



**HAL**  
open science

# Coefficients de fiabilité et approche hiérarchique pour la détection et le dénombrement de petits objets dans une vidéo

Valentina Pestova

► **To cite this version:**

Valentina Pestova. Coefficients de fiabilité et approche hiérarchique pour la détection et le dénombrement de petits objets dans une vidéo. Traitement des images [eess.IV]. Université de Bordeaux, 2018. Français. NNT : 2018BORD0462 . tel-02073635

**HAL Id: tel-02073635**

**<https://theses.hal.science/tel-02073635>**

Submitted on 20 Mar 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE PRÉSENTÉE  
POUR OBTENIR LE GRADE DE

**DOCTEUR DE**

**L'UNIVERSITÉ DE BORDEAUX**

ÉCOLE DOCTORALE

MATHEMATIQUES ET INFORMATIQUES

Par ValentinaPESTOVA

COEFFICIENTS DE FIABILITE ET APPROCHE  
HIERARCHIQUE POUR LA DETECTION ET LE  
DENOMBREMENT DE PETITS OBJETS DANS UNE VIDEO

Sous la direction de : Henri NICOLAS

Soutenue le 21 décembre 2018

Membres du jury:

M. DOMENGER Jean-Philippe	Professeur	Université de Bordeaux, LaBRI	Président
M. BEN AMAR Chokri	Professeur	ENIS de Sfax, Tunisie	rapporteur
M. JOLY Philippe	Professeur	IRIT, Toulouse	rapporteur
M. NICOLAS Henri	Professeur	Université de Bordeaux	examinateur

**Titre :** Coefficients de fiabilité et approche hiérarchique pour la détection et dénombrement des petits objets dans les vidéos

**Résumé :**

Le problème du dénombrement d'un grand nombre de petits objets en mouvement dans les vidéos est un contexte applicatif jusqu'à présent peu étudié. Dans ce cadre, la difficulté réside essentiellement dans le fait qu'en raison de leurs petites tailles apparentes dans la vidéo, il n'est pas possible de définir un modèle géométrique fiable de ces objets. Or, les travaux existants dans le domaine de la détection d'objets dans des vidéos, utilisent souvent un tel modèle des objets d'intérêt. Les méthodes de détection existantes ne sont de ce fait pas applicables directement dans le cadre de la détection de tels petits objets. Dans le cadre de cette thèse, on propose une méthodologie complète pour la détection de nombreux petits objets, avec un cadre applicatif visant plus particulièrement la détection et le comptage d'oiseaux migrateurs dans une vidéo. Le principe innovant, proposé comme une solution de ce problème, consiste à associer des coefficients de fiabilité de détection aux objets pour les dénombrer en évitant de prendre en compte de trop nombreuses fausses détections. Un algorithme hiérarchique analysant l'aspect spatio-temporel d'objets (leurs apparence et l'évolution dans le temps) dans une vidéo à l'aide de méthodes de traitement d'images, de statistique et de la logique floue est ainsi proposé. Le but des coefficients de fiabilité est d'estimer la probabilité que les paramètres d'une détection correspondent aux paramètres attendus pour les objets d'intérêt. Finalement, l'ensemble des coefficients est converti en une valeur qui évalue la séquence du traitement d'un objet. La somme de ces valeurs correspond au nombre d'objets d'intérêt dans une vidéo. Les résultats obtenus montrent que les bonnes détections sont pour la plupart comprises dans le dénombrement avec des coefficients de fiabilité égaux ou proche de 1, et les fausses détections sont supprimées ou sous-pondérées avec des coefficients de fiabilité plus faible. Les résultats de comptage dans des vidéos contenant de très nombreux oiseaux sont proches de la vérité terrain, ce qui prouve la validité de la solution proposée comme un moyen de dénombrement automatique d'objets dans des vidéos.

**Mots clés :** traitement vidéo, analyse d'objets, détection et comptage de petits objets, coefficients de fiabilité

**Title :** Reliability coefficients and hierarchical approach for the detection and counting of small objects in videos

**Abstract :**

The problem of counting of big volumes of small moving objects in videos is a domain, which was not studied to date. The difficulty of this application consists essentially in the fact, that because of small sizes of objects, apparent in the videos, definition a reliable geometric model of these objects is impossible. The existing researches for object detection in videos frequently use a geometrical model of objects of interest. For this reason, the existing methods of object detection cannot be applied for the detection of very small objects in the study case. This thesis proposes a complete methodology, allowing the detection of very small objects in videos, and designed particularly the detection and counting of migrating birds in videos. An innovative principle and the solution of this problem consist in association of coefficients of detection reliability to the objects, in order to count them, avoiding counting of many false detections. The solution proposes a hierarchical algorithm, which analyses the spatial and temporal aspects of objects (their appearance and evolution in time) in a video, by the means of methods of image processing, statistics, and fuzzy logic. The aim of the reliability coefficients is to estimate the probability, that the parameters of a detected objects conform to the expected parameters of the objects of interest. Finally, the coefficients are put together and converted into a value, which evaluates the sequence of processing, applied to detect an object. The sum of these values corresponds to the number of the objects of interest in a video. The results show, that the most of correct detections are characterized in the counting by the reliability coefficient equal or close to 1. The results show, that the most of correct detections have their reliability coefficients close to 1, and the false detection are deleted or have low reliability coefficients. The counting results in the videos with numerous groups of migrating birds are close to the ground truth. This validates the proposed solution as a method of automatic counting of objects in videos.

**Keywords :** video processing, image processing, small object detection, reliability coefficients, object counting

Unité de recherche : LaBRI - Laboratoire Bordelais de Recherche en Informatique

Numéro de l'unité de recherche : UMR5800

Adresse de l'unité de recherche :

LaBRI - Bâtiment A30

351 cours de la Libération

33405 TALENCE CEDEX

## Sommaire

<b>1. Introduction générale.....</b>	<b>6</b>
<b>1.1 Brève présentation des principes et des objectifs de l’algorithme proposé .....</b>	<b>8</b>
<b>1. Etat de l’art.....</b>	<b>11</b>
<b>2.1 Méthodes non numériques pour le comptage de très petits objets en mouvement... ..</b>	<b>12</b>
<b>2.2 Détection d’objets dans les images : un état de l’art .....</b>	<b>13</b>
2.2.1 Modélisation et soustraction d’arrière-plan.....	14
2.2.2 Seuillage d’images .....	14
2.2.3 Détection de contours d’objets.....	14
2.2.4 Modélisation d’objets.....	15
<b>2.3. Détection d’objets en mouvement dans les vidéos : un état de l’art .....</b>	<b>16</b>
2.3.1 Suivi de formes d’objets similaires .....	16
2.3.2 Différences entre les images consécutives .....	16
2.3.3 Modélisation d’un mélange Gaussien .....	16
2.3.4 Correspondance des points caractéristiques (« points d’intérêt »).....	17
2.3.5 Analyse des histogrammes .....	17
2.3.6 Flux optique .....	17
<b>2.4 Détection des très petits objets dans l’état de l’art .....</b>	<b>17</b>
2.4.1 Hypothèses sur la taille et le modèle d’objets .....	18
2.4.2 Méthodes existantes de détection d’oiseaux d’une taille normale dans le contexte de la détection d’oiseaux.....	18
2.4.3 Méthodes existantes de détection de petits objets dans le contexte de détection des oiseaux .....	18
2.4.4 Méthodes existantes pour la détection de très petits objets pour le suivi d’oiseaux migrateurs.....	19
<b>2.5 Conclusion sur l’état de l’art dans les domaines de détection et de suivi d’objets dans les vidéos .....</b>	<b>19</b>
2.5.1. La possibilité d’application des méthodes existantes pour le suivi et le dénombrement des oiseaux-migrateurs.....	20
2.5.2. Intérêt de la détection et du dénombrement en fonction de la fiabilité de détection .....	21
<b>3.1 Objectifs globaux du projet Octobre Bleu .....</b>	<b>22</b>
<b>3.2 Positionnement des travaux de recherche dans le projet.....</b>	<b>22</b>
<b>3.3 Présentation des corpus vidéo .....</b>	<b>23</b>
3.3.1 Standardisation de caractéristiques des vidéos.....	23
3.3.1.1 Caractérisation de la difficulté de traitement d’une vidéo.....	25
3.3.2 Analyse des critères des niveaux de complexité .....	29
<b>4. Méthodologie proposée pour la détection et le dénombrement de très petits objets.....</b>	<b>31</b>

<b>4.1 Hypothèses admises dans l'approche proposée .....</b>	<b>31</b>
<b>4.2 Présentation générale de l'approche proposée pour détection et dénombrement de petits objets dans des vidéos .....</b>	<b>35</b>
4.2.1 Formalisation de la solution proposée.....	41
<b>4.3 Formalisation du modèle de petit objet d'intérêt .....</b>	<b>42</b>
4.3.1 Forme d'objet.....	42
4.3.2 Texture .....	42
4.3.3 Comportement dynamique .....	43
4.3.4 Modèle d'objets pour les oiseaux migrateurs.....	43
4.3.5 Limites du modèle d'objets introduit dans la solution .....	45
<b>4.4 Présentation de l'aspect spatial de la solution proposée pour la détection de très petits objets.....</b>	<b>48</b>
4.4.1 Domaines scientifiques dans l'aspect spatial .....	49
4.4.1.1. Application de méthodes statistiques pour l'analyse de zones dans les espaces colorimétriques.....	50
4.4.2 Prétraitement de la vidéo pour l'initialisation de la première étape.....	52
4.4.2.1 Principe d'estimation du critère de classification de zones avec des objets d'intérêt .....	53
4.4.2.2. Analyse qualitative : corrélation entre les histogrammes .....	54
4.4.2.3. Analyse quantitative : la distance du $\chi^2$ entre les histogrammes .....	55
4.4.2.4. Combinaison de la corrélation et de la distance de $\chi^2$ entre les histogrammes sur l'étape de prétraitement.....	55
4.4.2.5. Formalisation de la méthode d'estimation du seuil de variation d'intensité lumineuse de couleurs par l'analyse des histogrammes appliquée en prétraitement .....	57
4.4.3 Première phase : Evaluation à priori de zones susceptibles de contenir des objets d'intérêt .....	58
4.3.3.1 Description de la phase d'évaluation à priori de zones .....	59
4.3.3.2 Coefficients de fiabilité sur la phase d'évaluation de zones d'intérêt .....	62
4.3.3.3 Objectif et principe d'application du coefficient d'intensité de contraste dans une zone .....	64
4.3.3.4 Expérimentations et analyse de la méthode d'évaluation de zones d'intérêt et du coefficient de fiabilité de détections dans une zone .....	66
4.3.3.5 Conclusion sur l'application des coefficients de fiabilité pour les zones d'image.....	70
4.4.4 Deuxième phase : Détection d'objets dans les zones sélectionnées.....	71
4.4.4.1 Bases théoriques de la méthode de détection de contours.....	71
4.4.4.2 Notions de statistique descriptive et de paramètres de distribution.....	72
4.4.4.2 Méthodes de détection des contours.....	73
4.4.4.3 Filtre adaptatif proposé : application du filtre Laplacien standard et analyse statistique.....	76

4.4.4.4 Première itération et étape d'adaptation du filtre Laplacien adaptatif.....	77
4.4.4.5 Estimation du seuil de l'intensité de contours d'objets d'intérêt en fonction de la caractéristique d'asymétrie .....	79
4.4.4.6 Deuxième itération du filtre Laplacien adaptatif : application de la combinaison linéaire des matrices paramétrées .....	81
4.4.4.7 Justification analytique d'application du filtre adaptatif .....	82
4.4.4.8 Seuillage par la fonction sigmoïde du deuxième masque de contours .....	84
4.4.4.9 Formalisation de la technique de seuillage proposée .....	86
4.4.4.10 Coefficients de fiabilité de détection de contours .....	88
4.3.4.11 Analyse de résultats d'application du coefficient de détection des contours .....	88
4.4.5 Conversion de contours de petits objets d'intérêt en régions continues.....	89
4.3.5.1 Présentation de l'approche de conversion de contours de petits objets en régions continuées .....	90
4.4.5.2 Coefficients de fiabilité de conversion des contours détectés en régions continues.....	92
4.4.5.3 Principe de classification de pixels d'une région continue.....	93
4.4.5.4 Principe de l'estimation du coefficient de fiabilité et son effet sur le résultat final.....	94
4.4.5.5. Exemples d'estimation du coefficient de fiabilité .....	98
<b>4.5 Aspect temporel et sa contribution dans l'aspect spatial.....</b>	<b>99</b>
4.5.1 Introduction de l'aspect temporel.....	99
4.5.2 Prise en compte des coefficients de fiabilité liés à l'aspect temporel .....	101
4.5.3. Troisième phase : Estimation des paramètres du modèle théorique de petits objets et évaluation de conformité.....	102
4.5.3.1. Application de la régression linéaire pour la modélisation de petits objets d'intérêt .....	103
4.5.3.2. Composition de la méthode de MCO et de la régression linéaire pour la modélisation de petits objets d'intérêt.....	104
4.5.3.3. Analyse de la variance résiduelles pour l'estimation de qualité du modèle .....	105
4.5.4. Estimation et rectification des coefficients de fiabilité de conformité au modèle d'objet d'intérêt.....	106
4.5.4.1 Estimation du coefficient de conformité au modèle.....	107
4.5.4.2 Approche proposée pour l'ajustement des coefficients de fiabilité de conformité au modèle.....	108
4.5.4.3 Sélection d'objets conformes au modèle par la couleur théorique.....	112
4.5.5 Coefficient final de conformité d'objet au modèle d'objet d'intérêt.....	113
<b>4.6 Phase de comptage en fonction des coefficients de fiabilité .....</b>	<b>114</b>
4.6.1 Principe de combinaison de coefficients de fiabilité d'objet d'intérêt en un coefficient final .....	114
4.6.2 Fonctions d'appartenance aux ensembles flous utilisés .....	115

4.6.3 Méthodes de combinaison de coefficients .....	116
4.6.3.1 Application du mécanisme d'inférence de Mamdani .....	117
4.6.3.2 Dénombrement d'objets détectés .....	121
<b>5. Expérimentations et résultats.....</b>	<b>122</b>
<b>5.1 Résultats expérimentaux sur les vidéos du corpus 1 .....</b>	<b>122</b>
<b>5.2 Analyse des étapes hiérarchiques de la solution proposée .....</b>	<b>125</b>
<b>5.3 Résultats expérimentaux sur les vidéos du corpus 2 et la conception du modèle de     « très petits objets » .....</b>	<b>129</b>
<b>5.4 Analyse des résultats d'application de la solution sur les vidéos typiques     d'observation de migrations pour leur suivi et dénombrement .....</b>	<b>129</b>
<b>6. Conclusion.....</b>	<b>134</b>
<b>7. Discussion.....</b>	<b>135</b>
<b>7.1 Contributions des travaux de recherches effectués .....</b>	<b>135</b>
<b>7.2 Perspectives et travail ultérieur .....</b>	<b>137</b>
<b>8 Architecture modulaire du projet Octobre Bleu .....</b>	<b>138</b>
<b>8.1 Perspectives d'extensions sur d'autres types d'applications .....</b>	<b>138</b>
<b>9 Annexe.....</b>	<b>140</b>
<b>9.1. Exemples de détections dans certaines vidéos du corpus 2.....</b>	<b>140</b>
<b>10 Bibliographie .....</b>	<b>149</b>

## 1. Introduction générale

La thématique principale des travaux de recherches présentés dans le cadre de cette thèse se situe dans le contexte du suivi et du comptage d'objets vidéo plus particulièrement dans le cadre d'applications de type vidéosurveillance. D'une manière générale, les techniques de vidéosurveillance trouvent des applications dans des domaines variés tels que, par exemple :

- La surveillance des lieux publics (affluence dans les transports en commun, vidéosurveillance dans les rues ...).
- L'analyse du trafic routier (densité ou fluidité, détection des embouteillages...).
- Les sciences naturelles, tels que la biologie ou l'ornithologie (e.g. suivi des oiseaux migratoires ou détection des insectes).

La capacité de détecter, analyser et dénombrer de très petits objets en mouvements (nommé par la suite « les objets d'intérêt ») est important dans le cadre de certaines applications pour lesquelles les appareils d'enregistrement de vidéos se trouvent à une grande distance, comparativement à la taille des objets d'intérêt. Citons par exemple des vidéos d'oiseaux volant à haute altitude, ou des colonies d'insectes. Typiquement dans ce type d'applications, il peut y avoir dans les images un nombre très importants d'objets d'intérêt (jusqu'au moins des centaines d'objets). Dans un tel contexte, il est très difficile, voire impossible, d'obtenir un modèle géométrique fiable des objets d'intérêt d'autant que les objets peuvent être déformables (tel un oiseau dont les battements d'ailes modifient la géométrie apparente dans l'image).

Dans ce contexte, le besoin de développer une approche générale de détection et de dénombrement de très petits objets et à la base des travaux de recherche développée dans le cadre de cette thèse.

Le contexte applicatif visé est plus particulièrement celui du dénombrement d'oiseaux migrateurs qui est le thème essentiel du projet Octobre Bleu au sein duquel cette thèse a été réalisée.

Ce projet est destiné à plusieurs applications dans des contextes variés du domaine de vidéosurveillance pour des différents objectifs (vidéosurveillance dans les transports en commun, contrôle du trafic routier, gestion des populations d'oiseaux migrateurs).

Le sujet de la thèse a été concentré sur la prise en charge des besoins de recherches ornithologiques liés aux traitements d'images et de la vidéo tel qu'exprimés par la Fédération Départementale des Chasseurs de la Dordogne (FDC) et du GIFS (émanation de fédérations des chasseurs départementales de France). Ces besoins concernent l'automatisation de l'observation et du suivi des migrations du pigeon ramier (palombe) au moyen de l'analyse de vidéos. L'objectif des travaux de recherches est de faciliter l'acquisition de données et d'améliorer les résultats d'observations. La difficulté de la conception et du développement d'un algorithme pour atteindre ces objectifs sont dues aux particularités des données vidéo obtenues sur le terrain :

- Les vidéos contenant les vols d'oiseaux migratoires sont enregistrées depuis des postes d'observation situés à des distances importantes des objets filmés : typiquement de 500m à 750m.
- Les vidéos sont enregistrées au format HD (avec la résolution 1920x1080 pixels) ce qui représente un volume significatif de données à traiter susceptible de générer une complexité opératoire élevée.
- Une distance importante entre la caméra et les objets filmés fait que la forme apparente des objets est peu précise. Ce phénomène ne permet pas d'appliquer les approches existantes



basées sur la reconnaissance de forme ou le suivi par des méthodes de mises en correspondance de points clés.

Les méthodes appliquées actuellement dans le cadre des recherches ornithologiques pour le suivi de migrations emploient soit des approches de comptage largement manuel des oiseaux, soit des techniques nécessitant des dispositifs spécifiques qui permettent de distinguer les objets suivis, tel que par exemple des bagues ou des émetteurs-radio. De tels dispositifs ne peuvent pas être utilisés pour le comptage des gros flux de migrations car il est bien évidemment impossible de mettre une bague ou un émetteur sur les centaines voire les milliers des oiseaux que peut comporter un seul vol. L'utilisation de techniques de traitement d'images apparait de ce fait comme une solution prometteuse. Néanmoins, les méthodes de traitement vidéo existantes ne sont le plus souvent pas applicables directement lorsqu'en raison de la très faible taille des objets leur forme change d'une image à l'autre de manière non prédictive. De plus, une forme spécifique pour de très petits objets ne pouvant être définie, le risque d'obtenir de très nombreuses fausses détections est très important.

En raison des caractéristiques mentionnées ci-dessus, la détection et le dénombrement de très petits objets en mouvement dans les vidéos (appliqués ici pour le cas des oiseaux dans des flux migratoires filmés à une grande distance) nécessite le développement d'approches algorithmiques innovantes qui permettant de gérer les limites imposées dans le contexte applicatif visé. Cet objectif a été placé au centre des recherches effectuées dans le cadre de cette thèse. Pour prendre en compte de manière structurée les risques de fausses détections, la notion de critère de fiabilité est introduite au sein de chaque phase algorithmique du processus de détection. Pour chaque objet détecté, ce critère, exprimé comme une probabilité, définit le poids de chaque détection.

Cette approche permet d'interpréter une détection en tant qu'une composition de fiabilité d'exécution des étapes d'un algorithme hiérarchique et en déduire le nombre final d'objets dans une vidéo. En cours d'exécution les entités traitées convergent du niveau global de toute une vidéo au niveau local d'un objet pour prendre en considération deux principes complémentaires : l'aspect spatial et l'aspect temporel de la représentation d'un objet dans une vidéo. L'application de cet aspect spatio-temporel permet de définir un modèle d'objet d'intérêt, qui caractérisera les détections dans une vidéo.

Les recherches menées dans le cadre de cette thèse concernent essentiellement le développement d'une approche prenant en compte une analyse spatiale de la scène. Les aspects temporels ont été traités par ailleurs dans le cadre du projet Octobre Bleu. Néanmoins, ces deux aspects sont clairement complémentaires et des résultats globaux intégrant les deux aspects sont présentés dans ce document.

Pour obtenir ces résultats, des méthodes issues des domaines du traitement d'images, de la logique floue et de la statistique sont utilisées. Les techniques du traitement d'images (tels que le filtrage et les opérations morphologiques) sont précédées par l'analyse statistique des zones d'image traitées. Les méthodes de la logique floue sont ensuite appliquées pour estimer la fiabilité du traitement effectué par rapport au modèle d'objet d'intérêt estimé depuis les données d'entrée.

Les résultats expérimentaux montrent que l'introduction de la notion de fiabilité estimée sous une forme numérique au moyen de l'utilisation de plusieurs domaines scientifiques permet d'affiner la représentation des très petits objets d'intérêt dans des vidéos et d'obtenir un comptage proche de la vérité terrain et peu sensible aux fausses détections.

L'algorithme proposé, bien qu'il a été développé plus spécifiquement pour le comptage des oiseaux migrateurs, est destiné à être généralisé à d'autres contextes d'applications, tel que la surveillance du trafic routier, le comptage de foules ou la détection/comptage des insectes.

## 1.1 Brève présentation des principes et des objectifs de l'algorithme proposé

Les travaux de recherches effectués au cours de cette thèse étaient concentrés sur le suivi et le dénombrement d'oiseaux au moyen du traitement et de l'analyse de vidéos. Les objectifs sont l'automatisation des processus permettant de contribuer efficacement aux recherches ornithologiques, tout en réduisant la main-d'œuvre impliquée, en améliorant les résultats du dénombrement par rapport au dénombrement manuel, et en améliorant la confiance dans le résultat obtenu. Pour répondre à ces objectifs, les recherches dans le cadre de la thèse consistaient en la conception d'une méthode novatrice basée sur le traitement d'images et vidéos pour fournir un moyen automatique de dénombrement d'oiseaux-migrateurs dont la précision serait comparable avec la vérité terrain basée sur le dénombrement manuel faite par plusieurs personnes qualifiées.

Pour obtenir un tel résultat, une méthode innovante a dû être proposée pour prendre en considération les particularités des vidéos filmées depuis des postes d'observation et les difficultés de détection dû aux caractéristiques des vidéos. La difficulté essentielle étant de garantir la détection du plus grand nombre possible d'objets d'intérêt, tout en évitant de générer de trop nombreuses fausses détections.

Pour cette raison, des approches basées sur la logique floue ont été choisies pour estimer les fiabilités de traitement propre à chaque étape de la méthode proposée pour obtenir une valeur globale de fiabilité pour chaque objet détecté. L'idée de base étant que les bonnes détections sont globalement plus proches du modèle d'objet retenu que les fausses détections.

La fiabilité de leurs détections est évaluée dans chaque frame et à chaque étape de l'algorithme. Un coefficient de fiabilité est ensuite associé à une étape et attribué à un objet correspondant. A la fin du traitement d'une vidéo pour chaque objet détecté ses coefficients de fiabilité sont combinés en un seul coefficient global pour représenter sa fiabilité. Ce coefficient de fiabilité caractérise le niveau de précision de détection d'un objet et contribue implicitement au dénombrement final.

Ainsi les détections « fiables » sont comprises intégralement dans le résultat de dénombrement. Les détections générées par le bruit ou sur des étapes intermédiaires, obtiennent le plus souvent des valeurs faibles de leurs coefficients de fiabilité et donc influencent d'une manière peu significative le nombre de détections finales.

Le comptage n'est alors pas réalisé en comptant directement le nombre d'objets d'intérêt. Le comptage final est obtenu plutôt en réalisant la somme des coefficients de fiabilité de l'ensemble des objets détectés.

Sur la base de ces principes généraux, l'approche proposée dans la thèse se compose de plusieurs étapes et combine des techniques différentes et des domaines scientifiques variés. Cette séquence de traitements est introduite ci-dessous.

