Computer Society, 2011.
- [5] A. Buchanan. Novel view synthesis for change detection. In *Proceedings of the 2009 Conference of Electro Magnetic Remote Sensing Defence Technology Centre*, 2009.
- [6] M. Carlotto. Detecting change in images with parallax. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6567, page 42, 2007.
- [7] H. Chui. *Non-Rigid Point Matching : Algorithms, Extensions and Applications*. Phd, Yale University, 2001.
- [8] C. Clifton. Change detection in overhead imagery using neural networks. *Applied Intelligence*, 18(2) :215–234, 2003.
- [9] S. Cluff. *A Unified Approach to GPU-Accelerated Aerial Video Enhancement Techniques*. Phd, Brigham Young University, 2009.
- [10] D. Crispell. *A Continuous Probabilistic Scene Model for Aerial Imagery*. Phd, Brown University, 2010.
- [11] G. Farneback. Two-frame motion estimation based on polynomial expansion. *Image Analysis*, pages 363–370, 2003.
- [12] G. Finlayson, S. Hordley, C. Lu, and M. Drew. On the removal of shadows from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 59–68, 2006.
- [13] M. Fischler and R. Bolles. Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6) :381–395, 1981.
- [14] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [15] R. Kumar, P. Anandan, and K. Hanna. Shape recovery from multiple views : A parallax based approach. *Proceedings of the International Conference on Pattern Recognition*, 1994.
- [16] S. Leutenegger, M. Lopez, and J. Edgington. Str : A simple and efficient algorithm for r-tree packing. In *Proceedings of the 13th International Conference on Data Engineering (1997)*, pages 497–506, 1997.
- [17] F. Nava, A. Nava, J. Lamolda, and M. Redondo. Change detection for remote sensing images with graph cuts. *SPIE, Image and Signal Processing for Remote Sensing XI*, 5982 :229–242, 2005.
- [18] T. Pollard. *Comprehensive 3-d Change Detection Using Volumetric Appearance Modeling*. Phd, Brown University, 2009.
- [19] K. Primdahl, I. Katz, O. Feinstein, Y. Mok, H. Dahlkamp, D. Stavens, M. Montemerlo, and S. Thrun. Change detection from multiple camera images extended to non-stationary cameras. *Proceedings of Field and Service Robotics*, 2005.
- [20] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms : a systematic survey. *IEEE Transactions on Image Processing*, 14(3) :294–307, 2005.
- [21] C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8) :747–757, 2000.
- [22] C. Stennett and R. Evans. Visual change detection for route monitoring. In *Proceedings of the 2009 Conference of Electro Magnetic Remote Sensing Defence Technology Centre*, 2009.
- [23] D. Toth, T. Aach, and V. Metzler. Illumination-invariant change detection. In *4th IEEE Southwest Symposium on Image Analysis and Interpretation*, page 3, 2000.
- [24] S. Watanabe and K. Miyajima. Detecting building changes using epipolar constraint from aerial images taken at different positions. In *Proceedings of the International Conference on Image Processing (2001)*, volume 2, pages 201–204, 2001.
- [25] B. Zitová and J. Flusser. Image registration methods : a survey. *Image and Vision Computing*, 21(11) :977–1000, 2003.