Afin de pouvoir traiter les vidéos sous les contraintes mentionnées ci-dessus dans le cadre du concept de fiabilité de détections, l'algorithme se compose de plusieurs étapes auxquelles est associé un coefficient. Les étapes principales de l'algorithme proposé sont les suivants (voir le schéma de la figure 1.1) :

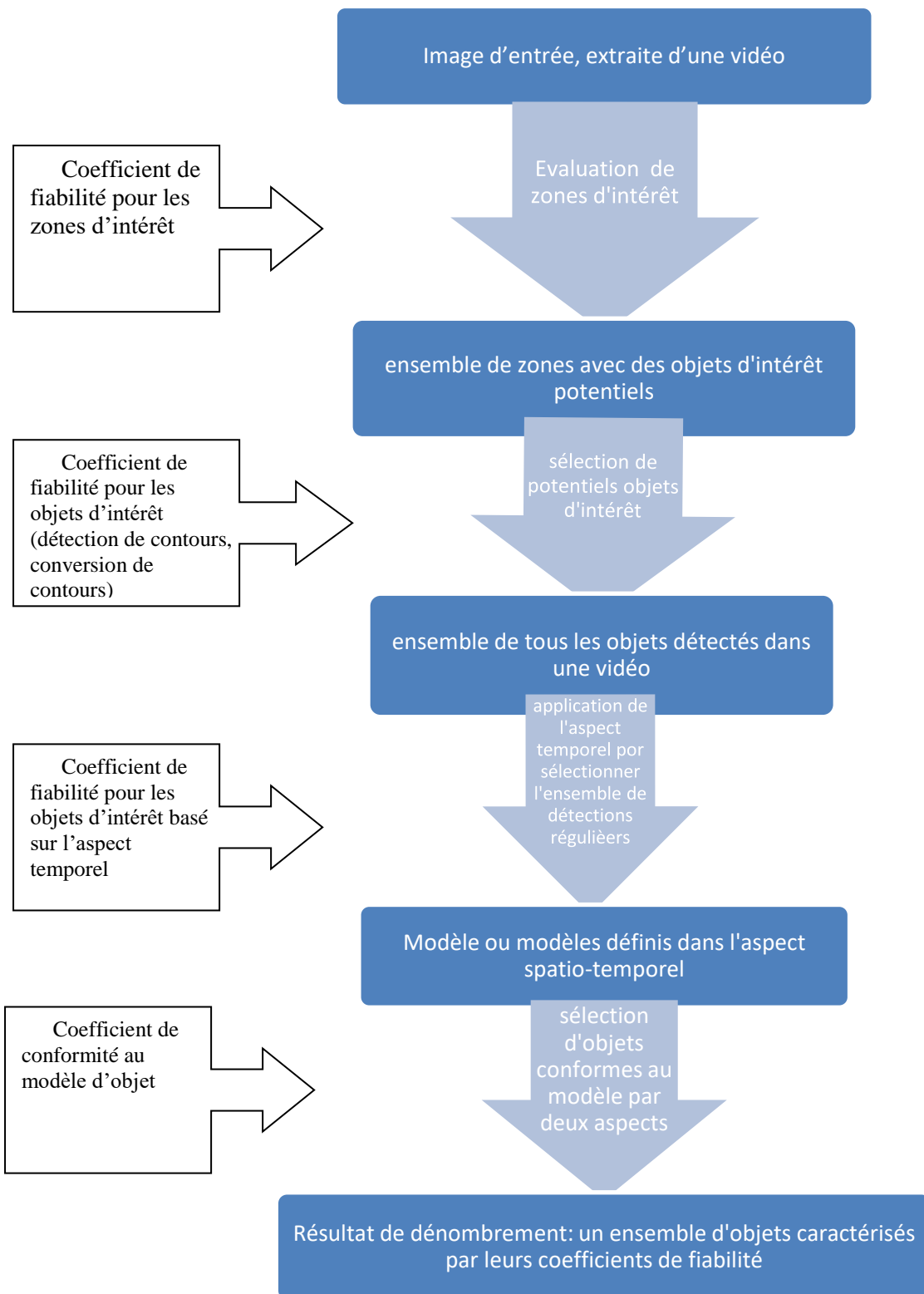


Figure 1.1 : Schéma conceptuel de l'algorithme proposée pour la détection et dénombrement de très petits objets dans les vidéos

- 1) **Modèle d'objets d'intérêt** : Un modèle spatio-temporel des objets d'intérêt est défini sur la base de leurs caractéristiques géométriques, colorimétriques et comportementaux. Ce modèle est utilisé tout au long du processus de détection pour identifier au mieux les objets d'intérêt. Les descripteurs de ce modèle sont soit fournis par l'utilisateur, soit estimés automatiquement. Dans ce dernier cas, cela nécessite une initialisation des paramètres du modèle. La méthode développée pour l'estimation des paramètres du modèle est détaillée dans la section 4.5.

2) **Caractérisation de la difficulté intrinsèque de la détection des objets d'intérêt dans chaque zone de l'image.** D'une manière générale, plus une zone de l'image contient une texture proche de celle des objets d'intérêt, plus le risque d'obtenir de nombreuses fausses détections est élevé. La fiabilité d'une détection dans de telles zones est alors par nature faible. Un coefficient de fiabilité est alors défini par zone (pour des raisons de facilité d'implémentation, une approche par blocs carrés a été retenue) en fonction de la proximité de leurs textures par rapport au modèle d'objet. Les zones pour lesquelles un coefficient de fiabilité égal à zéro sont de fait éliminer des phases d'analyse ultérieures.

3) **Détection des objets d'intérêt :** Une analyse spatio-temporelle des images est réalisée pour détecter les éventuels objets d'intérêt. La partie spatiale, plus directement développée dans le cadre de cette thèse, est basée sur un filtre Laplacien adaptatif permettant la détection de leurs contours. Les deux itérations de ce filtre (de la version standard du filtre Laplacien et de la version adaptée à la détection de contours fins de petits objets) sont présentées dans la section 4.4.4. Ce procédé permet d'améliorer la netteté des contours et d'atténuer les « faux contours » (détectés par le filtre Laplacien standard et dû au phénomène de numérisation d'images dans des zones contenant des transitions entre deux couleurs).

Une mise en correspondance entre les régions détectées et le modèle d'objets permet de caractériser les régions en tant qu'objets d'intérêt auquel est associé un critère de fiabilité.

4) **Définition des critères de fiabilité :** Dans le cas de vidéos pour lesquelles le processus de détection peut s'avérer complexe (très petits objets, images bruitées, textures dans le fond de l'image proche de la texture/couleur théorique du modèle d'objets), le modèle d'objet d'intérêt correspond non seulement aux objets d'intérêt réellement présents dans la scène, mais également à de nombreuses fausses détections (des branches ou des feuilles d'arbres qui bougent avec le vent, des détections « parasites » des insectes, le bruit numérique...). Cette confusion apparait dans le cas, où les paramètres spatiaux des objets d'intérêt et des fausses détections correspondent au modèle d'objets. Dans ce cas les bonnes et les fausses détections sont caractérisées par des valeurs élevées de coefficients de fiabilité issus de l'analyse spatiale. Pour résoudre ce problème et supprimer des fausses détections restantes, une approche combinée, basée sur l'analyse spatial de l'apparence d'objets et l'analyse temporelle de leurs déplacements a été développée.

Les coefficients de fiabilité pour chaque objet détecté sont progressivement estimés à chaque étape d'algorithme pour prendre en considération l'évolution du modèle estimé et y adapter les valeurs de coefficients. Les coefficients obtenus à chaque étape sont convertis en un coefficient final qui caractérise la fiabilité de la détection d'un objet d'intérêt.

5) **Comptage des objets :** Finalement le comptage est réalisé en additionnant les coefficients de fiabilité globaux. Globalement cela a pour effet de réduire significativement le poids des fausses détections, même si certaines bonnes détections sont également sous-évaluées.

## 1. Etat de l'art

Ce chapitre se propose de décrire brièvement les principales méthodes existantes aujourd'hui pour la détection et le comptage de petits objets en mouvement dans une vidéo et d'analyser la possibilité de leur utilisation pour des applications telles que le suivi et le comptage des oiseaux migrateurs.

L'ornithologie est un domaine dans lequel de nombreux contenus vidéo sont utilisés. Pourtant, les possibilités d'utilisation de méthodes d'analyse basées sur les méthodes récentes du traitement des images et de la vidéo ont été peu étudiées jusqu'à présent. Ce contexte applicatif se rapproche sur le plan scientifique d'applications liées à la vidéosurveillance telles que le contrôle du trafic routier ou le dénombrement d'une foule. Dans ces contextes, les techniques de traitement d'images les plus largement utilisées sont liées à la détection de mouvement, la reconnaissance des formes et la caractérisation d'objets dans les vidéos. Ces types de techniques sont, à l'exception des techniques de reconnaissance de formes (car les oiseaux sont souvent trop petits dans les images), applicables dans le cadre du suivi de migrations d'oiseaux. Le suivi des migrations au moyen d'analyse d'images et de vidéos peut être considéré comme une application spécifique de vidéosurveillance, dont les méthodes peuvent servir pour les besoins d'ornithologie.

Dans l'ornithologie, le suivi de migrations saisonnières, le dénombrement des oiseaux-migrateurs pour des données statistiques ou la surveillance de colonies d'oiseaux sont réalisés le plus souvent à l'aide de méthodes traditionnelles basées soit sur un opérateur humain soit sur des dispositifs physiques telles que des bagues. Une limite évidente de ces méthodes physiques, qui motive les recherches pour la conception d'une méthode numérique du suivi, est l'impossibilité d'équiper un grand nombre d'oiseaux de ces outils physiques.

Les points essentiels du travail sur l'état de l'art, consacré aux études de méthodes existantes pour le suivi de migrations aux moyens physiques et analytiques, consistent à :

- étudier les moyens physiques (non-numériques) existants pour le suivi de migrations et identifier leurs désavantages.
- étudier des approches analytiques (numériques) existantes dans le domaine de l'analyse de vidéos et du traitement d'images appliquées pour des recherches ornithologiques et pour les besoins de la vidéosurveillance.
- identifier les limites des approches analytiques dans le cas de détection de très petits objets en mouvement.
- analyser des techniques de traitement d'images pour l'application sur les étapes de l'algorithme conçu et leurs adaptations pour le cas la détection et dénombrement de très petits objets en mouvement dans une vidéo (ici : des oiseaux-migrateurs).

Dans ce contexte, cet état de l'art se décompose de la manière suivante :

Les méthodes non-numériques, appliquées en ornithologie, sont d'abord présentées dans la section 2.1. Elles nécessitent des outils spéciaux, telles que les bagues et les émetteurs radio, portés par les oiseaux-migrateurs. Leurs limites et inconvénients, qui motivent l'application des méthodes numériques, sont décrits dans la même section, ainsi que les avantages de ces dernières par rapport aux méthodes traditionnelles.

La section 2.2 présente les principales méthodes numériques existantes dans le domaine du traitement d'images et destinées à la détection d'objets dans une image statique. L'analyse de techniques du traitement d'images est essentielle dans l'état de l'art, car ce domaine est étroitement lié avec le traitement de vidéos, qui se compose d'une séquence d'images, et l'objectif final des travaux des recherches – la détection et le comptage de très petits objets dans les vidéos.

La section 2.3 est consacrée aux méthodes numériques de détection et du suivi d'objets en mouvement dans les vidéos. La différence essentielle entre l'application de techniques du traitement d'images sur des images elles-mêmes et sur les vidéos est l'ajout de critères supplémentaires concernant le comportement dynamique des objets dans la vidéo. Dans la même section les interactions entre les critères de détection d'objets dans une image et une vidéo sont également étudiées dans le contexte de leur application et dans le contexte de la détection et le suivi des oiseaux migrateurs.

La section 2.4 analyse les notions de « petit objet » et « très petits objets » en termes de leur apparence dans les vidéos, ainsi que les difficultés qu'elles impliquent pour la détection d'objets dans des images. Le cas de la détection et le comptage des oiseaux-migrateurs en tant que des « petits » et « très petits » objets est également étudié dans cette section.

Enfin, la section 2.5 résume les particularités et les limitations dans le cas de détection de « petits » et de « très petits » objets et présente l'intérêt de la conception d'une approche de détection et de comptage d'objets de ce type, basée sur la notion de fiabilité de détection. La section 2.5 fait également le lien entre l'état de l'art dans les domaines de détection d'objets dans les images et les vidéos, et la solution proposée, qui permet de positionner son innovation parmi les travaux existants.

## 2.1 Méthodes non numériques pour le comptage de très petits objets en mouvement

Le dénombrement des oiseaux migrateurs est un problème ancien que n'a bien sûr pas été historiquement réalisé à l'aide de méthodes numériques. Les principales méthodes non numériques existantes de suivi d'oiseaux-migrateurs sont les suivantes :

- **Observation directe** : dénombrement des migrations à l'œil depuis des postes d'observation [1][5].

Le comptage manuel depuis un poste d'observation est une des méthodes physiques de suivi des migrations d'oiseaux. Elle est la plus simple à réaliser techniquement. Le dénombrement se fait à l'œil par des techniques de comptage rapide sur la base de photos. Typiquement, le vol entier est visuellement segmenté en zones contenant des parties du vol approximativement de la même densité. Le comptage est ensuite réalisé sur certaines zones et extrapolé sur l'ensemble de l'image. L'avantage de cette méthode est une certaine rapidité de dénombrement, mais elle possède plusieurs inconvénients qui limitent son efficacité et sa précision :

- Nécessite un personnel formé en comptage rapide
  - Coût de main-d'œuvre important : généralement plusieurs personnes doivent dénombrer le même vol.
  - Nécessité de déplacements réguliers sur les postes d'observation.
  - Précision hétérogène et approximative du dénombrement.
- **Observation au moyen d'un matériel spécifique.** Dans ce contexte, plusieurs approches sont possibles [7]:
    - **Baguage d'oiseaux** avec une bague métallique portant un identifiant [2].

Le baguage d'oiseaux est fréquemment utilisé aujourd'hui pour le suivi des migrations. Cette méthode permet d'obtenir des informations diverses nécessaires pour les recherches ornithologiques, tels que les déplacements des oiseaux, ses inconvénients en limitent l'application pour le dénombrement d'oiseaux migrateurs. Les principaux inconvénients sont les suivants :

- Nécessité de capturer des oiseaux pour leur fixer une bague ou pour enregistrer les informations souhaitées.
- Utilisation d'une main d'œuvre qualifiée générant des coûts importants.
- Cette méthode ne donne pas d'information sur la trajectoire de migration, mais seulement des points de passage.
- Le nombre d'oiseaux pouvant être surveillé est limité (car il est difficile de baguer un nombre statistiquement significatif d'oiseaux-migrateurs).
- Risque de pertes de bagues par les oiseaux.

- **Suivi à l'aide d'un radar des vols d'oiseaux [3] :**

Ce type de techniques permet de suivre les trajectoires des vols d'oiseaux relativement important du moins dans certaines zones, mais ne permet pas d'en effectuer le dénombrement.

- **Utilisation d'un émetteur radio pour la détection satellitaire [4][6] :**

Des émetteurs-radio peuvent être utilisés à la place d'une bague sur certains oiseaux. Ces émetteurs sont portés par les oiseaux et les signaux qu'ils transmettent sont captés par des satellites ou des radars. Elles permettent donc le suivi des trajectoires des vols d'oiseaux. Ces méthodes nécessitent néanmoins un matériel plus coûteux.

En résumé, ces techniques utilisant un matériel spécifique sont assez répandues et toujours utilisées aujourd'hui mais ont des limites communes pour le cas de suivi des grands vols des oiseaux migrateurs :

- Elles permettent de suivre des oiseaux isolés ou un faible nombre d'oiseaux par rapport au volume d'oiseaux migrateurs lors d'une campagne de migrations. Pour obtenir une efficacité et une précision de dénombrement suffisantes, un grand nombre d'oiseaux devraient être équipé d'une bague ou d'une balise, ce qui demande une quantité significative de main-d'œuvre, des matériaux et des ressources financières.
- Elles nécessitent un personnel qualifié en ornithologie.

Les travaux de recherches dans le cadre de la thèse ont pour but de faciliter le travail du personnel et automatiser le suivi des migrations en appliquant des méthodes de traitements d'images sur des vidéos avec des vols d'oiseaux.

## **2.2 Détection d'objets dans les images : un état de l'art**

Le suivi d'objets dans les vidéos est indissociable de la détection de ces mêmes objets dans les images successives de cette vidéo. Il est donc nécessaire dans le contexte applicatif visé dans le cadre de cette thèse d'être capable de détecter dans chaque image les objets d'intérêt.

Pour obtenir un tel résultat, plusieurs approches méthodologiques existent avec, pour chacune d'entre elles, de nombreuses variantes possibles. De plus le contexte applicatif influence la méthode utilisée et la manière avec laquelle elle est mise en œuvre. Les paragraphes suivants décrivent brièvement ces principales approches pour en montrer l'intérêt et les désavantages dans le cadre du contexte de la cette thèse.

### 2.2.1 Modélisation et soustraction d'arrière-plan

Pour détecter des objets au moyen d'une méthode de soustraction d'arrière-plan, il est impératif d'avoir en possession l'information sur le contenu d'une scène du fond et de le mettre à jour régulièrement, si chaque nouvelle image est extraite d'une vidéo. Les principales étapes d'une méthode de soustraction de l'arrière-plan sont les suivantes :

- Modélisation de l'arrière-plan de la scène
- Soustraction de valeurs (couleurs, intensités lumineuses), associées aux pixels correspondants de l'image d'arrière-plan et d'une image, où les objets sont recherchés.
- Génération d'un masque d'objets détectés dans une image par le critère suivant :

$$I(Obj)_j = \begin{cases} 1, & \text{si } BG_j - I_j \geq S \\ 0, & \text{sinon} \end{cases}$$

où,  $S$  est la différence minimale entre les pixels pour considérer une différence considérée comme significative entre les pixels d'une image et l'arrière-plan.  $BG$  et  $I$  sont respectivement les valeurs de pixels de l'arrière-plan et de l'image.

De nombreuses variables existent selon que l'on considère des différences par pixels ou par zones/régions, ou que l'on analyse la différence sur les trois composantes ou simplement sur la luminance. Ce type d'approche est très au bruit particulièrement dans les zones texturées.

### 2.2.2 Seuillage d'images

Le seuillage d'une image et ses variantes, adaptées aux applications concrètes, est couramment appliqué pour la détection d'objets en tant qu'un ensemble de pixels dans une image, différent des autres par un certain critère. Fixe ou adaptatif, un seuil est appliqué généralement sur des images en niveaux des gris pour générer un masque des objets d'intérêt  $I(Obj)$  depuis l'image initiale  $I$  en comparant ses pixels avec un seuil  $S$  :

$$I(Obj)_j = \begin{cases} 1, & \text{si } I_j \geq S \\ 0, & \text{si } I_j < S \end{cases}$$

où  $I_j$  et  $I(Obj)_j$  sont les pixels de ces deux images sur les mêmes positions. Typiquement une telle méthode de seuillage est utilisée sur une image issue de la différence entre une image et l'arrière-plan de la scène (voir la section 2.2.1).

### 2.2.3 Détection de contours d'objets

Le principe de la détection des contours est de localiser les discontinuités dans une image, qui séparent des régions différentes [37][38]. Dans les images ces discontinuités représentent des forts changements de l'intensité lumineuse, qui correspondent aux contours d'objets – les régions qui les délimitent dans une image.

Les opérateurs de détections de contours se classifient en deux groupes [34] :

1. Opérateurs basés sur la dérivée première (le gradient), telle que l'opérateur de Sobel ou de Prewitt
2. Opérateurs basés sur la dérivée seconde, tels que l'opérateur Laplacien [22]



Pour les opérateurs, basés sur la dérivée première, l'indicateur de détection d'un contour est la magnitude du gradient qui s'exprime par la formule analytique :

$$|G| = \sqrt{G_x^2 + G_y^2}$$

où  $|G|$  signifie la magnitude du gradient  $G$ ,  $G_x$  – la valeur du gradient en direction horizontale (sur l'axe X) et  $G_y$  – la valeur du gradient en direction verticale (sur l'axe X).

Dans le cas discret du traitement d'images deux matrices de convolution (pour les directions sur les axes X et Y) correspondantes à un opérateur sont appliquées. Par exemples, les matrices pour l'opérateur de Sobel, respectivement pour les directions X et Y sont :

$$\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \text{ et } \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$$

Les opérateurs basés sur la dérivée seconde cherchent des passages par zéro pour identifier les locations de contours. Analytiquement ils s'expriment en formule suivant :

$$L(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$

et numériquement en matrice de convolution la plus utilisée pour l'exemple de l'opérateur Laplacien :

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{pmatrix} \text{ ou } \begin{pmatrix} 0 & -1 & 0 \\ -1 & -4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

La détection de contours sert d'un point de départ pour la détection d'objets, dont les formes sont définies par leurs contours [35]. Par exemple, l'approche dans [12] analyse aussi la direction du gradient pour modéliser la forme d'un objet et la comparer avec la forme d'un humain dans la vidéo.

Pour que les formes puissent être modélisées et identifiables pour une détection d'objets fiable, les contours devraient être nets, et l'image peu, ou pas bruitée. Ces conditions ne sont pas remplies que dans des vidéos d'une haute qualité et avec des objets visibles dans les vidéos (i.e. avec un bon contraste sur le fond), ce qui n'est pas le cas pour les vidéos fournies dans le cas d'application sur le suivi de migrations.

## 2.2.4 Modélisation d'objets

L'objectif d'un modèle d'objets consiste à définir des descripteurs capables de caractériser les objets d'intérêt à partir de descripteurs identifiables dans les images. Ces descripteurs peuvent être basés sur des caractéristiques :

- Géométriques (formes, points caractéristiques, dimensions...),
- Texturelles (couleurs dominantes, modèles de textures, histogramme [12] [13] [44]...)
- Dynamiques (vitesse, trajectoire, déformations...), lorsque la dimension temporelle est prise en compte.

L'algorithme de l'article [17], dont le sujet est la détection et la reconnaissance des espèces d'oiseaux, débute par la détection de ces contours et analyse les points extrêmes des régions englobées

pour appliquer la transformée de Fourier et évaluer le taux de similarité avec des modèles d'espèces, décrites par les paramètres du même type.

De même que la forme d'objet, sa texture – sa couleur et, éventuellement, son motif – se positionnent comme les descripteurs essentiels pour les modèles de référence et permettent une distinction par le critère de couleur. Les histogrammes d'images en couleur, basés sur les valeurs des composantes-couleur dans des différents espaces colorimétriques et ou les histogrammes, générés depuis les descripteurs, qui encodent les textures – « Local Binary Patterns » (LBP) [28] servent à la modélisation d'objets par l'aspect de couleur et leur détection dans un environnement complexe au moyen de comparaison des distances entre histogrammes [27].

L'efficacité des méthodes de trois groupes, présentées ci-dessus dépend des tailles d'objets dans une vidéo et de la qualité de ces textures. Elles influencent directement sur les descripteurs numériques associés aux objets et comparés à ceux du modèle et ainsi sur l'exactitude de la modélisation d'objets. Plus la taille d'objet est petit, plus il y a des risques de fausses alarmes – des fausses détections d'objets, correspondants aux modèles, mais d'autre nature que les objets d'intérêt. Le suivi d'objets dans une vidéo dépend alors de l'efficacité de détection dans chaque image de la séquence, qui compose une vidéo. Les méthodes existantes pour le suivi d'objets dans les vidéos et leurs connexions avec les méthodes de détection d'objets sont présentées dans la section 2.3 de l'état de l'art.

### **2.3. Détection d'objets en mouvement dans les vidéos : un état de l'art**

Les modifications entre les images successives d'une vidéo sont essentiellement dues aux mouvements des objets dans la scène et/ou au mouvement de la caméra. Ils proviennent également des modifications de la forme des objets, tels que les mouvements propres des bras et des jambes [14], des battements d'ailes d'un oiseau, ou des mouvements de la végétation dus au vent, des mouvements des nuages, des variations d'illumination ou encore du bruit d'acquisition. Un critère temporel permettant de caractériser les variations temporelles dues aux objets en mouvement des autres causes est donc crucial dans toutes méthodes de détection d'objet en mouvement. Les principales approches existantes pour la détection d'objets en mouvement dans les vidéos sont présentées dans les paragraphes suivants.

#### **2.3.1 Suivi de formes d'objets similaires**

Associés à des formes géométriques simples ou des courbes, qui modélisent les contours d'objets [18] les objets sont suivis au moyen de recherches dans une séquence d'images des régions similaires aux formes géométriques connues. Pour pouvoir appliquer les méthodes de ce groupe, il est impératif de disposer des informations sur l'ensemble d'objets, qui permettront d'associer ces régions dans le temps et de les suivre au cours d'une vidéo.

#### **2.3.2 Différences entre les images consécutives**

C'est une des méthodes de détection d'objets en mouvement les plus simples [19]. Leur principe de détection des mouvements est la soustraction des pixels correspondants dans les images consécutives pour y identifier des changements du moment  $t-1$  au moment  $t$ . De même que dans la soustraction d'arrière-plan pour la détection d'objets, les pixels dont la différence est supérieure au seuil introduit signifie des variations temporelles et ainsi des mouvements d'objets. La génération des masques des régions pour chaque paire consécutive de  $t-1$  à  $t$  permettra de suivre les trajectoires d'objets détectés.

#### **2.3.3 Modélisation d'un mélange Gaussien**

Le fait qu'une image représente un ensemble des pixels, avec des caractéristiques numériques (couleur, intensité lumineuse), permet d'analyser leurs distributions au sein d'une image. Dans une

image ces pixels peuvent être distribués dans au moins deux ensembles : ceux du fond (arrière-plan) et ceux, correspondant à des objets sur le fond. Or, une scène réelle généralement contient des objets des types variés et un arrière-plan hétérogène. Par conséquent, il peut y avoir plusieurs distributions de valeurs, associées à ces pixels. Des méthodes, basées sur la construction des mélanges Gaussiens, tel que [20], sont conçues pour détecter des différences de ces distributions, qui correspondent à des changements spatial ou temporels d'un frame d'une vidéo à l'autre.

### **2.3.4 Correspondance des points caractéristiques (« points d'intérêt »)**

Les algorithmes SIFT et SURF sont couramment appliqués dans des applications de reconnaissance d'objets dans des images par leurs caractéristiques, la modélisation 3D [23], l'assemblage d'images et dans le suivi d'objets dans les vidéos par la connexion de leurs points d'intérêt, dont les caractéristiques sont invariantes dans une vidéo [30]. Les déplacements de ces points d'intérêt correspondront aux déplacements d'objets d'intérêt.

### **2.3.5 Analyse des histogrammes**

La détection et le suivi des mouvements au moyen d'analyse des histogrammes considèrent une segmentation préalable d'images consécutives en blocs locaux [27][29]. Les différences entre histogrammes locaux, estimés pour les blocs sur les mêmes positions à chaque moment  $t$ , décrivent les différences, détectées dans ces zones au cours du temps. Pour générer les histogrammes plusieurs solutions sont possibles :

- Convertir une image en niveau de gris et utiliser les intensités lumineuses en tant qu'échantillons.
- Créer des histogrammes de couleurs ou en niveau de gris [49] [50].
- Analyser les textures dans les blocs et des méthodes d'encodage, pour décrire leurs contenus [28].

Ces histogrammes représentent les distributions dans des zones d'images, et les distributions similaires sont considérés comme correspondantes aux mêmes objets dans des images consécutives.

### **2.3.6 Flux optique**

Le flux optique est une approche qui détecte tous les changements temporels dans une séquence de trames d'une vidéo par l'analyse des pixels individuellement [21][8]. Le principe du flux optique est l'estimation du champ vectoriel en termes de vitesse de déplacements de ces pixels entre l'image  $t-1$  à l'image  $t$  [9]. Grâce aux analyses de chaque pixel individuellement, des moindres mouvements et des très petits objets peuvent être détectés par le flux optique sous conditions que la caméra conserve l'angle de vue de l'objet suivi et la luminosité ne change pas au cours d'une vidéo.

Ces deux vastes domaines scientifiques – de la détection d'objets dans une image, i.e. dans un contexte statique, et le suivi de leurs déplacements, i.e. dans un contexte dynamique – sont toujours associées [32]. Pour la détection d'objets d'intérêt dans une vidéo, qui par définition est composée d'une séquence d'images, l'étape de détection d'objets dans chaque frame est essentielle est repose sur leur forme apparente et leur taille. L'interaction entre les méthodes de ces deux domaines en fonction de tailles d'objets est détaillée ci-dessous, dans la section 2.4.

## **2.4 Détection des très petits objets dans l'état de l'art**

### 2.4.1 Hypothèses sur la taille et le modèle d'objets

La plupart des méthodes de détection d'objets présentées ci-dessus suppose qu'il est possible d'associer la forme d'un objet dans l'image à la forme de référence ou d'identifier et définir les correspondances de leurs points caractéristiques. La notion de la représentation d'objets dans les vidéos est alors essentielle pour le choix d'une méthode du traitement.

En fonction des dimensions apparentes des objets dans les images, trois situations principales peuvent être considérées :

- **Objets de taille normale :** Il s'agit d'objets dont la géométrie est clairement visible et identifiable dans les images. Il est possible de définir et modéliser leurs formes et d'identifier leurs points caractéristiques.
- **Objets de petites dimensions :** Il s'agit d'objets pour lesquels l'identification des points caractéristiques de forme n'est pas possible. Seule la dimension de l'objet est identifiable et devient un paramètre essentiel. Eventuellement elle permet aussi d'identifier une forme générique (e.g. approximée à une primitive géométrique) [15].
- **Objets de très petites dimensions :** Dans ce cas toute identification de forme -même approximative- devient impossible. La réduction de la taille des objets dans une vidéo, et l'augmentation de la distance entre les objets et la caméra atténue l'intensité du contraste de la couleur d'objet et celle du fond. Dans ces conditions la détection des contours, qui servent à délimiter un objet et à définir sa forme devient impossible. Cette configuration se retrouve dans certaines contenant des vols d'oiseaux migratoires.

Le suivi d'objets en mouvement dans les vidéos dépend de la précision de leurs détections dans chaque image de sa séquence. Certaines méthodes de suivi d'objets de ces différents groupes, pour l'application de détection des oiseaux, existent dans la littérature :

### 2.4.2 Méthodes existantes de détection d'oiseaux d'une taille normale dans le contexte de la détection d'oiseaux

Les articles [15], [16], [17] se servent des caractéristiques de la forme apparente des oiseaux, dont la modélisation n'est possible qu'en cas de leur taille suffisante : l'envergure des ailes pour la reconnaissance des espèces d'oiseaux [17], la fréquence de battement des ailes [16] ou l'analyse de courbe de contours et un filtre par taille [15]. Dans [16] un modèle temporel de battement des ailes est appliqué pour identifier les espèces dans une vidéo. Les objectifs de ces travaux sont également la détection et la reconnaissance des espèces précises d'oiseaux, qui n'est possible qu'à partir des caractéristiques déterminantes.

L'utilisation de ces méthodes n'est pas possible dans les vidéos dans lesquelles la géométrie des objets ne peut pas être définie d'une manière suffisamment discriminante.

### 2.4.3 Méthodes existantes de détection de petits objets dans le contexte de détection des oiseaux

La détection d'objets dont la taille ne permet que d'obtenir une représentation approximative de la forme apparente d'un objet détecté, est étudiée dans [24] et [31]. Les méthodes proposées fonctionnent dans le contexte prédéfini par l'environnement dans la scène d'une vidéo. Un petit objet – une balle de football – est détecté est suivi uniquement sur le terrain de jeu [24], ou repéré au moyen de la segmentation hiérarchique [31], qu'introduit aussi un contexte de recherches de petits objets.

#### **2.4.4 Méthodes existantes pour la détection de très petits objets pour le suivi d'oiseaux migrants**

Quant aux très petits objets, où les critères les discriminants dans une vidéo sont très limités, leur distinction dans un contexte prédéfini devient peu efficace, car la forme apparente, ainsi que les tailles de très petits objets d'intérêt, se confondent dans l'ensemble de détections, y compris des « objets-parasites » et le bruit numérique.

Le suivi et le comptage des oiseaux migratoires se positionnent notamment dans ce cas particulier de détection de très petits objets, où que très peu d'information sur leur forme apparente dans une vidéo n'est disponible. Les méthodes classiques du traitement d'images n'y sont pas applicables sans modifications.

Les algorithmes des travaux présentés ci-dessus ne sont pas applicables dans le cas de suivi d'oiseaux migrants pour plusieurs raisons :

- Une très petite taille d'objets d'intérêt limite la modélisation de leur forme (même approximativement) [15].
- Sans connaître la forme approximative d'un objet ces points caractéristiques ne peuvent pas être extraits [16].
- La détection dans un contexte prédéfini [24] est aussi limitée, car généralement les vidéos avec des grands vols des oiseaux migratoires contiennent un grand nombre des très petits objets d'intérêt et des « objets-parasites » dans le même contexte (dans le ciel). Leurs formes apparentes se rassemblent dans le même contexte, ce qui complique leur distinction, contrairement au cas de suivi de balle [24].

#### **2.5 Conclusion sur l'état de l'art dans les domaines de détection et de suivi d'objets dans les vidéos**

Les travaux existants dans l'état de l'art actuel sont souvent basés sur la sélection de zones, conformes aux modèles attendus dans leurs cas d'application, où ils sont déterminés par des descripteurs, propres aux objets d'intérêt, et où ces derniers peuvent se localiser (des points caractéristiques pour la forme d'objet d'intérêt ou la texture pour une zone d'intérêt). Or, dans les vidéos avec des vols d'oiseaux-migrants la modélisation de leur forme apparente dans une vidéo n'est pas possible à cause d'une très petite taille (souvent réduite jusqu'à quelques pixels seulement), et théoriquement, un grand nombre des zones dans une vidéo peut correspondre à des régions avec des objets d'intérêt en raison de manque de distinction entre les « objets-parasites », le bruit et les objets d'intérêt. La détection et le suivi de très petits objets en mouvement et son cas d'étude pour le suivi et le comptage des oiseaux-migrants se positionnent alors dans une application particulière et peu étudiée dans les travaux existants.

Les vidéos avec des oiseaux migrants sont enregistrées depuis des postes d'observation à des distances de 500m – 750m de la caméra, d'où une altération significative de la taille et de la forme apparente d'objets d'intérêt dans une vidéo de ce type. En outre, les variations locales d'intensité lumineuse ou du contenu dans des zones avec les objets d'intérêt affectent la perception de leurs formes qui peut être variable et irrégulière (altérée à cause d'un contraste insuffisant ou d'une taille variable à cause d'objets superposés). Sans aucun modèle de référence pour la détection de très petits objets en mouvement dans les vidéos (dont le cas d'application pour le suivi de migrations fait partie), elle est exposée à un risque élevé de fausses détections.

### 2.5.1. La possibilité d'application des méthodes existantes pour le suivi et le dénombrement des oiseaux-migrateurs

Certaines techniques existent pour la détection de petits objets dans les vidéos. Mais ces méthodes ne conviennent pas dans le cas d'application pour suivi de migrations au moyen de traitement de vidéos pour plusieurs raisons.

Les vidéos avec le jeu de football [24], enregistrées à une grande distance, peuvent se classer en tant que « des vidéos avec des petits objets d'intérêt », car l'objet traqué (la balle) peut se distinguer dans le contexte prédéfini du terrain de jeu. Contrairement aux vidéos avec des vols d'oiseaux, où les objets d'intérêt sont caractérisés comme « des très petits » et donc peuvent être similaires aux zones avec des fausses alarmes, la détection de la balle est possible au moyen de comparaison de la similarité de zones qui la contiennent. Par contre, dans le cas étudié de suivi de migrations le points d'observation sont plus éloignés d'objets observés, ce qui pose des limites sur les méthodes de comparaison de zones.

La méthode, proposée dans l'article [20], construit les mélanges Gaussiens pour les pixels dans les zones, associées dans la séquence d'images d'une vidéo, qui sert à estimer et suivre les variations entre les pixels dans le temps dans les mêmes zones. En termes des tailles d'objets, qui affectent la qualité de détections (voir la section 4.3.5), ces variations peuvent correspondre à des très petits objets en mouvement. Les particularités de l'apparence des très petits objets dans les vidéos (une très petite taille, un faible contraste sur le fond et une couleur, proche à celle du fond, section 2.4.1) imposent une valeur basse pour le seuil qui définit, s'il y a eu de variations dans un mélange Gaussien du moment  $t-1$  au moment  $t$ . Ainsi la sensibilité aux faibles variations dans les modèles des zones des pixels a mené à des nombreuses fausses détections, correspondantes pour la plupart au bruit numérique et à des transitions des couleurs dans des zones visuellement homogènes, qu'aux objets d'intérêt (un exemple d'un masque d'objets en mouvement, généré par la méthode de [20] est illustré sur les figures 2.1 et 2.2). D'autre part, la méthode de [20], est destinée à l'exécution sur les smartphones, où les tailles d'images sont inférieures et les résolutions sur plus faibles, que celles dans le cas des vidéos HD. Comme conséquent, le temps d'exécution et la complexité de l'algorithme de [20], appliqué sur les vidéos fournies, n'a pas été acceptable.

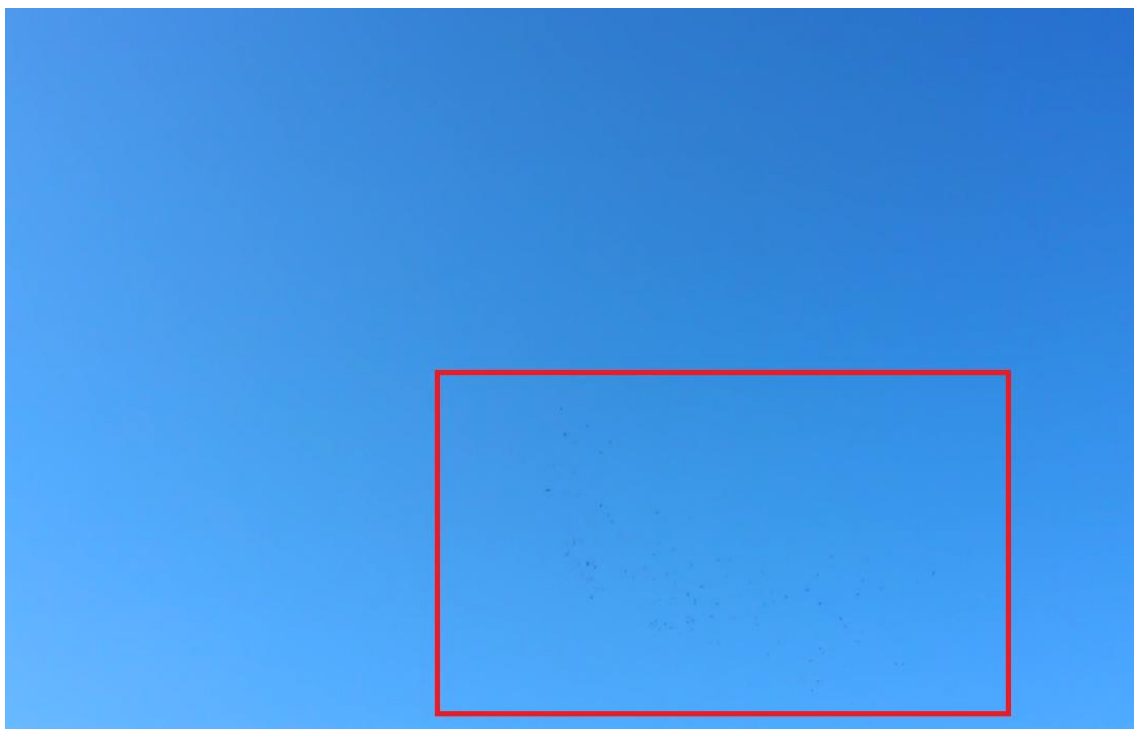


Figure 2.1 : Une image extraite d'une vidéo avec un vol d'oiseaux éloigné de la caméra et des objets avec un faible contraste sur le fond (le vol est encadré dans le carrée rouge)



Figure 2.2 : résultat d'application de la méthode [20]. Différences de l'intensité lumineuse produisent des fausses détections à gauche, en haut de l'image. En supposant que chaque région blanche dans les images générées est un objet d'intérêt potentiel, le dénombrement direct par la méthode de composantes connexes calcule tous les objets. Les objets d'intérêt sont alors confondus avec du bruit dans des images binaires générées par la même méthode pour représenter les objets en mouvement d'un instant à l'autre.

### 2.5.2. Intérêt de la détection et du dénombrement en fonction de la fiabilité de détection

Repérer et compter de très petits objets dans une vidéo à l'œil n'est pas une tâche facile, particulièrement en présence de très nombreux objets. La détection automatique de tels objets dans les vidéos est également une tâche complexe. En effet, si la vue humaine saura, par exemple, reconnaître immédiatement des vols éloignés constitués des groupes de plusieurs espèces d'oiseaux et les identifier en tant que des « objets d'intérêt », les méthodes numériques ne sont pas toujours aussi efficaces. En résumé, généralement, les très petits objets dans les vidéos sont caractérisées par les points suivants, qui compliquent le traitement et motive le développement d'une méthode novatrice pour leurs détection efficace :

- Un faible contraste entre le fond de la scène et les petits objets rend difficile, de les distinguer sans générer de nombreuses fausses détections.
- Une très petite taille d'objets qui ne permet pas de modéliser leurs formes apparentes
- Un arrière-plan texturé d'une scène complique également la détection des petits objets, qui peuvent facilement être confondu avec le fond de la scène.
- Les teintes des couleurs d'objets éloignés de la caméra sont proches à celle de la couleur du fond, et se distinguent uniquement par l'intensité des couleurs. Si les contraste d'objets est faible, les repérer sans confondre avec le bruit n'est pas possible.

Les méthodes numériques existantes dans la littérature, où la détection d'objets est conçue pour les catégories d'objets « d'une taille normale » ou des « petits objets », ont démontré les difficultés du cas particulier de « très petit objets » à cause de limites de modélisation d'une structure complète qui

permettrait sélectionner dans l'ensemble de détections similaires les objets d'intérêt. La solution, proposée dans cette thèse, se base alors non sur l'aspect quantitatif de détections (i.e. le nombre d'objets, dont les représentations dans une vidéo répondent à certains paramètres, tels que la taille ou la forme), mais sur l'aspect qualitatif (i.e. à quel point l'apparence d'un objet ressemble à l'apparence attendue pour un objet d'intérêt). Cet aspect qualitatif, fondamental dans l'algorithme implémenté, s'exprime dans des critères de fiabilité, appelés dans la suite du document « les coefficients de fiabilité », qui caractérisent un objet en tant qu'un objet d'intérêt ou une fausse alarme. Ce coefficient, compris dans l'intervalle  $[0, 1]$  est utilisé directement dans le comptage classique d'objets un par un et permet de supprimer implicitement les détections peu ou pas fiables. Son concept est détaillé dans la section 4.2.1.

Le besoin de l'Association de Chasseurs de la Dordogne (GIFS) a servi de cas d'études pour les travaux des recherches, effectuées dans le cadre de la thèse. Sous le facteur d'une grande distance, les paramètres d'objets suivis dans les vidéos fournies répondaient aux caractéristiques attendues pour l'apparence de très petits objets dans une vidéo.

### **3. Présentation du projet Octobre Bleu**

Le projet Octobre Bleu est un projet de recherche et développement qui se propose de développer une bibliothèque d'algorithmes de traitement d'images et de vidéos pour des applications diverses (des recherches ornithologiques, la vidéosurveillance, le contrôle du trafic etc.). Les partenaires du projet Octobre Bleu sont les suivants : Sogeti, porteur du projet auprès de la région aquitaine, le Labri (université de Bordeaux), et le GIFS (Groupe d'Investigations sur la Faune Sauvage). Le LaBRI est responsable de la partie recherche du projet. Sogeti est en charge de l'ensemble des aspects ingénierie du projet. Le GIFS intervient comme client pour la partie détection et comptage des oiseaux migrateurs et a fournis le corpus de données vidéo contenant les vols d'oiseaux migrateurs.

Les travaux de recherches dans le cadre du projet Octobre Bleu ont été initialisés par le besoin du GIFS (Groupe d'Investigations sur la Faune Sauvage) suite à son besoin d'automatiser le suivi de migrations du pigeon ramier (palombe ou *Columba palumbus*) dans le Sud-Ouest de la France. L'objectif de fonctionnalités du projet pour cette application est de permettre à un utilisateur – observateur ou ornithologue – de filmer avec un support accessible (caméra, tablette ou smartphone) des flux migratoires et procéder à leur traitement automatique par le système développé par le projet Octobre Bleu sur le serveur pour obtenir le nombre d'oiseaux dans les flux. Cette application a pour vocation à être généralisée pour la détection et le dénombrement d'autres espèces d'oiseaux, d'animaux ou d'insectes, ainsi que d'autres objets en mouvement dans les vidéos.

#### **3.1 Objectifs globaux du projet Octobre Bleu**

L'objectif global du projet consiste dans le développement d'une bibliothèque d'algorithmes de traitement d'images pour la détection et le dénombrement d'objets en mouvement variés dans des vidéos. Cette bibliothèque est destinée à de nombreux cadres applicatifs d'analyse des vidéos : vidéosurveillance dans des lieux publiques, analyse d'affluence dans des transports en commun, analyse de trafic routier, pour en citer quelques-unes dans le milieu urbain. La détection et le dénombrement d'objets en mouvement trouvent leurs applications également dans des recherches scientifiques, telles que le suivi de migration d'oiseaux ou de population des insectes. Les algorithmes du projet Octobre Bleu sont aussi destinés à ces domaines.

#### **3.2 Positionnement des travaux de recherche dans le projet**

Les travaux abordés dans le cadre de la thèse sont centrés sur l'application du projet Octobre Bleu pour les recherches ornithologiques, définies par GIFS, i.e. pour le suivi de migrations de la palombe au Sud-Ouest. Les objectifs du projet sont toutefois plus vastes et comprennent des applications dans des autres domaines, que le suivi de migrations. Afin de permettre ces futures extensions et des



adaptations pour des autres cas d'application, Octobre Bleu devrait fournir des solutions génériques. Dans ce contexte trois axes de recherches ont été abordés lors de travaux dans le cadre de la thèse :

*Premier axe de recherche* - Sélection de zones d'intérêt dans une vidéo :

- identification de régions dans une vidéo, susceptibles de contenir des objets en mouvement, indépendamment du contexte applicatif.

*Deuxième axe de recherche* - Conception du modèle d'objets d'intérêt :

- Estimation et modélisation des caractéristiques attendues pour les objets d'intérêt en fonction du contexte d'application et des conditions dans une vidéo. Par exemple, dans le cas de suivi des oiseaux-migrateurs, la distance entre les objets d'intérêt et la caméra aura un effet sur leur forme apparente dans une vidéo. Des modèles différents d'objets d'intérêt seront alors nécessaires pour pouvoir les distinguer parmi les autres objets présents dans une scène filmée.

*Troisième axe de recherche* – Choix d'une méthode de dénombrement d'objets détectés :

- En fonction du nombre d'objets filmés (et la densité du vol pour le cas des flux migratoires)
- En fonction du taux approximatif de fausses détections potentielles.

Le concept principal de l'algorithme de la bibliothèque Octobre Bleu se concentre sur une solution générale destinée à être utilisée pour des applications variées. La solution proposée devrait s'y adapter sans en modifier les étapes fondamentales. La séquence algorithmique conçue pour le suivi de migrations pourrait être modifiée en une séquence d'analyse d'affluence dans le transport commun ou en séquence de détection et dénombrement de véhicules sur une autoroute (en changeant le modèle d'oiseaux au modèle de voyageur ou de voiture).

Ainsi l'objectif des travaux de recherche était d'implémenter ces fonctionnalités du projet Octobre Bleu et de proposer une solution, basée sur des techniques de traitement d'images, paramétrable pour des contextes d'application et capable de s'adapter à des conditions dans des vidéos fournies. Les travaux réalisés dans le cadre de cette thèse se focalise sur la détection et le comptage des oiseaux migrateurs, plus particulièrement concernant les phases d'analyse spatiales.

### **3.3 Présentation des corpus vidéo**

#### **3.3.1 Standardisation de caractéristiques des vidéos**

Comme il est mentionné dans la section précédente, une augmentation de la distance entre les objets et la caméra résulte en une réduction de la dimension apparente des objets d'intérêt. Pour cette raison il devient très compliqué, voire impossible, de distinguer à une grande distance les objets d'intérêt d'autres objets, présents dans une scène ou même de les distinguer du fond de la scène. En pratique, la difficulté de la détection dépend de plusieurs facteurs liés tout d'abord à cette distance, mais aussi aux caractéristiques dynamiques et texturales du fond de la scène sur laquelle se trouve l'objet. Il est alors intéressant de définir différentes classes de vidéos en fonction de ces caractéristiques. Le corpus de vidéos, fournies par le GIFS, a été classé en fonction des caractéristiques suivantes :

*1. Type de prise de vue et stabilité d'une vidéo :*

- Caméra fixe (installée sur un point d'observation) : prise de vue stable, absence de mouvement due à la caméra.
- Caméra globalement fixe mais peu stable : Ce manque de stabilité peut se produire à cause du vent (caméra positionnée sur un support sensible au vent, caméra portée par un opérateur immobile). Cela génère des tremblements / légers mouvements de la caméra dans les vidéos.

- Caméra en mouvement (suivi des déplacements d'objets d'intérêt) : Existence d'un mouvement important et continu entre la scène et la caméra, et donc d'un mouvement apparent dans toute la vidéo.
- Distance entre la caméra et les objets d'intérêt observés :
  - Grande (500-750m) : le cas le plus fréquent pour les vidéos enregistrées depuis des postes d'observation
  - Proche : vidéos ont été enregistrées avec un smartphone dans des conditions spontanées (i.g. avec des vols qui surviennent dans les champs)

## 2. Type et complexité d'arrière-plan :

- Arrière-plan peu ou pas texturé : typiquement il s'agit du ciel d'une couleur unie (e.g. bleu clair ou gris voilé) ou du ciel en plusieurs couleurs proches au niveau de leur teinte (avec des nuages, au coucher du soleil).
- Arrière-plan texturé : présence ou non d'objets immobiles sur l'arrière-plan (végétation, forêt, etc.).
- Présence d'objets en mouvement différents des objets d'intérêt. Il peut s'agir de très petits objets tels que des insectes, ou même de la poussière qui peuvent avoir des tailles apparentes dans l'image de l'ordre de celle des objets d'intérêt.

## 3. Caractéristiques d'objets en mouvement :

- Si les objets d'intérêt se déplacent en groupes, la *densité* de groupes peut être caractérisée :
  - Objets proches les uns des autres (e.g. un vol groupé d'oiseaux, probablement avec des chevauchements). Ils forment alors un groupe d'objets.
  - Objets espacés (i.e. éloignement géographique dans l'image et/ou trajectoires différentes).
  - Présence de plusieurs groupes d'objets espacés.
- *Contraste* entre le fond de la scène et la couleur dominante des objets d'intérêt. :
  - Faible contraste entre le fond et la couleur dominante des objets d'intérêt : fort risque de nombreuses fausses détections.
  - Fort contraste entre le fond et la couleur dominante des objets d'intérêt : faible risque de fausses détections.
- *Trajectoires et vitesses* apparentes de déplacement d'objets :
  - Vitesse lente, notamment lorsque la distance objet/caméra est grande

- Déplacement rapide d'objets (essentiellement s'ils sont proches de la caméra)
- Type de mouvement : erratique ou linéaire

#### 4. Conditions météorologiques, schématiquement deux situations peuvent être retenues :

- *Calmes/beau temps* : Absence de mouvements parasites du au vent susceptible de faire bouger des objets fixes sur le fond (principalement la végétation) ou même la caméra. Le fond de la scène est de ce fait davantage fixe ce qui réduit les risques de fausses détections.
- *Vents/pluies* : Le vent et la pluie génèrent des mouvements locaux -essentiellement au niveau de la végétation- d'objets susceptibles de créer de nombreuses fausses détections

#### 5. « Objets parasites » dans la prise de vue ou leur absence

Les combinaisons des différents critères présentés ci-dessus caractérisent les difficultés de comptage des objets d'intérêt. Une classification et une standardisation des vidéos par ces paramètres permet donc de concevoir pour chaque classe une combinaison de traitements adaptés. D'une manière globale les vidéos ont été réparties en quatre corpus, représentant leurs niveaux de complexité (voir le tableau 3.1).

##### 3.3.1.1 Caractérisation de la difficulté de traitement d'une vidéo

Selon les interactions des caractéristiques d'une vidéo donnée, le paramétrage et l'algorithmique elle-même doivent souvent être modifiés. Pour pouvoir obtenir cela d'une manière automatisée, il est intéressant de pouvoir extraire des catégories de vidéos en fonction de ses caractéristiques essentielles ayant un impact direct sur la facilité de détection et de comptage des objets d'intérêt. Pour cela, nous avons pris en compte les éléments suivants :

- Caractéristiques de texture et de mouvement d'objets dans une scène (autres que les objets d'intérêt) contribuent à des variations du fond de la scène.
- Caractéristiques de taille (dimension apparente) et de texture des objets d'intérêt dans la vidéo.
- Caractéristiques et similarité d'apparence d'objets d'intérêt avec d'autres objets dans une scène

Dans des vidéos réelles toutes les combinaisons des caractéristiques ci-dessus sont possibles. Pour des raisons de simplicité de classification pour l'utilisateur non spécialiste, seul les quatre classes répertoriées dans le tableau 3.1 sont utilisées. En pratique, un utilisateur non spécialiste répondra aux questions suivantes :

- Vidéo acquise avec une caméra fixe, tremblante ou en mouvement ?
- Arrière-plan semble d'être uniforme ou hétérogène ?
- Objets d'intérêt petits ou très petits ?
- Couleur des objets d'intérêt distinguable du fond de la scène ?

Les réponses à ces questions permettent de classer chaque vidéo selon une des quatre catégories mentionnées dans le tableau 3.1.

Tableau 3.1: Répartition du premier ensemble de vidéos en quatre complexités – corpus (illustrations sont données sans les figures 3.1 – 3.4).

<b>Corpus</b>	<b>Mouvement de l'arrière-plan</b>	<b>Texture de l'arrière-plan</b>	<b>Caractéristiques couleur des objets en mouvement</b>	<b>Caractéristiques couleur de taille des objets en mouvement</b>
1	Aucun (fond fixe)	Uniforme	Couleur dominante des objets d'intérêt différentes du fond de la scène sur lequel ils se trouvent	Objets des tailles petites ou moyennes
2	Mouvement du fond très faible (due notamment à un mouvement de tremblement de la caméra)	Faiblement texturé (par exemple zones nuageuses)	Couleur dominante des objets d'intérêt proches du fond de la scène sur lequel ils se trouvent	Objets des tailles petites ou moyenne
3	Présence d'objets en mouvement autre que les objets d'intérêt	Fond partiellement ou totalement texturé	Couleur dominante des objets d'intérêt différentes du fond de la scène sur lequel ils se trouvent	Objets de très petites tailles.
4	Fond de la scène en mouvement (Caméra en mouvement)	Fond partiellement ou totalement texturé	Couleur dominante des objets d'intérêt proches du fond de la scène sur lequel ils se trouvent	Objets de très petites tailles.

La classification présentée ci-dessus introduit l'attribution d'un niveau général de complexité pour une vidéo en fonction de son contenu. Or, une vidéo classée dans le corpus « simple » se retrouve plus difficile à traiter car ces caractéristiques qualitatives (e.g. netteté ou la taille d'objets d'intérêt) compliquent la détection d'objets d'intérêt sans un certain nombre des fausses détections.

Ainsi, une vidéo avec un fond homogène du ciel bleu, mais avec des objets d'une petite taille, flous et difficilement visibles même sur un fond uni nécessitent des ajustements des paramètres d'algorithme pour éviter une perte considérable d'objets d'intérêt. Ces valeurs seront suffisantes pour détecter des faibles changements dans des zones d'intérêt, mais elles entraînent aussi la détection du bruit (présent surtout dans des vidéos de basse qualité).

Les deux figures ci-dessous illustrent que les vidéos avec un bon contraste d'objets d'intérêt sur le fond seront plus simples à traiter (figure 3.1). Cependant, certaines vidéos du corpus 2, comme celui sur la figure 3.2, pourraient logiquement être classées en corpus 1.

Au contraire, la vidéo, dont une image est illustrée sur la figure 3.5, est un exemple d'un cas, ou une vidéo avec les caractéristiques du corpus 1 (le fond homogène d'une couleur uniforme) aura un niveau de difficulté de traitement élevée en raison d'un faible contraste d'objets sur le fond. Souvent dans des vidéos de ce type il est difficile de distinguer les objets d'intérêt même à l'œil.



Figure 3.1 : Corpus 1 - un fond uni, petits objets d'intérêt à une couleur différente du fond



Figure 3.2 : Corpus 2 - un fond légèrement texturé, la moitié de très petits objets d'intérêt ont une couleur proche du fond



Figure 3.3 : Corpus 3 - un fond avec une texture complexe, des petits objets d'intérêt et des objets en mouvements sur le fond (nuages)



Figure3.4 : Corpus 4 - un fond avec une texture complexe, beaucoup d'objets d'intérêt, confondus dans la texture du fond, et d'autres objets, similaires aux objets d'intérêt (des branches d'arbres)



Figure 3.5 : Exemple d'un cas, où une vidéo répondant aux caractéristiques du corpus 1 est classée dans le corpus 2.

### 3.3.2 Analyse des critères des niveaux de complexité

L'analyse des résultats d'exécution de l'algorithme sur les vidéos classées par les critères listés dans le tableau 3.1 a révélé qu'ils n'ont pas le même effet sur la séquence des traitements :

- 1) Le critère de la caméra fixe est moins important si le fond de vidéo est uni et s'il n'y a pas d'objets fixes sur l'arrière-plan, qui pourraient suivre les tremblements de la caméra et introduire des « objets-parasites ».
- 2) La distance entre les objets et la caméra a un effet significatif sur la précision de la détection, car elle définit la taille qu'une région, associée à un objet en mouvement, aura dans la vidéo. Plus la distance est grande, plus petites seront ces régions, est plus proche seront leurs couleurs à celle du fond d'image. Cela signifie que les paramètres d'objets d'intérêt et « d'objets-parasites » peuvent se rapprocher, jusqu'au point où le modèle théorique ne pourra pas les distinguer.
- 3) La présence de plusieurs types d'objets dans une vidéo d'augmente pas la complexité du traitement, car la conception hiérarchique de l'algorithme permet d'appliquer pendant le traitement plusieurs modèles d'objets
- 4) La prise de vue dans chaque frame est importante pour choisir une méthode de dénombrement final (si tous les frames d'une vidéo contiennent l'intégralité d'objets en mouvement, la moyenne des résultats de dénombrement obtenues par frames peut être employée, mais s'il s'agit d'un déplacement passant, i.e. quand dans chaque nouveau frame des nouveaux objets entrent dans le champ de vue, une approche plus complexe est nécessaire pour pouvoir de suivre l'évolution du nombre de détections)

Les quatre corpus de complexité ont été alors fusionnés en seulement deux, où la notion de contraste d'objets sur le fond d'image est le critère principal de classification. Les vidéos, réparties dans ces deux corpus (présentés dans le tableau 3.2), peuvent avoir des contenus hétérogènes, mais leurs complexités estimées de traitement seront similaires (e.g. si des objets ont un bon contraste sur un fond uni, sans aucun obstacle, mais la caméra a des tremblements à une faible amplitude, ses mouvements n'auront pas d'effet considérable sur leur représentation spatiale, ni sur la trajectoire de leurs mouvements ; i.e. le résultat d'exécution de l'algorithme ne sera pas altéré).

Cependant, dans certaines conditions certains critères ne sont pas pertinents, tandis que des autres (et leurs combinaisons) sont essentiels pour le déroulement de l'algorithme :

1. La distance entre les objets et la caméra
2. La complexité de l'arrière-plan
3. Le contraste entre les objets et le fond
4. L'intégralité du vol est comprise dans chaque frame

Ces critères ont été transformés en deux corpus de vidéos, qui ont servi dans deux phases des expérimentations (section 4.1) et en hypothèses (section 4.1) afin de concevoir les étapes de l'algorithme qui pourrait être généralisé aux niveaux de difficultés variés.

Les critères des deux corpus finaux sont résumés dans le tableau 3.2.

Tableau 3.2 : Les corpus de complexité définitifs pour l'ensemble de vidéos fournies

<b>Corpus</b>	<b>Mouvement de l'arrière-plan</b>	<b>Texture de l'arrière-plan</b>	<b>Contraste visuel d'objets d'intérêt sur le fond</b>	<b>Forme apparente d'objets d'intérêt sur le fond</b>
1	Aucun ou très faible	Uniforme	Couleur dominante des objets d'intérêt est visuellement assez contrastée de celle du fond pour les distinguer facilement à l'œil nu	Objets de tailles petites ou moyenne, avec des formes similaires
1	Fixe ou légèrement tremblante			
2	Présence d'objets en mouvement autre que les objets d'intérêt	Fond partiellement ou totalement texturé	Couleur dominante des objets d'intérêt est proche de celle du fond. Visuellement les objets ne sont pas facilement repérables	Objets de très petites tailles ou sans des formes apparentes nettes et régulières
2	Fond de la scène en mouvement (Caméra en mouvement)			



## 4. Méthodologie proposée pour la détection et le dénombrement de très petits objets

Les méthodes de détection d'objets en mouvement existantes dans l'état de l'art actuel risquent de produire des nombreuses fausses détections dans le cas particulier de la détection de petits ou très petits objets. L'analyse des techniques existantes dans le domaine du traitement d'images, telle que proposée dans le chapitre état de l'art, a d'ailleurs permis de tirer cette conclusion. Dans ce contexte, l'objectif principal des travaux de recherche menés dans le cadre de cette thèse consiste alors à concevoir une méthode de comptage des petits et très petits objets dans une vidéo prenant explicitement en compte les contraintes et difficultés qu'entraîne leurs tailles apparentes dans les images.

En effet, en pratique, dans le cadre du type de vidéo que nous souhaitons pouvoir traiter, même à l'œil nu il est parfois impossible de discerner les oiseaux ou de les distinguer les uns des autres. Un comptage visuel étant très difficile, il est a priori plus compliqué encore de les distinguer par des paramètres numériques (e.g. les couleurs de pixels dans une image d'une vidéo, associées aux objets ou au fond d'image, surtout si sa texture est complexe). L'analyse des données d'entrée – des vidéos de vol d'oiseaux migratoires enregistrées à une distance de 500 à 750 mètres depuis des postes d'observation fixes – a permis d'identifier les interactions entre les paramètres d'une scène dans une vidéo, ayant un effet sur sa complexité, et en particulier le contraste entre les oiseaux et le fond de l'image sur lequel ils se trouvent.

Dans ce contexte, ce chapitre décrit et justifie la solution, proposée pour le cas étudié de vidéos d'observation de migrations d'oiseaux. En particulier, la notion de fiabilité des détections est introduite de manière à lutter contre les problèmes de fausses détections.

Ce chapitre se décompose de la manière suivante : Tout d'abord, la section 4.1 introduit les hypothèses utilisées dans le cadre de la méthode proposée ici. La section 4.2 en présente globalement le fonctionnement de la méthode. Ainsi qu'une formalisation théorique de cette méthode. La section 4.3 présente le modèle d'objet utilisé. Les sections 4.4 à 4.6 présentent plus en détails les phases d'analyse spatiale, temporel, et de comptage en illustrant les phases d'analyse par des résultats expérimentaux.

### 4.1 Hypothèses admises dans l'approche proposée

L'approche proposée a pour objectif le comptage de très petits objets en mouvement, avec un contexte applicatif plus particulièrement dédié au comptage d'oiseaux migrateurs. Pour résoudre les difficultés de mise en œuvre d'une telle méthode de comptage, un modèle d'objet d'intérêt est mise en œuvre (voir la section 4.3).

La taille très faible des objets empêche la mise en œuvre d'un modèle géométrique précis. Le principe fondamental de la solution proposée dans ce travail de recherche se base alors sur la notion d'une « similarité » spatiale (l'apparence d'objets dans une vidéo), une « régularité » temporelle (les déplacements d'objets dans une vidéo) et sur la fiabilité de détections elles-mêmes. Sur cette base, des hypothèses de travail sont introduites pour permettre la mise en œuvre concrète de la solution proposée.

Les figures 4.1 et 4.2 mettent en évidence la difficulté de la tâche de détection de petits objets d'intérêt. De plus, la figure 4.2 représente un cas, où les variations entre les couleurs correspondent à des petits objets d'un faible contraste sur le fond, ainsi qu'à des transmissions de couleurs du fond (des variations de la teinte bleue du ciel, dans le cas illustré). Numériquement, ces variations peuvent être confondues et résultent en fausses détections. Afin de prendre en considération ces particularités de la tâche de détection de petits objets, la solution proposée base sur plusieurs hypothèses, listées ci-dessous :

### Conditions météorologiques :

- **Hypothèse 1** - Les migrations d'oiseaux sont un phénomène, qui se produit dans des conditions météorologiques calmes. Cela signifie, qu'il est possible d'avoir dans des vidéos le ciel clair ou couvert, mais pas de pluie ni de tempête, qui changeraient la scène significativement d'un moment à l'autre avec donc un fond de scène également bien plus complexe.

### Condition de prise de vue :

- **Hypothèse 2** : La distance entre les volées d'oiseaux et la caméra est constante (pas de rapprochement ni d'éloignement d'oiseaux significatif du point d'observation)
- **Hypothèse 3** : L'intégralité d'une volée d'oiseaux migratoires est contenu dans une vidéo. De ce fait, le nombre d'objets d'intérêt est constant dans chaque frame de vidéo. Si le nombre d'objets détectés varie significativement tout au long de la vidéo, cela signifie l'existence de fausses détections, ou d'objets non détectés dans tout ou partie des images. A noter que si cette hypothèse n'est pas vérifiée, elle peut être compensée par la prise en compte par le comptage des objets entrant ou sortant de la scène.

### Hypothèse sur les objets d'intérêt :

Ces hypothèses sont à la base de la construction du modèle d'objets d'intérêt. Elles varient bien sûr selon le type d'objets considérés.

- **Hypothèse 4** : La vitesse du vol est constante et sa trajectoire est linéaire, où les variations de la distance entre les objets et la caméra sont négligeables.
- **Hypothèse 5** : Les oiseaux d'un même vol sont considérés comme étant similaire en taille apparente dans une image. On suppose que cette hypothèse est satisfaite pour les oiseaux de la même espèce filmé à des distances similaires. Cette dernière hypothèse est globalement vérifiée pour un même vol d'oiseaux.
- **Hypothèse 6** : Les oiseaux d'un même vol sont considérés comme ayant une (ou plusieurs) couleur dominante similaire. On suppose que cette hypothèse est satisfaite pour les oiseaux de la même espèce. Néanmoins, les conditions de prises de vues peuvent générer des variations de luminance significative entre les oiseaux. Dans ce cas cette hypothèse n'est donc plus respectée ce qui ne peut que générer des non détections.

Les hypothèses ci-dessus sont directement introduites dans la méthode de détection proposée sous formes d'hypothèses numériques, appliquées lors de l'exécution de la solution proposée. Elles servent en tant que caractéristiques, ou paramètres, attendus pour identifier les objets d'intérêt avec un certain niveau de fiabilité. Ces hypothèses, appliquées dans le contexte de suivi de migrations peuvent s'adapter également à d'autres cas d'étude. La notion d'objets d'intérêt correspond ici à des oiseaux migratoires dans une vidéo



Figure 4.1 : Exemples de vidéos avec de petits objets, où leur contraste sur le fond permet de les distinguer visuellement et numériquement dans la séquence de traitements



Figure 4.2 : Exemple d'une vidéo avec des petits objets d'intérêt d'une forme supérieure à ceux sur la figure 4.1, mais avec un contraste sur le fond plus faible. Il n'est pas suffisant pour une distinction visuelle et numérique précise dans la séquence de traitements, d'où l'intérêt d'introduction de la notion de fiabilité de détection



Figure 4.3 : Illustration de l'hypothèse 2.1, où les petits objets sont suffisamment contrastés sur un fond du ciel voilé



Figure 4.4 : Les petits objets visuellement distinguables sur un fond texture du ciel nuageux. Dans cet exemple la détection numérique par la séquence de traitement est compliquée par des variations de couleurs dans la région avec des oiseaux est dans les régions avec les nuages.

#### **4.2 Présentation générale de l'approche proposée pour détection et dénombrement de petits objets dans des vidéos**

L'approche proposée pour la détection et dénombrement des petits objets consiste en un concept innovant de l'aspect spatio-temporel. Il sert à caractériser les objets détectés en deux côtés :

- Aspect spatial : caractérise le côté statique d'objet – son apparence dans une vidéo, qui comprend l'analyse du contraste d'un objet sur le fond, de sa forme et sa couleur
- Aspect temporel : caractérise le côté dynamique d'objet – la trajectoire et la régularité de son déplacement dans une vidéo

Le fait de combiner l'aspect spatial avec l'aspect temporel permet d'avoir un critère complémentaire permettant de différencier les vrais et les fausses détections obtenues à l'issue de la phase d'analyse spatiale (les fausses détections étant au moins approximativement conforme aux descripteurs spatiaux du modèle d'objets). L'aspect spatio-temporel est intégré dans la séquence de traitements complète (voir la figure 4.5) et est étroitement lié avec la notion de fiabilité de détections, fondamentale dans la solution implémentée.

En effet, l'aspect spatial permet de sélectionner des objets qui correspondent à l'apparence attendue des objets d'intérêt dans une image, mais il ne suffit pas pour analyser leurs caractéristiques temporelles dans la vidéo. Sans condition de déplacements, de petits objets d'une nature différente de celle des objets d'intérêt peuvent se confondre avec eux à cause de leurs paramètres spatiaux similaires à ceux d'objets d'intérêt (voir la figure 4.10). L'aspect temporel complète donc l'aspect spatial par l'analyse de déplacements d'objets d'intérêt tout au long d'une vidéo par leur côté dynamique. Il permet ainsi, d'exploiter l'ensemble du modèle d'objets en mouvement, qui se base sur plusieurs paramètres implicites d'objet (son contraste sur le fond d'image, la netteté de ses contours, les caractéristiques de son déplacement) et sur la fiabilité de traitement. Ces caractéristiques générales en combinaison avec l'estimation de fiabilité de traitement construisent une approche pour la sélection de petits objets d'intérêt flexible et adaptative. Cette particularité est essentielle, dans le cas d'étude,

où les valeurs numériques de descripteurs de la forme apparente d'objets sont très petites et sont compliquées à comparer. La section suivante formalise la détection et dénombrement de très petits objets sous forme d'interaction de fonctions, associées à chaque étape de traitement.

Deux notions principales sont à la base de l'approche pour la détection et le dénombrement de petits objets dans des vidéos proposée dans cette thèse :

1. **Introduction d'un modèle d'objet d'intérêt** : Le modèle des objets d'intérêt repose sur les descripteurs suivants :
  - **La taille des objets** qui doit être en théorie la même pour tous les oiseaux (cela suppose que les oiseaux soient tous de la même espèce, et que leurs distances à la caméra est sensiblement la même pour tous).
  - **La couleur des objets**. Elle est supposée être identique pour tous les oiseaux (espèce identique), et être donnée, d'une manière approximative par l'utilisateur.
  - **Le mouvement et la trajectoire des objets**. Les oiseaux sont supposés être en mouvement et suivre une trajectoire globalement linéaire (au moins sur une fenêtre temporelle limitée). En d'autres termes, cela signifie des objets similaires sont régulièrement détectés dans des frames consécutifs et leurs trajectoires de déplacement peuvent être approximativement tracée dans la séquence de frames analysés. Ce traitement permet d'éliminer le bruit occasionnel proche aux objets d'intérêt par les paramètres spatiaux du modèle (taille et couleur).
  - **Le mouvement en groupe des oiseaux** : Globalement, les oiseaux migrateurs vol en groupe avec donc des trajectoires et des vitesses similaires pour chaque oiseau. Cela signifie une direction et une vitesse constantes du déplacement d'un objet, ainsi que l'absence de l'accélération. Cette condition permet de supprimer les fausses détections d'objets fixes sur l'arrière-plan mais mouvants à cause de tremblements de la caméra.
2. **Introduction de critères de fiabilité** : La détection de très petits objets en mouvement rend très probable la détection de nombreuses fausses détections dues au bruit et à des mouvements très localisés du fond de la scène et s'avérant relativement proche du modèle d'objet. De plus les méthodes de détection s'avèrent bien souvent très sensibles aux seuils de détections qui sont fixés empiriquement. Pour prendre en compte ce problème, des critères de fiabilité sont associés à chaque phase algorithmique. Au final, le comptage est alors réalisé grâce à une somme pondérée par ces coefficients de fiabilité.

L'approche de détection comporte alors deux phases essentielles : une phase de détection intra-image, suivi d'une analyse temporelle.

Les travaux de recherche et de développement réalisés dans le cadre de cette thèse ont été centrés sur l'aspect spatial du modèle d'objets d'intérêt (le modèle et ces deux aspects sont présentés dans la section 4.3) et l'estimation de la fiabilité des différentes phases de traitements. Un module existant d'analyse temporel a été ensuite incorporé pour obtenir les résultats finaux de l'analyse spatio-temporelle.

La méthode de détection et de comptage proposée se découpe en trois phases principales. Afin d'intégrer la notion de fiabilité de traitement, fondamentale dans le concept développé, un **coefficient de fiabilité** est estimé au cours de chaque phase. Leurs principales caractéristiques sont les suivantes :

**Première phase** : Evaluation à priori de la fiabilité de détection par zone d'image.

L'objectif consiste à affecter un coefficient de fiabilité à chaque zone carrée dans l'image en fonction de la facilité d'identification des objets d'intérêt dans la zone considérée en évitant les fausses détections.

**Deuxième phase** : Détection d'objets d'intérêt dans les zones, où la détection d'objets est évaluée :

La phase de détection proprement dite se décompose en deux phases d'analyse, l'une spatiale, l'autre temporelle. Les principales caractéristiques de ces méthodes sont les suivantes :

- Analyse spatiale. Cette analyse se décompose selon les phases d'analyse suivantes :
  - Détection des contours d'objets dans les zones d'intérêt.
  - Evaluation de la fiabilité de détection des contours.
  - Détection des régions d'intérêt à partir des contours obtenus et des descripteurs spatiaux du modèle d'objets.
  - Evaluation de la fiabilité de la construction des régions à partir des contours
  
- Analyse temporelle. Cette analyse se décompose en trois phases principales :
  - Suivant temporelle des trajectoires d'objets
  - Sélection des objets conformes aux descripteurs temporels de mouvement
  - Evaluation de la fiabilité de l'analyse temporelle.

**Troisième phase :** Mise à jour du modèle d'objets.

Il est important de noter qu'il est nécessaire de faire évoluer le modèle d'objets en fonction de l'évolution des caractéristiques des objets détectés dans la vidéo. En effet, ces caractéristiques peuvent évoluer au cours du temps en fonction des variations d'éclairement des objets (et donc de leur changement du niveau de gris), de la distance vol d'oiseaux/caméra, ou de leur vitesse ou direction.

Les paramètres du modèle sont initialisés lors du traitement de la première image de la manière suivante : La couleur dominante des objets d'intérêt est supposée être fournie par l'utilisateur. La taille des objets est ajustée en fonction de la taille majoritaire des régions détectées. Les vitesses et directions du mouvement sont obtenues à partir de la deuxième image à partir d'une analyse et du suivi des objets détectés.

Les descripteurs du modèle sont ensuite mis à jour à chaque nouvelle image en tenant compte des estimations obtenues à l'instant précédent.

Le schéma de la figure 4.7 illustre la procédure de mise à jour du modèle d'objets d'intérêt.

**Quatrième phase :** Comptage des objets d'intérêt détectés.

La combinaison des coefficients de fiabilité définis pour chaque phase permet les coefficients de fiabilité finaux pour les objets détectés. Ces coefficients sont directement utilisés pour obtenir l'estimation finale du nombre d'objets d'intérêt détectés dans une vidéo.

La figure 4.6 présente la structure générale de la méthode proposée et les figures 4.7 – 4.9 sont des extraits d'une vidéo, qui illustre les son principe sur un cas d'application. Les schémas plus détaillés et associés individuellement à chaque phase seront donnés dans les sections qui leurs sont dédiées.

# Vidéo originale

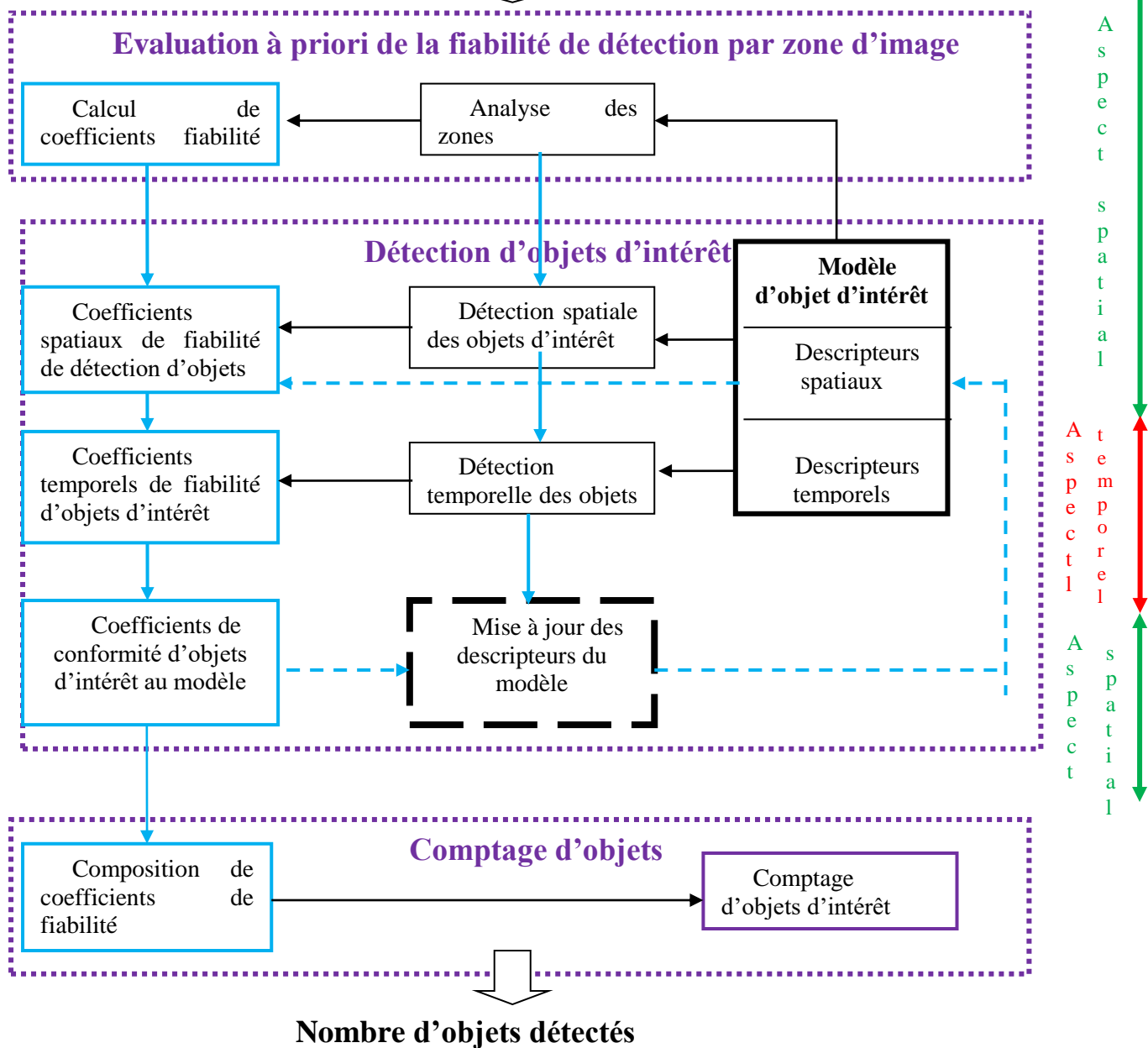


Figure 4.5 : Schéma général de la solution proposée.



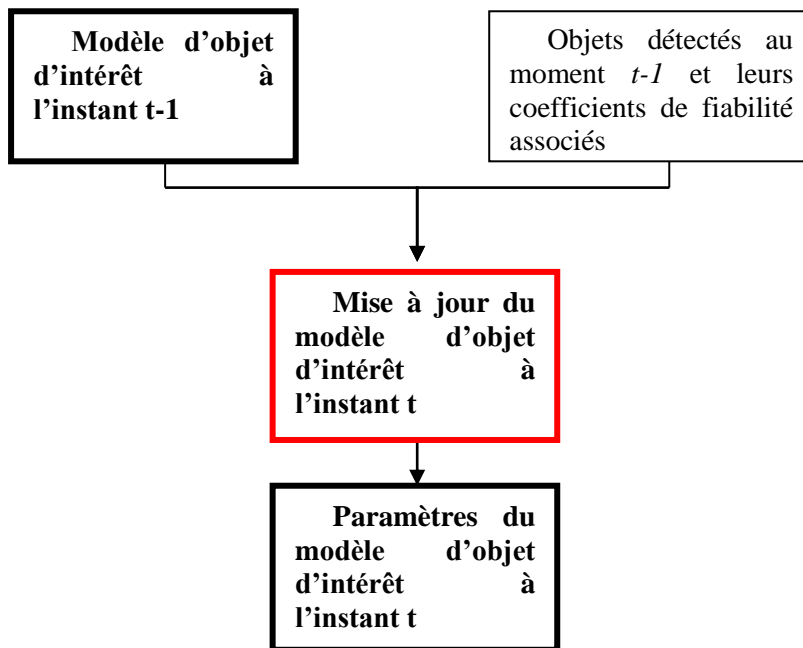


Figure 4.6 : Mise à jour du modèle d'objet. L'ensemble de tous les objets sont triés par les critères de l'aspect temporel et servent à calculer ensuite les paramètres du modèle depuis ceux d'objets en mouvement (i.e. répondant au critère de l'aspect temporel) et avec le meilleur contraste dans l'ensemble de détections.

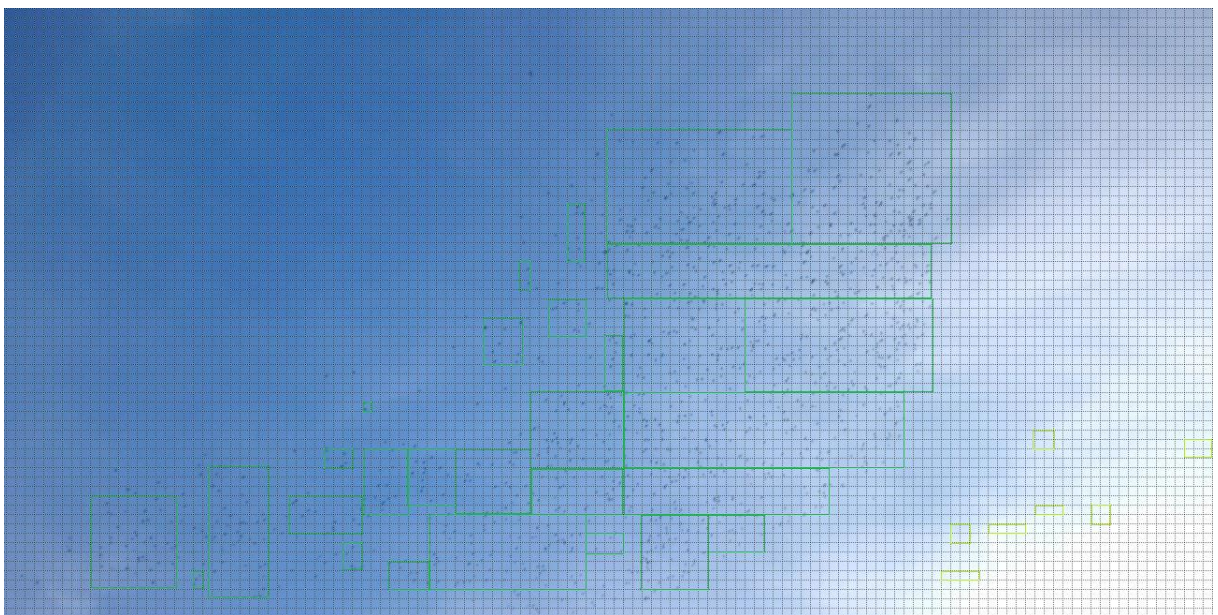


Figure 4.7 : Première phase d'évaluation à priori de la fiabilité de détection objets d'intérêt par zone d'image. L'image est segmentée en zones carrées et chaque zone est analysée à la présence d'objets d'intérêt. Des exemples de zones, où la détection d'objets d'intérêt est évaluée comme possible (i.e. les zones, qui obtiennent des coefficients de fiabilité différents de 0 à cette étape) sont encadrée dans des carrés en couleur : verte : les zones avec des vrais objets d'intérêt ; jaune : les zones avec des fausses détections, mais considérées comme des zones d'intérêt en raison des variations de couleurs générées par des autres phénomènes (ici : une voile nuageuse).

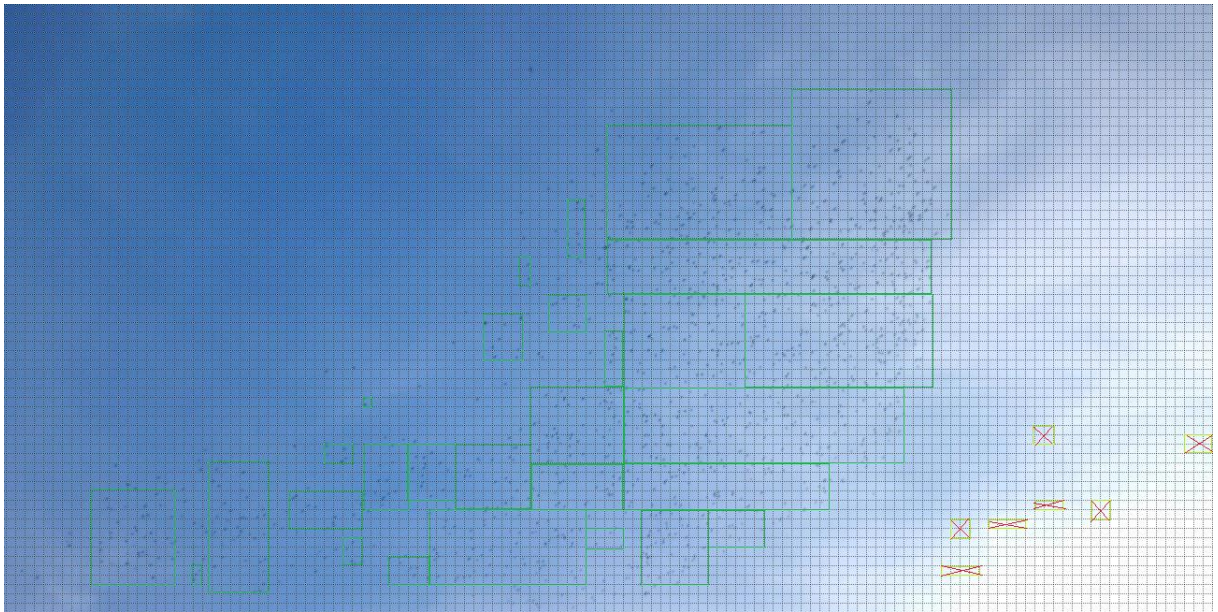


Figure 4.8 : Deuxième phase de détection et analyse d'objets d'intérêt dans les zones potentielles (avec des coefficients de fiabilité, estimés à l'étape précédente d'évaluation à priori de fiabilité de détections, différents de 0). Cette figure explique, que lors de l'exécution de la deuxième phase de détection d'objets d'intérêt, les coefficients de fiabilité, associés aux objets d'intérêt dans ces zones, sont également différents de 0. Pour les zones avec des coefficients de fiabilité de détection à priori, différents de 0, mais avec des objets, qui ne correspondent pas aux objets d'intérêt, les coefficients de fiabilité sur la deuxième phase vaut 0 (ou sont proche de 0). Cela permet la caractérisation de ces zones (et ainsi les objets contenus) comme peu fiables ou pas fiables, si le coefficient associé vaut 0 (i.e. la suppression implicite de ces zones).

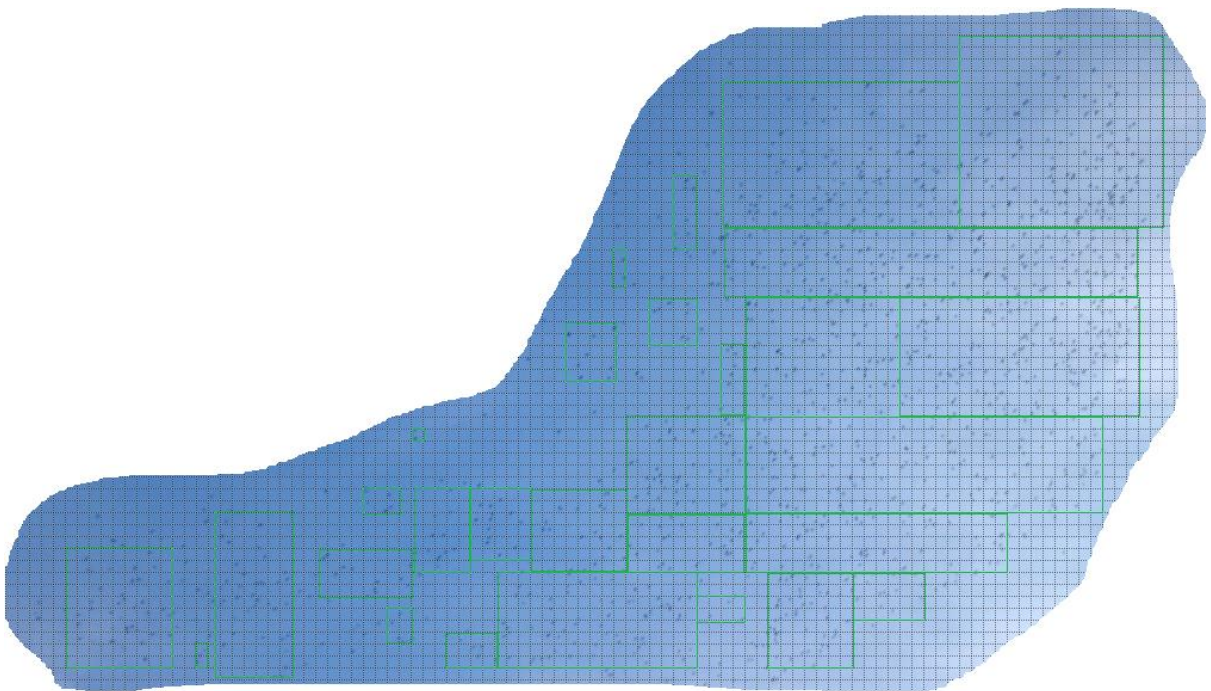


Figure 4.9 : Troisième phase de dénombrement d'objets d'intérêt. La combinaison de coefficients de fiabilité permet d'avoir à la fin d'exécution un ensemble d'objets uniquement, ou majoritairement avec des objets d'intérêt. Sur cette phase les objets sont caractérisés par leurs coefficients de fiabilité dans l'intervalle  $[0, 1]$  et sont calculés dans la somme finale. Cette figure montre, que ce sont les objets dans des zones fiables, qui composent l'ensemble final de détection (cela permet notamment de faire une simple addition pour avoir le résultat final de dénombrement).



Figure 4.10 : Illustration d'intérêt d'application de l'aspect spatio-temporel. Les carrés verts contiennent des objets, qui sont caractérisés comme des objets d'intérêt par où l'aspect spatial (l'exemple est non-exhaustif). Les carrés rouges contiennent des objets d'intérêt (des oiseaux), similaires à des fausses détections (des branches d'arbre). Les carrés jaunes contiennent des objets d'intérêt et des fausses détections. L'apparence d'objets dans les carrés jaunes et rouges est similaire à celle d'objets d'intérêt. Dans ce cas, sans application de l'aspect temporel ces fausses détections seront comprises dans l'ensemble final de détections, et le résultat de dénombrement sera très différent de la vérité terrain.

#### 4.2.1 Formalisation de la solution proposée

Dans la solution proposée les traitements dans la séquence s'enchaînent dans les trois phases, mais ils restent connectés au moyen des coefficients de fiabilité. En d'autres termes, le concept dans la solution proposée représente un concept hiérarchique d'exécution, où chaque étape suivante analyse une entité d'un niveau hiérarchique inférieur. La séquence des traitements sur une étape dépend alors de l'exécution de l'étape précédente, au niveau hiérarchique supérieur. Les coefficients de fiabilité représentent ici des fonctions, qui se superposent au fur et à mesure pour converger en une valeur qui désignera la fiabilité de détection d'un objet. Cette valeur finale définira également la probabilité, qu'un objet détecté représente un objet d'intérêt.

D'une manière formelle la combinaison de coefficients de fiabilité représente une combinaison de fonctions, qui expriment les fiabilités d'exécution des étapes de la solution, comprises dans les trois phases. Schématiquement la séquence proposée est illustrée sur la figure 4.11.

Formellement cette approche s'inscrit dans une combinaison de  $N$  fonctions, où la dernière fonction  $F_N$  applique une méthode d'analyse de toutes les étapes précédentes ensemble, et les fonctions  $f_0(O), \dots, f_{N-1}(\dots)$  sont les fonctions d'analyse de  $n$  étapes de traitement  $i \in [0, N-1]$  :

$$C(O) = f_N\{f_{N-1}[f_{N-2}(D_{N-2}), D_{n-1}], D_N\}$$

$$\text{où } C_1 = f_1(D_0) = D_1,$$

$$C_2 = f_2(C_1, D_2) = f_2(f_1(D_0), D_2),$$

$$C_3 = f_3(C_2, D_3) = f_3(f_2(f_1(D_0), D_2), D_3)$$

La solution d'estimer la fiabilité de chaque phase permet d'adapter les traitements à une vidéo d'une manière flexible, sans fixer explicitement les paramètres de méthodes des traitements. En comparaison avec le dénombrement d'objets détectés par des méthodes existantes, en cas du dénombrement par les coefficients de fiabilité la marge d'erreur se réduit de 200% à 10%.

A la fin du traitement ces coefficients de fiabilité sont assemblées en une caractéristique d'un objet – son coefficient de fiabilité final, qui servira pour le comptage dans les sommes :

$$Nb_t = \sum_{O_{i,t}} C(O_{i,t}) Nb = \frac{1}{T} \sum_T Nb_t$$

où  $Nb_t$  est la somme de coefficients de fiabilité  $C(O_{i,t})$  de chaque objet  $O_i$ , détecté à l'instant  $t$  et  $Nb$  est la moyenne d'objets détectés dans toute la vidéo de  $T$  frames.

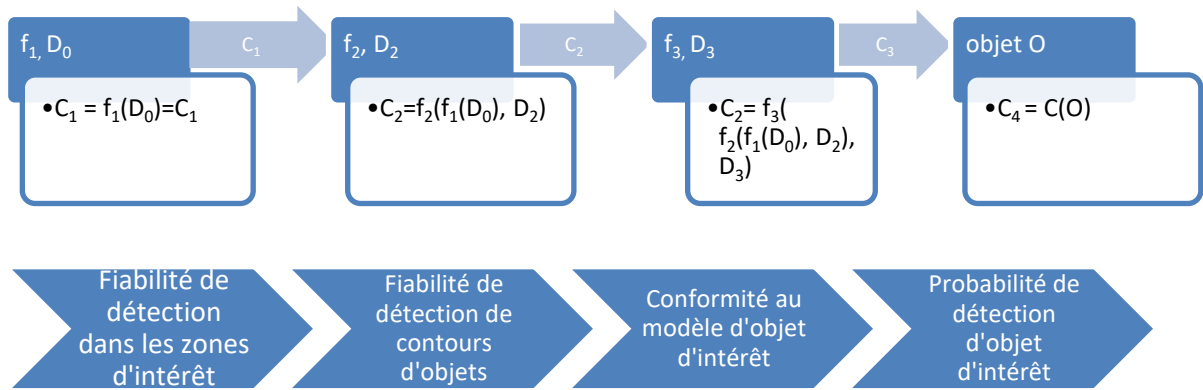


Figure 4.11 : Exécution de phases de la solution proposée, où  $f_i$  signifient les traitements propres à la phase  $i$ ,  $D_i$  – les coefficients de fiabilité de la phase  $i$  et,  $C_i$  signifient les coefficients de fiabilité de la phase  $i$  en fonction de la fiabilité de la phase  $i-1$

### 4.3 Formalisation du modèle de petit objet d'intérêt

D'une manière générale, l'apparence d'un objet dans une vidéo peut être caractérisée par les points suivants :

#### 4.3.1 Forme d'objet

- Contours exacts : les contours d'une forme irrégulière, mais la plus proche à celle d'un objet, qui l'englobent le plus précisément (typiquement en tant qu'une courbe ou un polygone non-convexe).
- Modèle de contours (primitive géométrique) : une forme d'objet approximative, mais paramétrable par des caractéristiques que peuvent la décrire d'une manière non-équivoque.
- Taille d'objet : typiquement sa superficie, i.e. la région à l'intérieur des contours d'objet

#### 4.3.2 Texture

- Couleur dominante : souvent suffisante pour présenter un objet d'une couleur globalement homogène, ou un objet éloigné, quand sa couleur est altérée (e.g. les nuances de ces couleurs fusionnent dans la même couleur)
- Modèle de texture : simple, d'une couleur uniforme, ou complexe, avec un motif et plusieurs couleurs.

### 4.3.3 Comportement dynamique

- Trajectoire de déplacement : linéaire, circulaire ou erratique
- Similarité entre objets : changement des formes des objets, dû à leurs propres mouvements (une personne qui gesticule, ou un oiseau qui batte des ailes).
- Régularité de présence d'objets : les mêmes objets se retrouvent systématiquement aux moments consécutifs et d'une manière régulière

### 4.3.4 Modèle d'objets pour les oiseaux migrateurs

Dans le cas d'étude de suivi des oiseaux migratoires leur apparence dans une vidéo est fortement affectée par la distance entre les objets d'intérêt et la caméra : une distance plus grande l'apparence d'objets est altérée et ses détails (tels que la texture ou les particularités de la forme) deviennent moins distinguables. La représentation d'oiseaux migratoires dans une vidéo rentre ainsi dans le cas de petits ou très petits objets. L'ensemble des paramètres, sur lequel se base la modélisation de l'apparence des oiseaux dans une vidéo est réduite aux caractéristiques principales suivantes :

- Forme : l'ellipse englobante les contours d'objet et/ou la taille d'objet
- Taille : Dimension, en pixels, de la région détectée
- Couleur dominante : en pratique dans le cas d'oiseaux migrateurs, les oiseaux sont de couleurs blanche, grise, ou noire.
- Trajectoire linéaire, pas d'accélération
- Trajectoire similaire pour tous les objets

L'apparence de très petits objets altérée dans une vidéo par rapport à la perception par l'œil humain, qui ne peut être décrite que par un nombre réduit de paramètres. Cela mène à des confusions fréquentes d'objets- « parasites » ou du bruit avec des objets d'intérêt - les oiseaux migrateurs –lors de la détection dans une vidéo. Le modèle théorique du petit et très petit objet d'intérêt sert à estimer la fiabilité de détection, en termes de la notion centrale de la solution proposée – le coefficient de fiabilité. Notamment, le coefficient de fiabilité, qu'interprète la conformité d'objet détecté au modèle théorique (voir section 4.4.2 pour une introduction détaillée) permet de distinguer des objets d'intérêt des fausses détections, typiques pour les vidéos faites depuis les postes d'observation et les traitements d'une image à une basse qualité :

1. Végétation sur l'arrière-plan : des branches d'arbres fines ou des petites feuilles en mouvement à cause du vent ou des tremblements de la caméra, dont l'apparence est similaire à celles des objets d'intérêt - des oiseaux dans le cas d'étude (voir Figure 4.17 pour des exemples).
2. Bruit sur une image : des très petites régions de nature différente, mais similaires aux objets d'intérêt (le bruit numérique dans des images initiale d'une basse qualité, des régions générées sur les étapes intermédiaires de la solution proposée).
3. Régions de transmission des couleurs (perceptibles lors des traitements d'images): des changement des composantes-couleur d'une valeur à une autre contribuent à l'attribution d'un coefficients de fiabilité élevé pour les zones d'intérêt par variation d'intensité des couleurs et à la détection des limites entre les couleurs en tant que contours d'objets

Par exemple, sur les figures 4.12 et 4.13 il est très difficile de distinguer visuellement les objets d'intérêt (les oiseaux migrateurs) dans une vidéo (figure 4.12), car leur couleur fusionne avec celle du

fond. Numériquement, dans le cas d'une basse qualité d'une vidéo, le bruit présent est similaire par son apparence aux objets d'intérêt (figure 4.13). Dans ces cas la comparaison d'objets détectés avec un modèle théorique au moyen de coefficients de fiabilité permet d'éviter un choix excessif d'objets dans l'ensemble d'objets d'intérêt



Figure 4.12 : Exemple de difficulté de traitement de vidéos avec des très petits objets – d'un vol d'oiseaux très éloigné. La distance entre les oiseaux et la caméra altère la qualité de la vidéo. Dans ce cas il est très difficile d'estimer le nombre d'oiseaux visuellement. Numériquement, ils se confondent avec des faibles variations sur le fond : celles des reflets du soleil ou des légers changements de teintes de la couleur bleue du ciel. Le suivi d'objets individuellement n'est pas possible, car ils fusionnent en un objet d'une forme irrégulière.



Figure 4.13 : Agrandissement de la zone d'intérêt sur la figure 4.12. Cet exemple montre, qu'une grande distance entre les objets d'intérêt – les oiseaux-migrateurs – et la caméra altère leurs apparences dans des zones d'intérêt et complique les traitements sur les trois phases de la solution proposée. Ici, les objets n'ont pas de contours nets et fusionnent visuellement.

### 4.3.5 Limites du modèle d'objets introduit dans la solution

Grâce à la combinaison de notions générales (la taille et la couleur) et celle de fiabilité de traitement le modèle de petit objet peut être appliqué pour les cas, où d'autres méthodes existantes ne peuvent pas être utilisées efficacement. Par exemple, il est difficile de reconstruire la forme d'un objet, dont la taille ne fait que quelques pixels et la comparer avec une forme de référence. C'est le cas de la plupart d'objets dans les vidéos du corpus 2, où une distance significative entre le vol d'oiseaux et la caméra réduit considérablement leurs tailles dans une vidéo (voir la figure 4.14). Plus cette distance est grande, plus petite est la taille d'objets, et ils peuvent être confondus avec le bruit. La figure 4.15 est une illustration de ce phénomène, où des très petits objets d'intérêt avec une faible contraste sur le fond et des légères variations de composantes colorimétriques résultent en contours similaires et porte à confusion dans la suite de traitements.

Pour résoudre cette ambiguïté, on estime une taille minimale, au-dessous duquel une détection est considérée comme du bruit. Cette limite est dérivée depuis la formule de distance approximative entre un objet et la caméra :

$$Distance_{camera \rightarrow objet}(mm) = \frac{F(mm) \cdot H_{obj}(mm) \cdot H_{img}(px)}{H_{obj}(px) \cdot H_{sensor}(mm)}$$

où  $F(mm)$  est la distance focale, estimée en millimètres,  $H_{obj}(mm)$  est la hauteur réelle d'objet d'intérêt,  $H_{obj}(px)$  est la hauteur d'objets dans l'image, estimée en pixels,  $H_{img}(px)$  est la hauteur de l'image en pixels et  $H_{sensor}(mm)$  est la taille de sensor, dépendante de la caméra.

La hauteur approximative d'un objet est alors estimée comme :

$$H_{obj}(px) = \frac{F(mm) \cdot H_{obj}(mm) \cdot H_{img}(px)}{Distance_{camera \rightarrow objet}(mm) \cdot H_{sensor}(mm)}$$

Généralement, si les paramètres de la caméra et les conditions d'enregistrement d'une vidéo sont renseignés, les variables de la formule peuvent être définies :

- Hauteur réelle d'objet d'intérêt (estimée en millimètres) :  $H_{obj}(mm)$ . En fonction d'une application concrète (le suivi des oiseaux migratoires, le contrôle du trafic routier, la vidéosurveillance de lieux publics), cette variable peut être définie par la taille d'une espèce d'oiseaux dans le vol observé, la taille moyenne d'une voiture légère ou d'un poids lourd, ou la taille moyenne d'une personne.
- Distance entre les objets et la caméra (estimée en millimètres) :  $Distance_{camera \rightarrow objet}(mm)$  En acceptant que la distance entre les objets et la caméra ne varie pas d'une manière à modifier la forme apparente d'objets dans une scène, sa valeur peut être définie en tant qu'un paramètre fixe de la distance entre le poste d'observation et le vol (pour le cas de suivi de migrations) où la hauteur de la position d'une caméra de vidéosurveillance au-dessus d'une autoroute ou un lieu public.
- La distance focale  $F(mm)$  et la taille de sensor  $H_{sensor}(mm)$  : sont les caractéristiques d'un appareil utilisé [46]

- Hauteur de l'image  $H_{img}(px)$  (estimée en pixels) : est une caractéristique d'une vidéo enregistrée (e.g., pour le type de vidéo HD la résolution est 1920x1080 pixels, d'où la hauteur d'image de 1080 pixels)

Sachant ces paramètres, on peut estimer la taille d'un oiseau dans une vidéo avec un vol éloigné à 750 mètres – la distance maximale, où les oiseaux seront encore visibles à l'œil. C'est également la distance maximale entre les vols d'oiseaux migratoires et le point d'observation. Dans ce cas concret l'estimation par la formule ci-dessus, le seuil pour la taille minimale d'un objet d'intérêt dans une vidéo est de 2 pixels.



(a)



(b)

Figure 4.14 : Différences en apparences d'objets d'intérêt en fonction de la distance de la caméra : (a) 50 mètres entre les objets d'intérêt et la caméra (b) environ 600 mètres



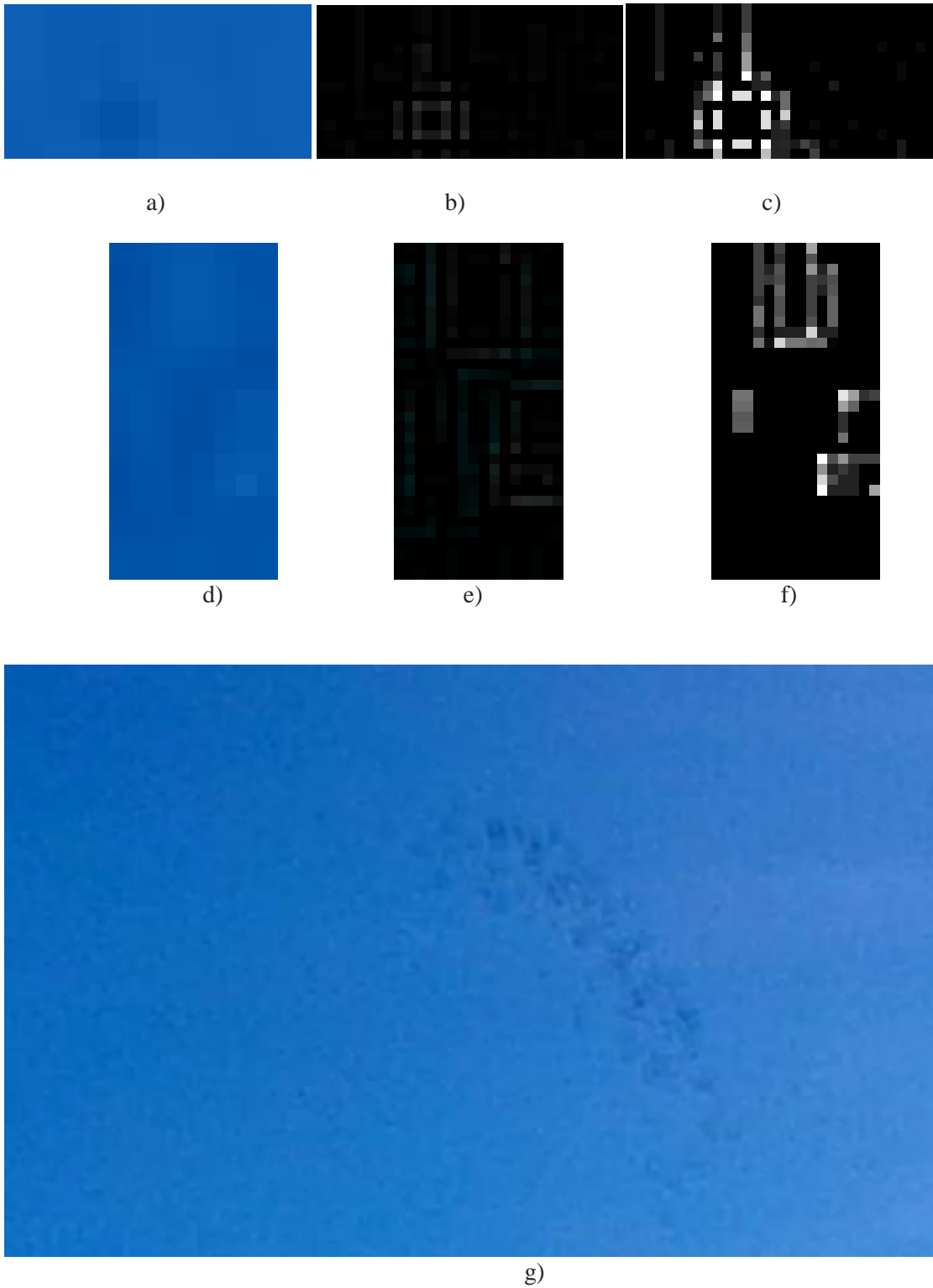


Figure 4.15 : Illustration de difficulté de distinction de petits objets à une grande distance de la caméra. Les deux images a) – f) représentent des zones, extraites depuis l'image g) contenant des très petits objets d'intérêt et des régions de variations de composantes colorimétriques, leurs contours détectés par l'opérateur Laplacien standard et sa version adaptée. Cet exemple met en évidence qu'à partir d'une certaine distance la distinction des très petits objets et de fausses détections devient très difficile, voire impossible, d'où l'intérêt de limiter les paramètres du modèle d'objet.

#### **4.4 Présentation de l'aspect spatial de la solution proposée pour la détection de très petits objets**

L'aspect spatial de la solution est destiné à reconstruire la forme apparente de petits objets d'intérêt dans une vidéo et d'évaluer la conformité des formes d'objets d'intérêt à cette forme attendue. Dans le cadre de l'objectif principale de la solution proposée : la détection et le dénombrement de petits objets des natures différentes dans des vidéos, la reconstruction de l'apparence attendue d'objets d'intérêt est une tâche difficile. Le principe proposé pour la reconstruction d'une forme d'objets sans information préalable sur leurs apparences est alors de procéder par l'approche descendante (des notions globales vers les notions locales). Dans le contexte du cas d'étude du suivi d'oiseaux migratoires cette approche descendante se présente sur le schéma de la figure 4.16.

Le premier niveau hiérarchique (le niveau supérieur) est une image, extraite de la vidéo, où le dénombrement d'objets doit être fait. Le niveau suivant, correspondant aux zones d'intérêt, correspond à la première phase de traitement d'évaluation a priori de la fiabilité de détection par zone d'image. Les deux niveaux inférieurs et donc le plus locaux correspondent à la deuxième phase de détection d'objets d'intérêt. Cette présentation hiérarchique permet de mieux illustrer comment les coefficients de fiabilité intègrent les phases de traitement et comment ils sont assemblés pour la phase finale de dénombrement.

Chaque étape à l'intérieur de phases de traitement résulte en un coefficient de fiabilité associé à chaque niveau hiérarchique. Les coefficients sont estimés conformément aux notions qui servent aux traitements des entités hiérarchiques. Par exemple, la première étape de l'algorithme classe des potentielles zones d'intérêt par l'estimation de variation d'intensité lumineuse des couleurs de ces pixels, qui caractérisera le contraste entre eux. Le coefficient de fiabilité dans ce cas est associé à toute une zone d'intérêt. Il signifie l'intensité du contraste globale des objets d'intérêt probablement présents dans cette zone.

Ainsi, grâce à l'application de la notion de fiabilité de détection, il est possible d'introduire abstraction d'un cas d'application, c'est-à-dire d'une nature concrète d'objets d'intérêt et de l'environnement dans une vidéo. Ce principe permet alors d'atteindre l'objectif de la solution proposée et l'adapter pour des applications variées.

Les sections de ce chapitre sont consacrées à la description d'étapes des solutions et de coefficients de fiabilité associés.

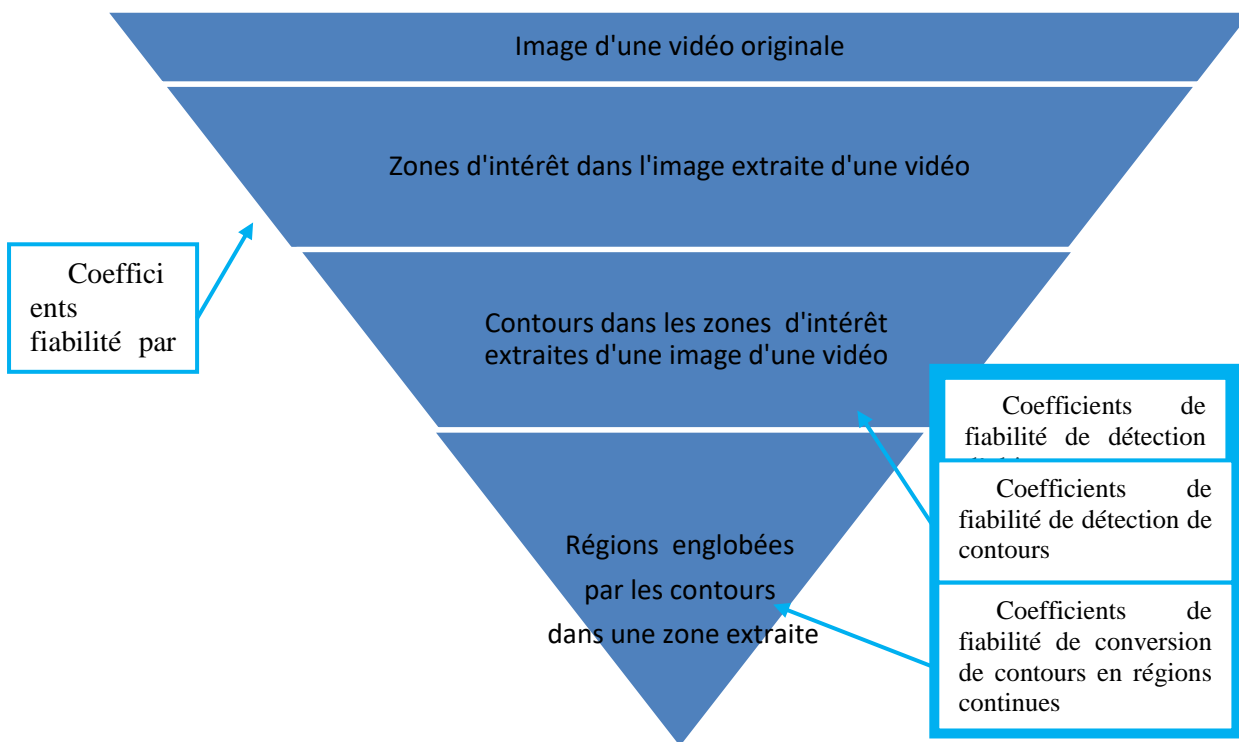


Figure 4.16 : Schéma des niveaux hiérarchiques dans la solution proposée

#### 4.4.1 Domaines scientifiques dans l'aspect spatial

Les domaines scientifiques appliquée pour la conception de l'aspect spatial sont celles de :

- **Colorimétrie** : Des différents espaces colorimétriques pour l'analyse de couleurs : YUV, HSV et CIELa\*b\*
- **Statistique** : la moyenne et la variance, observés dans l'ensemble de données

La colorimétrie intervient sur la première phase de la solution proposée. Les notions et les techniques de ce domaine permettent d'analyser des plusieurs natures de couleurs de pixels, exprimées par des différentes valeurs dans des différents espaces colorimétriques :

- Espace YUV : consiste des composantes de luminance (Y) et les composantes de chrominance (U et V). Les trois composantes sont dérivées de celles dans l'espace RGB, où la chrominance encode l'information sur la couleur, et la luminance – son intensité i.e. l'information sur la couleur est séparée de l'information sur sa vivacité. L'espace colorimétrique YUV est largement utilisé pour encoder les vidéos, car il reflète plus exactement la perception visuelle d'une scène et ne dépend pas de caractéristiques techniques des appareils utilisés (à la différence de l'espace RGB, où les paramètres physiques d'une caméra utilisée font varier les valeurs de composantes RGB). La composantes de luminance (Y) est dérivée des composants-couleur dans l'espace RGB et exprime, à quel point la couleur d'un pixel dans une zone est vive.
- Espace HSV : consiste de composante de teinte H (« Hue »), dont la valeur représente un angle sur le cercle de couleurs, de saturation S et de la valeur V. Les deux dernières

composantes définissent respectivement la profondeur de la couleur et sa brillance (i.e. la quantité de blanc). Le cercle de couleurs à son tour est séparé en intervalles - groupes de couleurs similaires. Chaque intervalle correspond à des couleurs et leurs nuances visuellement distinctes, même si les deux autres composantes dans l'espace HSV ont des valeurs proches. L'espace colorimétrique HSV permet d'analyser les différents aspects de couleur indépendamment et donner ainsi des caractéristiques complémentaires pour les méthodes d'analyse, conçues dans la solution

- Espace CIE  $La^*b^*$  : dans cette espace colorimétrique les composantes couleur expriment sa clarté L (le dérivé de luminance) et les relations  $a^*$  et  $b^*$  entre composantes couleur dans l'espace RGB. Comme l'espace YUV,  $La^*b^*$  est indépendant de l'appareil utilisé et permet d'exprimer la couleur d'une manière uniforme [45]. Espace  $La^*b^*$  est conseillé par CIE pour l'analyse de couleurs, qu'y devraient être converties pour la comparaison par la formule CIE2000. Cette formule est précise, mais coûteuse en complexité des calculs. Dans le contexte de détection de petits objets son application est justifiée, car leurs tailles ne dépassent pas quelques dizaines de pixels.

Les composantes dans des espaces colorimétriques sont des nombres, qui représentent de différents aspects d'une couleur. L'analyse de ces aspects dans l'ensemble de pixels d'une zone permet d'y observer les distributions de différentes composantes de couleurs et faire une hypothèse, si une zone représente un intérêt pour les traitements suivants. Puisque les zones d'images représentent des ensembles nombreux de pixels. Ce phénomène permet d'avoir des ensembles d'échantillons et d'appliquer des notions et de méthodes de la statistique.

#### 4.4.1.1. Application de méthodes statistiques pour l'analyse de zones dans les espaces colorimétriques

Les notions centrales de statistique appliquées dans la solution sont celles de la *moyenne* et de la *variance*. Elles sont largement utilisées pour l'analyse des grands volumes de données [52]. Par exemple, elles trouvent leurs applications dans le domaine du traitement d'images, où un tel volume de données peut être fourni. Par exemple, une image HD de la résolution 1080x1920 en niveaux de gris donne 2 073 600 échantillons et 6 220 800 si c'est une image en couleurs avec trois composantes. Dans le cas du traitement des vidéos le nombre d'échantillons disponibles est encore plus élevé.

Ayant un grand ensemble d'échantillons sous forme des caractéristiques de couleurs, leurs distributions peuvent être construites et analysées. Les valeurs de paramètres de ces distributions -de la moyenne et de la variance - permettent de supposer, si une zone d'image contient ou pas des objets d'intérêt. Par exemple, si les variances de composantes de couleur sont faibles, on suppose que la zone a une couleur visuellement uniforme. Cette zone peut être considéré comme sans objets d'intérêt (homogène). A l'inverse, s'il existe des variations de couleurs dont certaines sont proches de la couleur des objets d'intérêt, la probabilité qu'il ait un (ou plusieurs) objets d'intérêt est nettement plus élevée.

La *variance* d'intensités de leurs couleurs est appliquée sur la composante Y (luminance) en espace colorimétrique YUV. Elle permet de séparer les notions de l'intensité de couleurs (i.e. à quel point une couleur est claire) et de la différence entre les couleurs elles-mêmes (i.e. bleu et vert). Ainsi avec la composante Y il est possible d'estimer la variation d'intensités de couleurs dans une zone globalement, sans prendre en considération des faibles variations possibles des teintes de couleurs au sein du même intervalle de teintes. Ce traitement dans l'espace YUV permet ainsi de classer une zone en tant que

- **Zone hétérogène** : une zone, où un ensemble de pixels ont des couleurs visuellement distinctes, qui correspondent probablement à des objets d'intérêt. Une zone de ce type contient

typiquement des pixels des couleurs différentes ou des mêmes couleurs, mais visuellement différents par sa vivacité (plus claires ou plus sombres). Les régions possédants ce type de caractéristiques auront un coefficient de fiabilité d'autant plus élevé que l'on retrouve des couleurs proches du modèle d'objet.

- **Zone homogène (ou zone uniforme)** : une zone avec des faibles variances entre les intensités des couleurs des pixels. En pratique, ces variations dans des zones homogènes ne représentent pas des potentielles objets d'intérêt, mais des faibles variations d'une teinte d'une couleur dans un même intervalle (i.e. des nuances du bleu dans la couleur du ciel). Les zones de ces types sont considérées comme vides, i.e. sans objets d'intérêt. Elles obtiennent alors les coefficients d'évaluation à priori de fiabilité de détection par une zone d'image proche ou égaux à 0.

Ainsi la notion statistique de la *variance*, appliquée à l'ensemble de valeurs de la composante Y permet de présenter les différences entre les échantillons d'un ensemble des couleurs des pixels sous une forme numérique (un nombre) et de caractériser le degré de variation des intensités couleurs dans chaque zone d'intérêt. La *variance* d'intensité de couleurs représente alors une notion importante de la première phase de traitement et du prétraitement, qui sert à estimer la variance attendue pour des zones avec des objets d'intérêt.

#### 4.4.2 Prétraitement de la vidéo pour l'initialisation de la première étape

Afin d'évaluer quelle valeur caractérisera une variance dans les zones avec des objets d'intérêt potentiels, une analyse préalable de contenus de zones est nécessaire. La figure 4.17 montre, que choisir une valeur, fixée pour toute une vidéo ou toute une image, en tant que seuil de la variance minimale pour les zones avec des objets d'intérêt, ne suffit pas pour considérer certaines zones comme de zones avec des objets d'intérêt. L'effet sur le résultat final de dénombrement d'une valeur du seuil trop basse ne sera pas important, car les fausses détections seront caractérisées en tant que telles par les coefficients de fiabilité. Par contre, si la valeur du seuil est trop élevée, certains objets d'intérêt peuvent être perdus, car la variance lumineuse de couleurs dans les zones, qui les contiennent, sera inférieure au seuil (voir la figure 4.17).

L'objectif du prétraitement, présenté dans ce chapitre, est d'avoir un moyen flexible d'estimation du seuil de variation d'intensités lumineuses de couleurs dans les zones d'image. La méthode proposée est conçue pour s'adapter aux différences d'apparences d'objets d'intérêt dans les zones (leurs contrastes sur le fond) ainsi que les conditions globales dans une vidéo (changement de la luminosité dans une scène).

Pour avoir un critère d'évaluation à priori de détection d'objets d'intérêt dans les zones d'image (i.e. leur classification), qui s'adaptera à la scène dans une vidéo, la solution proposée applique une méthode conçue pour le cas de petits objets dans les vidéos et basée sur la comparaison d'histogrammes.

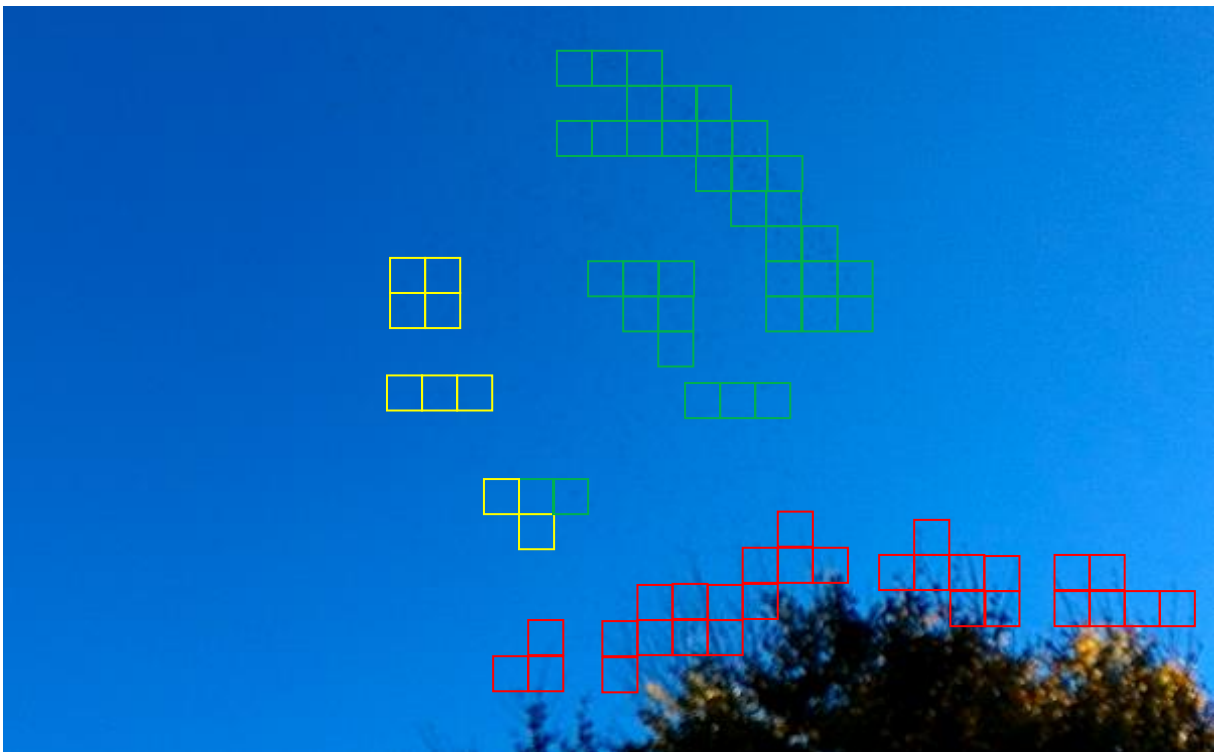


Figure 4.17 : Exemple d'une image, où le prétraitement a un effet sur la classification des zones. Carrés rouges : zones avec des objets sur l'arrière-plan (des branches d'arbres), où la variance d'intensité lumineuse de couleurs sont supérieures à celle de zones vertes avec les objets d'intérêt (les oiseaux). Carrés jaunes : zones avec des objets d'intérêt, mais une variation d'intensité lumineuse inférieure à celle dans les zones vertes. Avec trois valeurs différentes de la variance d'intensités de couleurs une valeur fixe en tant qu'un critère de classification de zones par la variance peut être trop basse. Elle ne permettra pas d'attribuer des coefficients de fiabilité élevés aux zones avec des objets d'intérêt plus faiblement contrastés (les zones dans les carrés jaunes, par rapport à celles dans les carrés verts).

#### 4.4.2.1 Principe d'estimation du critère de classification de zones avec des objets d'intérêt

En référence avec les niveaux hiérarchiques sur la figure 4.16 les entités analysées à l'étape de prétraitement (les images d'une vidéo) appartiennent au niveau supérieur et le plus global parmi tous les niveaux hiérarchiques. Le principe d'estimation du critère pour la classification de zones d'intérêt se base sur la notion d'histogramme, qui donne une aperçue de l'ensemble de pixels dans une image, ou dans une zone d'image, et correspond alors à ce niveau supérieur.

Dans le contexte d'hypothèses citées dans la section 4.1 les conditions météorologiques sont calmes et constantes tout au long d'une vidéo. En pratique, les changements de certaines caractéristiques, difficilement repérables à l'œil (telles qu'un léger changement de la luminosité dans une scène qui fera varier les intensités lumineuses des couleurs dans la vidéo) auront un effet sur la valeur du critère de classification. Pour la classification de zones avec des très petits objets les moindres différences de variations d'intensités lumineuses de couleurs peuvent être significatives, et il est alors important de mettre à jour le critère de classification de zones au cours du traitement.

Pour prendre en considération cette particularité, la méthode conçue analyse des intervalles égaux de taille  $T$ , qui composent tout l'ensemble des frames d'une vidéo. Pour chaque intervalle la première image – l'image  $t$  de référence – caractérise l'intervalle correspondant de  $T$  frames. La variation de l'intensité lumineuse, estimée dans chaque image  $t$  ainsi sélectionnée sera utilisée pour le calcul du coefficient de fiabilité associée à zones bloc dans tout l'intervalle de  $T$  frames qui lui est associé.

L'idée d'estimation de ce critère par intervalles d'images est de proposer une méthode adaptative de classification de zones d'intérêt pour la première étape et une abstraction du contenu d'une vidéo (i.e. sans prendre en considération la forme apparente d'objets dans la zone). Les histogrammes notamment permet d'exprimer des phénomènes globaux observés (dans ce cas : l'intensité lumineuse des pixels d'une zone) et de les comparer. Pour cette comparaison on admet, que les zones avec des potentiels objets d'intérêt de très petites tailles, ont des contenus similaires, qui peuvent être représentés par les distributions des intensités lumineuses de leurs pixels, i.e. par leurs histogrammes. Ils permettent de comparer les zones d'images sans analyser les objets qu'ils contiennent.

La méthode proposée pour l'estimation du critère de classification des zones d'intérêt fonctionne en plusieurs étapes :

1. Segmentation d'une image en zones, pour des questions de simplicité, un découpage en blocs carrés est utilisé.
2. Génération de l'histogramme de l'intensité lumineuse pour chaque bloc
3. Estimation des différences qualitatives entre les histogrammes pour toutes les paires de blocs d'image : comparaison de corrélations entre les histogrammes [39][51].
4. Estimation des différences quantitatives entre les histogrammes : complaisons de distances  $\chi^2$  entre les histogrammes des zones d'images
5. Estimation du seuil de variation d'intensité lumineuse pour la classification de zones d'intérêt

Dans le cas d'adaptation du seuil de variation lumineuse pour la classification de zones d'intérêt on s'intéresse à la marge de différences entre toutes les zones présentes dans une image. Sans aucun moyen de renseigner une variation initiale, elle ne peut être évaluée qu'approximativement, par une estimation qualitative. L'analyse des corrélations entre les histogrammes de distribution d'intensités lumineuses sur la troisième étape de prétraitement permet d'appliquer cette estimation et évaluer les différences qualitatives entre les zones d'images. Elles définiront alors sur l'étape suivante un seuil de variation, à partir duquel une zone d'image sera considérée comme une zone d'intérêt. Ensuite, dans

l'ensemble des zones mutuellement différentes, une caractéristique quantitative de la distance  $\chi^2$  entre les histogrammes pourra ajuster le paramètre numérique du seuil de variation, nommé par la suite  $Y_{\text{texture}}$ .

Les deux sections qui suivent expliquent les bases théoriques de cette approche analytique.

#### 4.4.2.2. Analyse qualitative : corrélation entre les histogrammes

La corrélation entre deux histogrammes permet de définir, à quel point ils sont similaires. En d'autres termes, cette corrélation exprime la similarité de distributions de données, représentées par les histogrammes. En s'exprime en une valeur limitée, comprise dans l'intervalle entre 0 et 1. Plus les deux histogrammes sont similaires, plus la valeur de leur corrélation est proche à 1, plus ils sont différents, plus cette valeur s'approche à 0. Elle est estimée par la formule suivante :

$$\text{Correlation}(H_1, H_2) = \frac{\sum_I (H_1(I_1) - \overline{H_1})(H_2(I_2) - \overline{H_2})}{\sqrt{\sum_I (H_1(I_1) - \overline{H_1})^2 \sum_I (H_2(I_2) - \overline{H_2})^2}}$$

où  $\overline{H_k} = \frac{1}{N} \sum_J H_k(J)$ ,  $H_1(I_1)$  et  $H_2(I_2)$  sont les histogrammes d'images  $I_1$  et  $I_2$  respectivement.

L'avantage de la comparaison de deux histogrammes au moyen d'un coefficient de corrélation est la possibilité de détecter des différences entre les distributions (dans ce cas : des intensités lumineuses entre de leurs pixels) sans connaître les valeurs des échantillons. Dans le prétraitement cette classification préalable de zones d'intérêt au moyen de corrélation entre les histogrammes permet de caractériser les zones par un paramètre qualitatif, dont la valeur est comprise dans un intervalle limité. Cette approche permet de détecter les différences entre les variations d'intensité de couleurs dans des zones, qui seront suffisamment importantes afin d'évaluer un seuil de variation d'intensité – c'est à dire avec un paramètre quantitatif.

Pour cette première analyse on admet que les paires de zones très similaires, i.e. dont les corrélations entre les histogrammes des intensités de couleurs correspondants, sont comprises dans l'intervalle de [0.9, 1] sont des zones homogènes du ciel, car il est peu probable d'avoir deux zones avec des objets d'intérêt presque identiques. Les observations et les expérimentations ont prouvé que pratiquement toutes les zones avec les valeurs de corrélation dans l'intervalle de [0.9, 1] sont des paires de zones avec un fond visuellement homogène ou avec des très faibles variations d'intensités de couleurs. Ces variations se réfèrent à des légères transmissions de couleurs ou à des objets qu'il n'est pas possible de distinguer sans un agrandissement.

L'autre intervalle des valeurs extrêmes : [0, 0.3] sert à caractériser les paires de zones dont les histogrammes des intensités de couleurs de pixels (et donc de leurs contenus) sont différents. Les zones de ce type ne sont pas utilisées dans le prétraitement afin de ne pas augmenter la valeur du seuil estimé. La variation d'intensité de couleurs dans ces zones sera supérieure à celle qui sera définie lors du prétraitement. Ainsi, dans le cas où ces zones contiennent des objets d'intérêt, elles auront des coefficients de fiabilité de détection élevés sur les étapes suivantes de la solution.

Pour construire un ensemble de zones, où les variations d'intensités de couleurs de pixels peuvent correspondre à des petits objets d'intérêt, l'intervalle [0.3, 0.9] des valeurs de corrélations entre les histogrammes a donc été fixé expérimentalement.



#### 4.4.2.3. Analyse quantitative : la distance du $\chi^2$ entre les histogrammes

Un autre moyen de comparer deux histogrammes est d'appliquer les propriétés de leur distance  $\chi^2$  : plus la distance  $\chi^2$  entre les deux histogrammes est petite, plus ils sont similaires, et plus elle est élevée, plus les histogrammes sont différents. Le  $\chi^2$  s'écrit de la manière suivante :

$$\chi^2(H_1, H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)}$$

La distance  $\chi^2$  égale à 0 signifie deux histogrammes identiques, et sa valeur, différente de zéro indique la présence de différences entre les deux histogrammes. Par contre, en fonction d'histogrammes comparés la distance peut varier dans l'intervalle  $(0, +\infty)$ , ou, en d'autres termes, elle n'a pas de limite numérique, s'il s'agit de deux histogrammes différents. Dans le contexte d'application, cela signifie, qu'uniquement la valeur de distance  $\chi^2$  entre les histogrammes pour deux zones aléatoires d'une image, ne peut pas définir si elles sont similaires ou différentes.

L'objectif du prétraitement est d'estimer un seuil, qui pourrait identifier des zones avec des faibles variances, probablement référées à des zones avec des petits objets d'intérêt. Sachant que la scène d'une vidéo peut avoir des objets des natures variées, les variances dans les zones avec des petits objets d'intérêt ne sont pas forcément les maximales parmi toutes les zones d'image. Dans ces conditions il faut avoir la valeur maximale et minimale pour les distances entre les histogrammes de paires de zones d'une image.

Une façon de le faire est d'appliquer un calcul couteux de toutes les distances  $\chi^2$  pour toutes les paires de zones d'une image. Autre solution, appliquée dans le prétraitement, est l'application simultanée de la corrélation de deux histogrammes et de leur distance  $\chi^2$ . La première est limitée par l'intervalle  $[0, 1]$  et peut classer les deux histogrammes (et ainsi deux images) comme parfaitement identiques ou complètement différentes. Cette caractéristique est ajoutée sur celle de la distance  $\chi^2$  afin d'estimer sa limite approximative. Les sections 4.4.2.4 et 4.4.2.5 expliquent le principe de cette estimation.

#### 4.4.2.4. Combinaison de la corrélation et de la distance de $\chi^2$ entre les histogrammes sur l'étape de prétraitement

Si la corrélation entre les deux histogrammes sert à présélectionner un ensemble des zones, mutuellement différentes les unes des autres, elle ne suffit pas pour estimer la valeur de seuil de variation d'intensité. La corrélation entre deux histogrammes représente un nombre flottant dans l'intervalle  $[0, 1]$  et elle s'applique à des paires d'images. Mais sur la première phase d'évaluation de zones d'intérêt, c'est un nombre qui caractérise le niveau de variation des intensités de couleurs et qui s'applique à une seule zone. Pour cette raison la corrélation entre les histogrammes n'est pas utilisée pour estimer le seuil de variation directement, mais pour choisir les zones d'images les plus appropriées pour son estimation.

Les deux mesures : la corrélation et la distance  $\chi^2$  entre les deux histogrammes sont alors complémentaires et l'application de ces deux mesures ensemble permet d'avoir deux interprétations de cette comparaison :

- Qualitative : la corrélation entre les paires d'histogrammes dont les valeurs sont limitées dans l'intervalle  $[0, 1]$

- Quantitative : la distance  $\chi^2$  qui interprète la corrélation entre les deux histogrammes dans une mesure de différence entre deux images. Les valeurs de distance  $\chi^2$  varient de 0 à l'infini. En pratique dans l'approche proposée elles sont toujours limitées, car la distance  $\chi^2$  est estimée uniquement pour l'ensemble de zones d'images, ou corrélations associées entre les histogrammes sont comprises dans l'intervalle (0.3, 0.9)

Ces deux interprétations expliquent l'intérêt d'utiliser les deux caractéristiques pour estimer et ajuster le seuil de variation d'intensité de couleurs dans les zones d'intérêt, dont la classification aura un effet sur toutes les autres étapes de la solution conçue.

Une autre raison d'application simultanée de ces deux mesures est de réduire le nombre de zones d'intérêt potentielles. Autrement, le fait de conserver les zones homogènes, avec les variations d'intensités de couleurs peu significatives affecterait l'estimation du critère de classification de zones d'intérêt.

La figure 4.18 illustre trois exemples de paires de zones similaires et différentes d'après la corrélation entre leurs histogrammes.

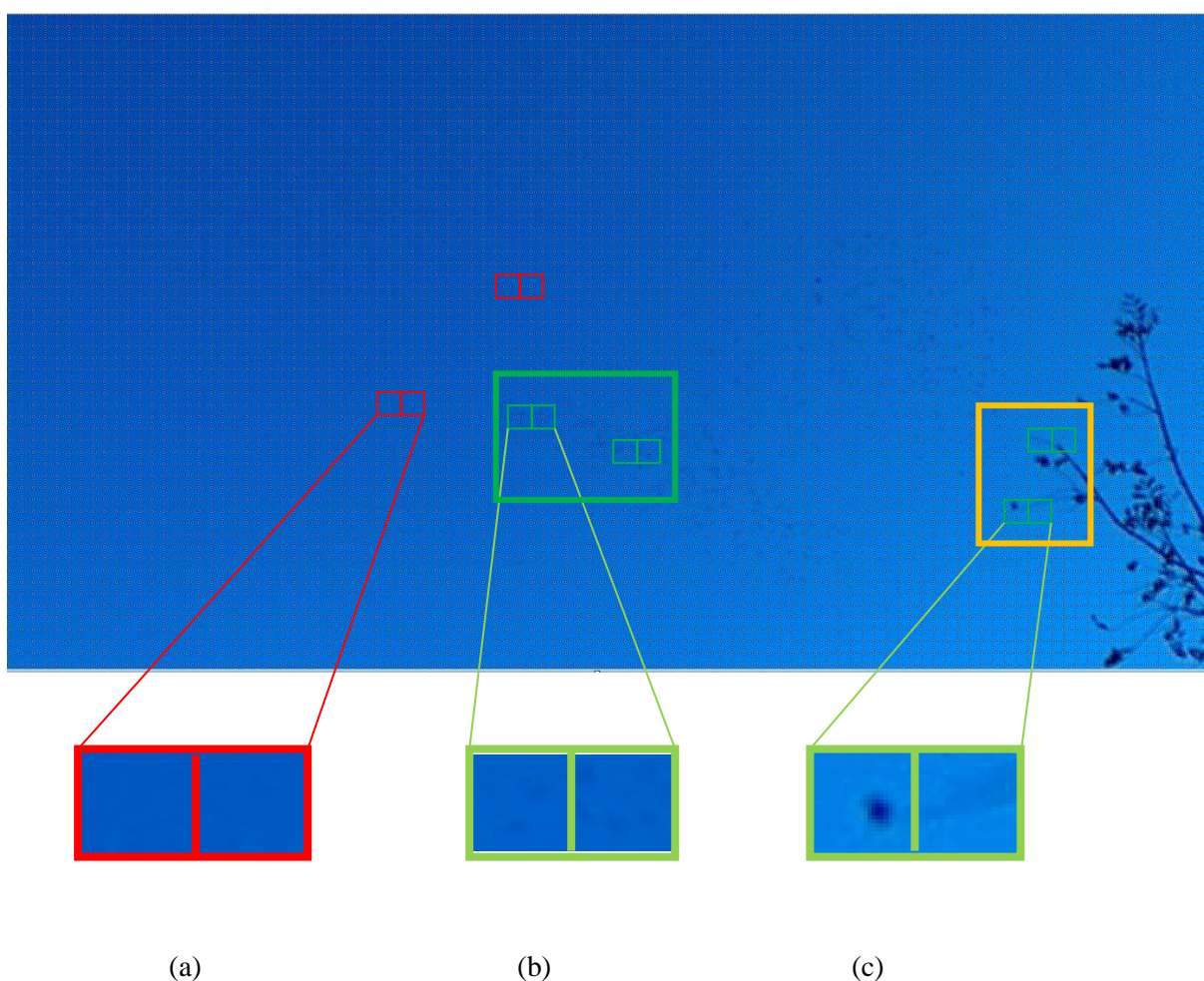


Figure 4.18 : Exemples d'importance d'application simultanée de mesures de corrélation et de distance  $\chi^2$  pour l'estimation de différences entre les zones d'images sur l'étape de prétraitement : (a) et la deuxième zone en rouge - pair de zones du fond sans objets d'intérêt visuellement identiques, avec la corrélation d'histogrammes proche de 1. (b) et (c) sont des paires de zones, pour lesquelles les valeurs de corrélation sont supérieures de 0. Les zones encadrées du carré vert (b) contiennent des objets d'intérêt difficilement visibles, et les zones dans le carré orange (c) – des zones avec des objets de différents contrastes. Dans les trois cas la distance  $\chi^2$  sera supérieure de 0, mais sans connaître son minimum et maximum, il ne sera pas possible de définir quelle valeur est associée à des zones avec des petits objets d'intérêt. La valeur de corrélation d'histogrammes sert d'une référence pour sélectionner de zones, où la distance  $\chi^2$  sera importante dans le prétraitement.

#### 4.4.2.5. Formalisation de la méthode d'estimation du seuil de variation d'intensité lumineuse de couleurs par l'analyse des histogrammes appliquée en prétraitement

Dans cette section on introduit les paramètres :

- $SM$  : mesure de similarité de deux zones, caractérisée par la corrélation entre leurs histogrammes
- $TUM$  : mesure d'uniformité de la texture d'une zone
- $Th_{Var}(Y)$  : seuil de variation d'intensité de couleurs

Ces paramètres sont utilisés ensuite à la première étape en appliquant les conditions suivantes :

- Une zone d'image est considérée comme homogène (sans objets d'intérêt), si  $TUM < Y_{texture}$
- Une zone d'image est considérée comme hétérogène (i.e. avec des couleurs différentes entre ces pixels, et donc avec des potentiels objets d'intérêt), si  $TUM \geq Y_{texture}$

L'algorithme suivant, dont le principe est basé sur les approches introduites dans les articles [27][33], formalise la méthode d'estimation et ajustement du seuil de variation d'intensité de couleurs :

- 1) *Calcul de la mesure de similarité pour chaque paire de blocs.* Dans l'image  $t$ , associée à un intervalle de  $T$  frames, la mesure de similarité  $SM$  est estimée pour chaque paire de blocs  $(B_{i,t}, B_{j,t})$  par la formule de distance de corrélation entre leurs histogrammes  $(H_{i,t}, H_{j,t})$  :

$$SM(B_{i,t}, B_{j,t}) = \frac{\sum_k (H(B_{i,t})(k) - \bar{H}(B_{i,t})) (H(B_{j,t})(k) - \bar{H}(B_{j,t}))}{\sqrt{\sum_k ((B_{i,t})(k) - \bar{H}(B_{i,t}))^2 \sum_k (H(B_{j,t})(k) - \bar{H}(B_{j,t}))^2}}$$

$$\text{où } \bar{H}(B_{i,t}) = \frac{1}{N} \sum_k H(B_{i,t})(k)$$

- 2) *Estimation des mesures de similarité de blocs minimale et maximale* ( $SM_{min} \in (0, 1)$  et  $SM_{max} \in (0, 1)$  respectivement) pour l'image donnée. Les histogrammes  $H_{i,t}$  et  $H_{j,t}$  (et par conséquent, les zones  $B_{i,t}$  et  $B_{j,t}$  dont ils caractérisent) sont considérés comme similaires, si la mesure de similarité  $SM$  correspondante est proche de 1. Ils sont considérés comme différents, si  $SM$  est proche de 0.
- 3) *Construction d'ensembles de blocs.* Pour une image  $t$  donnée, un ensemble de blocs avec leurs mesures d'uniformité de couleurs associées ( $TUM$ ), qui servira pour l'ajustement du seuil de variation d'intensité, est construite par la condition suivante :

$$SM(B_{i,t}, B_{j,t}) \in (SM_{min}, SM_{max})$$

L'ensemble ainsi construit contient les blocs, dont les contenus ne sont pas identiques (e.g. des zones avec des très petits objets ressemblants visuellement, mais différents au niveau de la présentation numérique de pixels qui les composent), et qui n'ont pas de différences considérables entre leurs contenus non plus (e.g. une zone avec une branche d'arbre d'une couleur proche du noir et une zone avec du ciel bleu).

- 4) Dans l'ensemble ainsi construit, la valeur initiale de  $Y_{texture}$  est définie comme variation maximale de composante  $Y$  dans la paire de blocs  $B_{i,t}$  et  $B_{j,t}$ , pour lequel la mesure de similarité est minimale :  $SM(B_{i,t}, B_{j,t}) = SM_{min}$ .

Cette étape permet d'estimer une valeur de variance  $Y_{texture}$  qui définira le seuil maximal des variations attendues pour des zones potentielles avec des petits objets d'intérêt. Cette valeur initiale de  $Th_{Var}(Y)$  est nommée par la suite  $Th_{Var}(Y)_{init}$ .

- 5) Pour estimer la valeur d'ajustement du seuil de variation initiale  $Y_{init}$ , la mesure quantitative de la distance de Chi-carré  $\chi^2$  entre les histogrammes est estimée pour chaque pair de zones  $(B_{i,t}, B_{j,t})$  tels que  $SM(B_{i,t}, B_{j,t}) \in (SM_{min}, SM_{max})$  :

$$\chi^2(H(B_{i,t}), H(B_{j,t})) = \frac{1}{2} \sum_k \frac{(H_{i,t}(k) - H_{j,t}(k))^2}{H_{i,t}(k) + H_{j,t}(k)}$$

- 6) Estimation du seuil de variation d'intensité : Finalement, le seuil de variation d'intensité de couleurs  $Th_{Var}(Y)$  pour un intervalle de  $T$  frames est estimé depuis sa valeur initiale  $Th_{Var}(Y)_{init}$  et la distance  $\chi^2$  moyenne dans l'ensemble de zones comme :

$$Th_{Var}(Y) = Th_{Var}(Y)_{init} - \overline{(\chi_2(B_{i,t}))}$$

où  $\overline{\chi^2(B_{i,t})} = \frac{1}{J} \sum_t^J \{\chi^2(B_{i,t}, B_{j,t})\}$  et  $J$  est le nombre de blocs dans l'ensemble.

La valeur  $Y_{texture}$  ainsi estimée est utilisée sur la première étape de la méthode principale de détection et dénombrement de petits objets d'intérêt, présentée dans le chapitre suivant.

#### 4.4.3 Première phase : Evaluation à priori de zones susceptibles de contenir des objets d'intérêt

La variation de couleurs  $Th_{Var}(Y)$  estimée pour les zones avec des potentiels objets d'intérêt, est utilisée en tant que seuil de variation sur la première phase d'évaluation de zones à priori pour la détection d'objets d'intérêt. L'objectif de cette première étape consiste en deux points principaux :

- Effectuer une classification de zones dans une vidéo, qui probablement contiennent des objets d'intérêt
- Estimer la probabilité, que les contenus de zones correspondent à des objets d'intérêt (i.e. d'évaluer la fiabilité de détectons dans ces zones)

Un autre avantage de l'évaluation à priori de zones est une réduction potentielle du volume de données à traiter. En pratique, pour détecter des petits objets, il est nécessaire d'analyser l'intégralité des pixels d'une image. Lorsque seule une petite partie de l'image peut contenir des objets d'intérêt (la figure 4.19 illustre la proportion entre le volume de données réel et utile), une évaluation des zones où de petits objets peuvent se trouver, peut réduire d'une manière importante le volume de données à traiter et permet d'optimiser la suite de traitement (les zones avec un coefficient nul ne sont pas traitées).



Figure 4.19 : Un exemple où la sélection de la zone utile sur la première phase de la solution permet d'éviter les traitements de zones vides (le ciel bleu) ou avec des objets d'autres natures que celle des objets d'intérêt (la végétation et le sol).

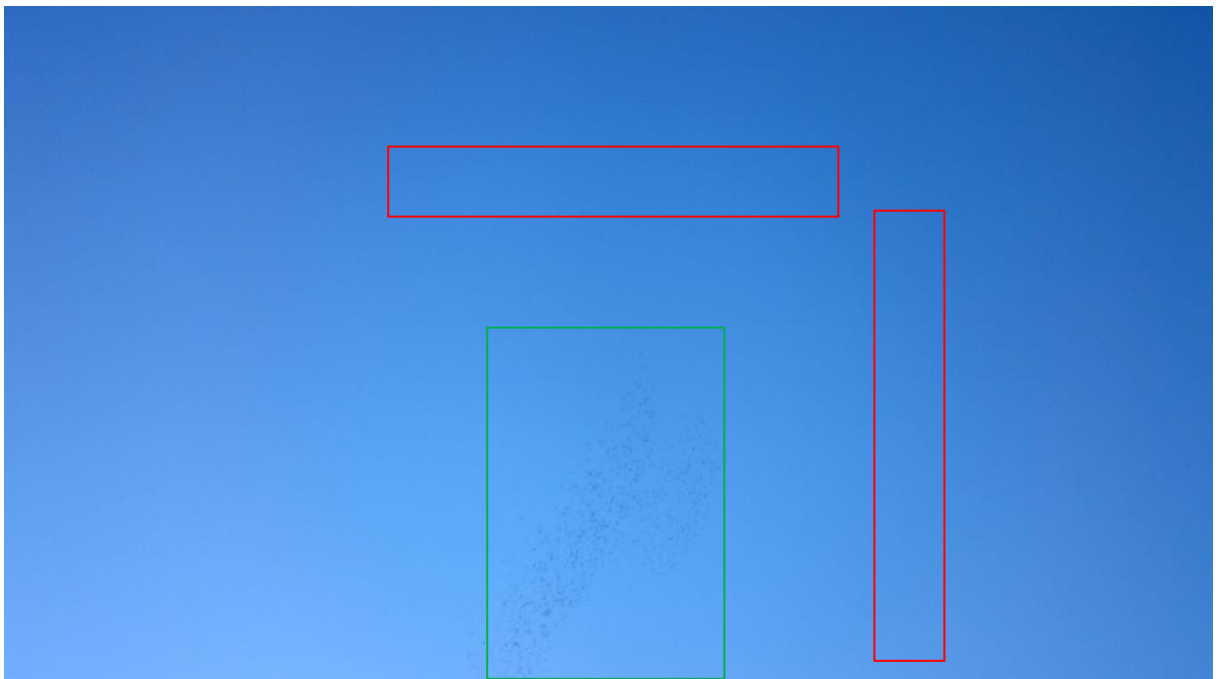


Figure 4.20 : Un exemple où l'évaluation de la zone utile (carré vert) sur la première phase de la solution en tant qu'une zone fiable pour les détections d'objets d'intérêt permet d'éviter les traitements de zones avec des très faibles variations d'intensités de couleurs, invisibles à l'œil, mais détectables par des méthodes numériques (carrés rouges).

#### 4.3.3.1 Description de la phase d'évaluation à priori de zones

L'article [21], qui a servi de base pour la conception de la première phase de la méthode proposée dans le cadre de cette thèse, se situe dans les domaines de la robotique et du traitement d'images. L'idée proposée dans [21] consiste à associer des zones d'images, où des fortes variances de couleurs sont observées, à des régions contenant des obstacles sur le chemin d'un robot, (elles sont donc à éviter par le robot). Dans le contexte de l'évaluation à priori des zones avec des objets d'intérêt, les zones avec ces variances, au contraire, sont à conserver pour les besoins des traitements suivantes. En d'autres termes, si dans [21] les zones avec des fortes variations de la composante Y sont les zones à éviter sur le chemin d'un robot, pour l'évaluation de zones avec des potentiels objets d'intérêt, se sont,

au contraire, les zones avec des variations élevées, qui doivent avoir des coefficients de fiabilité de détection élevés pour la suite des traitements.

Un seuil minimal de variation de la composante  $Y$  -  $Th_{var}(Y)$  - estimé lors du prétraitement, sert d'un critère de classification de ces zones en tant que zones d'intérêt avec des potentiels objets d'intérêt. Le traitement de la phase d'évaluation à priori de zones d'intérêt dans une image se base sur son organisation préalable en structure hiérarchique de type quadtree [21]. La figure 4.21 présente le schéma du processus d'organisation de zones d'image en structure hiérarchique de quadtree et la classification de ces zones par le critère basé sur la variance d'intensités de couleurs de pixels. La figure 4.22 illustre visuellement ce principe de segmentation.

La classification des zones d'intérêt sur la première étape d'exécution se fait par l'analyse globale de contenus de ses zones. A ce stade uniquement le fait que certaines zones d'une image peuvent contenir des objets sera pris en considération, sans spécification de la nature de ces objets. Ce concept permet une abstraction du contenu d'une vidéo, la sélection des objets des natures variées et la généralisation d'applications de la solution.

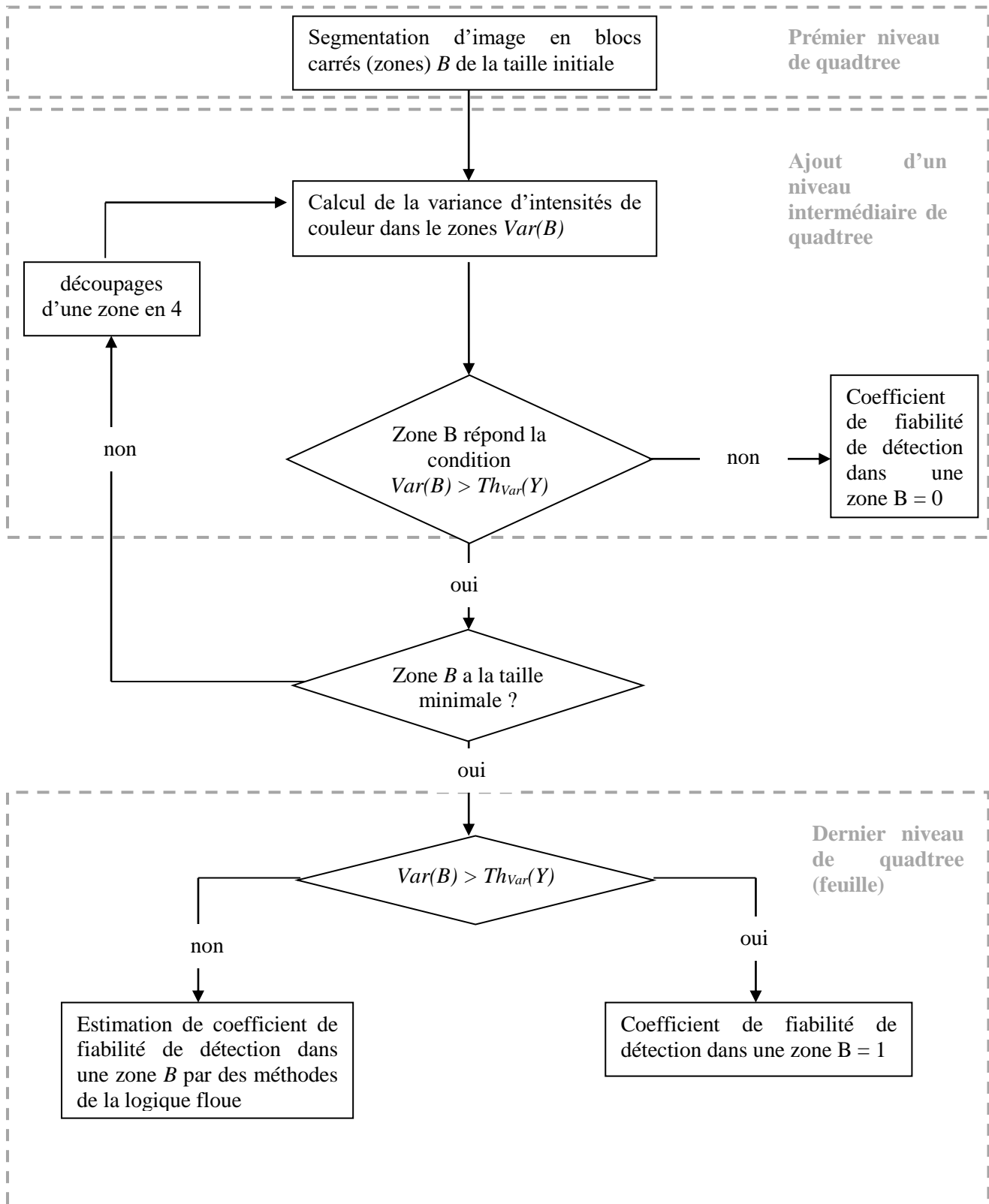


Figure 4.21 : Schéma du processus d'organisation de zones d'image en structure hiérarchique de quadtree et de la classification des zones par le critère de la variance d'intensités de couleurs de pixels.



Figure 4.22 : Exemple de segmentation et de sélection de zones d'une image (deux extraits de zones avec des branches d'arbre et une partie d'un vol d'oiseaux). Les zones, entourées des carrés rouges sont des zones, considérés globalement homogènes et ne sont pas retenues pour les traitements suivants. Celles, entourées des carrés verts, ont une variation d'intensités supérieures au seuil et elles sont analysées dans la suite de la solution proposée. Les extraits de la segmentation sur cette figure sont une partie de toute la segmentation du vol et de la végétation et un exemple typique de classification de zones en tant que homogènes ou hétérogènes.

#### 4.3.3.2 Coefficients de fiabilité sur la phase d'évaluation de zones d'intérêt

Comme cela est mentionné dans les sections précédentes, lors de la détection de petits objets la difficulté principale est de les différencier du bruit sur le fond d'image, si leur contraste y est faible ou les tailles sont petites. Dans le cas d'étude pratiquement toutes les vidéos sont enregistrées depuis une poste d'observation à 500-750 mètres d'objets observés, où ils sont difficilement détectables visuellement et numériquement. Dans le contexte de l'évaluation de zones par les valeurs de la variance d'intensités de couleurs de pixels, elles peuvent être légèrement inférieures au seuil  $Th_{var}(Y)$  sélectionné, mais toujours contenir des objets d'intérêt à peine visibles sur le fond. Le coefficient de fiabilité  $HSV_w(B)$ , qui intervient sur la première phase de l'algorithme, sert à caractériser les zones de ce type par la probabilité qu'elles contiennent des objets d'intérêt. Ce premier coefficient contribue à l'estimation des coefficients de plusieurs objets ou à la suppression d'une zone entière avec tous les objets qu'elle contient, puisqu'une zone peut contenir plusieurs objets d'intérêt. Son objectif est d'évaluer le contraste d'objets d'intérêt potentiels dans une zone sélectionnée. Plus loin dans le texte il est appelé « coefficient d'intensité de contraste ».

Des petits objets d'un faible contraste sur le fond ont souvent des composantes couleur dans l'espace RGB proches à celles du fond. Il est donc difficile d'en déduire un phénomène d'intensité de ces variations, qui pourrait caractériser une zone en tant qu'homogène ou hétérogène.

Par exemple, sur les figures 4.23 (a) et (b) toutes les moyennes de toutes les trois composantes dans l'espace colorimétrique RGB sont différentes, mais dans l'espace HSV deux sur trois sont très proches. Cela signifie que l'analyse dans l'espace RGB ou YUV, dont la composante Y sert pour classifier les zones d'intérêt ne présentent pas non plus une méthode fiable d'estimation de coefficients de fiabilité. La composante Y se calcule sous forme d'une combinaison linéaire de composantes dans l'espace RGB, et donc reflète les variations de composantes RGB, difficiles à analyser. Dans l'espace



HSV, quant à lui, des valeurs proches de la composant de teinte H signifiera des couleurs visuellement proches dans le même intervalle de de teintes. Dans ce cas il suffit à comparer les composantes S (saturation) et V (valeur) qui représente notamment l'intensité lumineuse d'une couleur. Si les valeurs de la composante H sont scientifiquement différentes (e.g. ne différence de 30 entre elles donne deux couleurs visuellement distinctes), cette composante seule permettra de différencier les objets d'intérêt sur le fond d'image (voir la figure 4.24).

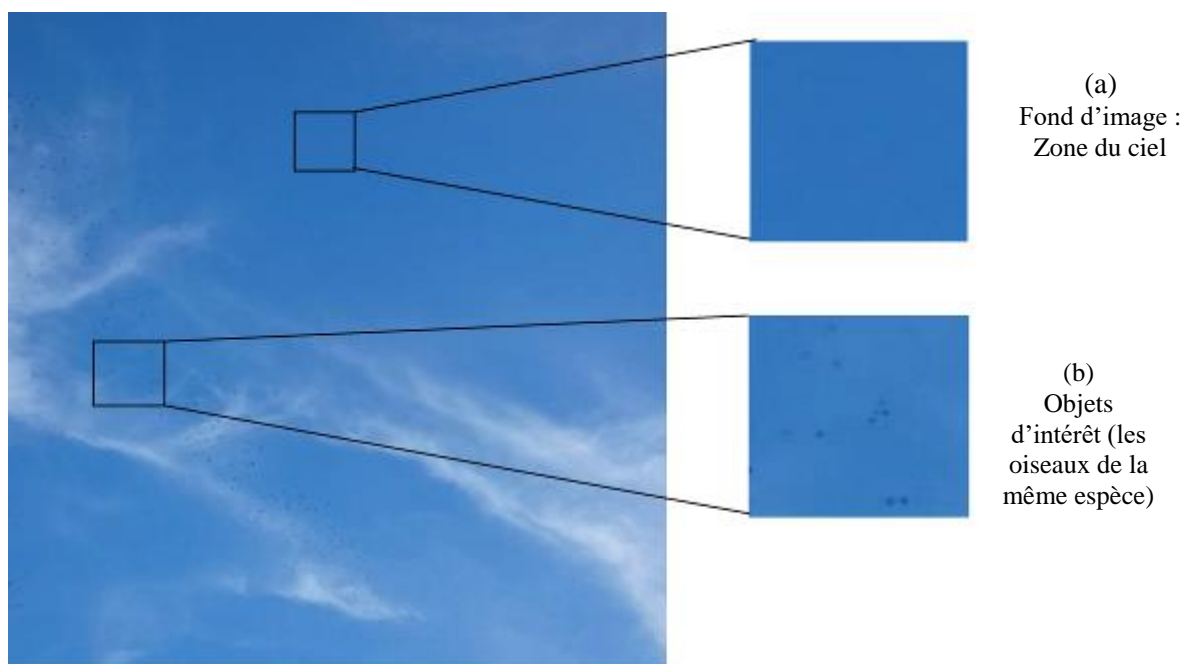


Figure 4.23 Exemples où la comparaison de couleurs en espace HSV permet de détecter des variations de couleurs dans les zones mieux, que dans l'espace RGB : (a) les composantes de la couleur dominante du fond sont (210, 72, 71) dans l'espace HSV et (51, 115, 180) dans l'espace RGB. (b) les composantes de la couleur d'objets d'intérêt sont (210, 70, 78) dans l'espace HSV et (62, 129, 200) dans l'espace RGB. Dans ces deux zones la teinte de couleur du fond et d'objets d'intérêt est la même, mais leurs intensités sont différentes (notamment, la valeur de saturation pour la figure (b) est inférieure à celle de (s), qui correspond aux couleurs plus foncées d'objets d'intérêt).



Figure 4.24 Exemples des couleurs avec différentes valeurs de composantes de teinte (30 pour l'orange, 0 pour le rouge), mais les mêmes pour les composantes de la saturation et la valeur (255 les deux). Cette figure explique, que deux couleurs peuvent se différencier visuellement et numériquement par les valeurs d'une seule composante dans un espace colorimétrique.

#### 4.3.3.3 Objectif et principe d'application du coefficient d'intensité de contraste dans une zone

De faibles variations des composantes RGB d'un pixel à l'autre comme, par exemple, dans le cas d'une transition entre deux couleurs dans une zone, contribuent au changement des composantes-couleur de pixels dans cette zone, même si elle ne contient pas toujours d'objets d'intérêt et paraît visuellement uniforme.

Le coefficient de fiabilité  $HSV_w(B)$  estimé à cette étape d'algorithme sert à résoudre cette ambiguïté et à évaluer la fiabilité de détection dans les zones d'intérêt. Il se base sur l'analyse simultanée dans l'espace YUV, qui est utilisé pour sélectionner des zones avec les variations d'intensités lumineuses de couleurs dans des zones, et HSV pour caractériser la nature de ces variations et d'estimer les coefficients de fiabilité de détection dans ces zones.

L'espace colorimétrique RGB est le plus utilisé pour la présentation numérique des couleurs, mais il ne convient pas pour les besoins d'analyse de l'intensité lumineuse de couleurs. Par exemple, une zone logiquement (et visuellement) homogène, i.e. sans objets d'intérêt, peut avoir des pixels avec des teintes de la couleur dominante du fond légèrement différentes, mais qui se trouvent toujours dans le même intervalle de teintes (e.i. des différentes nuances de bleu dans le ciel). Dans l'espace RGB cette situation s'exprimera dans des valeurs variées de composantes couleurs, qui générera des variations des intensités de couleurs. Ainsi les zones de ce type seront classifiées en tant que zones d'intérêt, quand elles devraient être considérées comme uniformes. L'avantage de la méthode proposée est d'utiliser des traitements dans plusieurs espaces colorimétriques, qui donnent à la sortie un coefficient de fiabilité de détection dans une zone et reflètent des différents aspects de couleurs dans des deux espaces colorimétriques définis par la Commission Internationale de l'Eclairage (CIE).

Dans le contexte de l'évaluation à priori des zones d'intérêt l'estimation de coefficients de fiabilité est intégrée dans les étapes d'évaluation de zones d'intérêt. Elle est schématiquement présentée sur la figure 4.25. Les étapes d'estimation du coefficient d'intensité de contraste dans une zone, s'enchaînent dans la séquence suivante :

- 1) Construction d'un ensemble de zones, où des variations d'intensité lumineuses sont observées et l'estimation de coefficient de fiabilité par rapport au seuil de la variance lumineuse dans une zone
- 2) Estimation de caractéristiques de la couleur dominante d'une zone : sa vivacité et sa teinte. La vivacité de la couleur est estimée sous forme de somme des rapports entre les composantes de saturation et de la valeur (où la valeur est privilégiée car elle exprime la brillance - l'intensité d'une couleur, qui est proche à la notion de vivacité) avec la formule suivante :

$$BR(B, bg) = \frac{1}{N_b} \sum_j^{N_B} \left( value(p_j) + \frac{1}{saturation(p_j)} \right)$$

La teinte globale  $HUE(B, bg)$  d'une zone est estimée par le même principe :

$$HUE(B, bg) = \frac{1}{N_b} \sum_j^{N_B} hue(p_j)$$

où  $BR(B, bg)$  définit la vivacité de la couleur dominante dans la zone  $B$ ,  $p_j$  est un pixel de la zone et  $N_B$  est le nombre de pixels dans la zone  $B$ .

- 3) Le contraste de teinte de la couleur théorique d'objets et celle de la couleur dominante d'une zone est estimé en fonction de la répartition de teintes dans les intervalles du cercle de couleurs de Munsell [40] et de la quantité fixe des teintes similaires contenues dans un intervalle (égale à 30 par définition du cercle de couleurs) :

$$Contrast(HUE(B, bg)) = \begin{cases} \frac{|HUE(B, bg) - HUE_{object}|}{30}, & \text{si } |HUE(B, bg) - HUE_{object}| < 30 \\ 1, & \text{sinon} \end{cases}$$

Si la teinte de la couleur dominante dans une zone  $HUE(B, bg)$  et la teinte de la couleur théorique d'objets se trouvent dans des intervalles du cercle de couleurs différents, les couleurs associées seront visuellement distinctes. Par conséquence, elles seront numériquement différentes dans l'algorithme. Dans ce cas, même si les deux couleurs ont la même vivacité, le contraste de leurs teintes sera détecté et compris dans l'estimation du coefficient d'intensité de contraste pour une zone d'intérêt. Le contraste entre les teintes est estimé sous forme de la valeur de leurs différences par rapport à la quantité de teintes similaires, contenues dans un intervalle par la définition du cercle de couleur de Munsell (cette quantité est égale à 30). Le contraste de vivacité de couleurs  $Contrast(BR(B, bg))$  est estimé par rapport à sa valeur maximale 255, où  $BR(B, bg)$  est l'intensité lumineuse de la couleur dominante de la zone  $B$  et  $BR_{object}$  est l'intensité lumineuse de la couleur théorique d'objet :

$$Contrast(BR(B, bg)) = \frac{|BR(B, bg) - BR_{object}|}{255}$$

- 4) Dans le contexte de l'hypothèse sur l'apparence de petits objets, leur contraste sur le fond peut être faible et la couleur proche de celle du fond. Par conséquent, la variation de teinte de la couleur dans une zone et de son intensité lumineuse sont également importantes pour estimer le coefficient de fiabilité d'une zone pour les traitements suivants (i.e. si une des deux valeurs est suffisamment élevée, les détections dans cette zone d'intérêt seront fiables). En d'autres termes les objets qui seront détectés dans des zones fiables, seront très probablement des objets d'intérêt. Le coefficient d'intensité de contraste est estimé en fonction des valeurs de ratios de deux contrastes et de leur proximité à la limite de visibilité, exprimée par la valeur 0.5 de deux contrastes. Puisqu'il est important de conserver les objets d'intérêt avec un faible contraste sur le fond, cette approche permet de ne pas trop baisser le coefficient d'intensité, quand il s'approche à la limite de visibilité. L'estimation de coefficient se fait par le système suivant, qui applique la fonction d'appartenance de la logique floue : la fonction gaussienne en espace 2D ou un ratio, en fonction d'intensité de contraste de teintes de couleur et de leurs luminosités :

$$HSV_w(B) = \begin{cases} \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\left(\frac{(x-\mu_x)^2}{2\sigma_x^2} + \frac{(y-\mu_y)^2}{2\sigma_y^2}\right)\right] & \text{si } Contrast(HUE(B, bg)) \leq 0.5 \\ & \text{et } Contrast(BR(B, bg)) \leq 0.5 \\ \frac{Contrast(HUE(B, bg)) + Contrast(BR(B, bg))}{2}, & \text{si } Contrast(HUE(B, bg)) \leq 0.5 \\ & \text{ou } Contrast(BR(B, bg)) \leq 0.5 \end{cases}$$

où  $x=Contrast(HUE(B))$  et  $y=Contrast(BR(B))$  sont les valeurs de contrastes des teintes et de l'intensité lumineuse d'une zone respectivement,  $\mu_x$  et  $\mu_y$  sont leurs valeurs moyennes dans une zone, fixées à 0.5;  $\sigma_x$  et  $\sigma_y$  sont leurs déviations standards dans la même zone, fixées à 0.2.

#### 4.3.3.4 Expérimentations et analyse de la méthode d'évaluation de zones d'intérêt et du coefficient de fiabilité de détections dans une zone

Cette section présente des exemples d'évaluation des zones d'image et d'estimation de coefficients de fiabilité de détection d'objets d'intérêt dans ces zones, présentés ci-dessus. La figure 4.26 regroupe trois situations :

- 1) Le fond d'une zone est homogène et ne contient pas d'objets intérêt
- 2) Le fond d'une zone est texturé et ne contient pas d'objets intérêt
- 3) Le fond d'une zone est homogène et contient des objets intérêt

Les cinq images (a) – (e) de la figure 4.26 sont les zones avec la même teinte de couleur pour l'intégralité de pixels : celle du ciel et celle des couleurs des oiseaux, représentés par des groupes de pixels plus sombres. Dans ce cas l'intensité lumineuse globale d'une zone est définie par la majorité de pixels du fond bleu clair. La couleur d'objets dans ces zones est suffisamment contrastée de celle du fond pour contribuer à la variation d'intensité de couleurs, caractéristique pour des zones avec des objets d'intérêt.

La figure 4.26 (a) est un exemple d'une zone de ciel sans objets d'intérêt, avec une faible variation d'intensité lumineuse de couleurs de pixels. Cela signifie un contenu homogène dans cette zone, i.e. un fond uni. Contrairement à la figure 4.26 (a), la figure 4.26 (d) contient deux oiseaux très éloignés de la caméra, et très peu visibles à l'œil. En pratique les zones avec des objets de ce type se caractérisent par des faibles variations d'intensités de couleurs, au-dessous du seuil estimé. Il est alors difficile de les évaluer uniquement par le critère de variation d'intensité lumineuse sans en perdre, si le seuil est trop élevé, ou sans en retenir trop, s'il est trop bas. L'attribution d'un coefficient de fiabilité pour une zone permet de caractériser la fiabilité de détection d'objets pour la suite de la séquence des traitements. Les objets, qu'elles contiennent ne sont pas considérés comme des détections fiables. Ainsi, dans le cas de fausses détections leurs contribution dans le résultat final sera graduellement diminuée sur les étapes suivantes de la solution.

L'exemple sur la figure 4.26 (d) est une illustration, où la variation d'intensité lumineuse dans une zone est très proche à celle dans la zone vide sur l'exemple de la figure 4.26 (a). Les zones vides (telles que sur la figure 4.26 (a)) auront le coefficient de fiabilité égal à 0, mais les zones avec des objets peu contrastés sur le fond (la figure 4.26 (d)) seront conservées avec un coefficient dans l'intervalle (0, 1), caractérisant la fiabilité de détection d'objets d'intérêt dans une zone. Ainsi cette approche permet d'éviter une perte d'objets d'intérêt, ayant un faible contraste, et en même temps de supprimer implicitement les zones réellement vides par le coefficient de fiabilité proche ou égal à 0.

Les exemples sur les figures 4.26 (b), (c) et (e) sont les zones d'intérêt, ou la détection d'objets est fiable. Elles sont caractérisées par un coefficient de fiabilité proche de 1. Leurs combinaisons avec les coefficients de fiabilité sur les étapes suivantes contribueront au comptage d'objets détectés par des valeurs, également proches de 1 et caractérisant des détections finales fiables.

Les exemples sur la figure 4.26 (k) – (o), représentent les zones, qui ont des fonds hétérogènes, avec des couleurs différentes, qui contribuent à une variation d'intensité supérieure au seuil estimé. Mais elles ne répondent pas à l'hypothèse dans le contexte d'application de suivi de migrations d'oiseaux, qui se produisent dans le ciel à une certaine hauteur. Cette hypothèse permet de supposer que le fond attendu pour une zone avec des objets d'intérêt a une couleur du ciel relativement homogène. Sur l'arrière-plan des zones sur les figures 4.26 (k) – (o) des parties du sol et de la végétation occupent la majorité de la zone, qui signifie une certaine proximité du sol. Les zones de ce type ont des coefficients de fiabilité de détection bas pour la suite de traitement, car d'après l'hypothèse ne peuvent pas s'y produire.

Les figures 4.26 (f) – (g) présentent un ensemble de zones, pour lesquelles leurs coefficients signifient des faibles détections. Mais les formes apparentes d'objets détectés sont différentes de celles d'objets d'intérêt (e.g. comme celles sur les figures 4.26 (b), (c), (e)). Cet exemple illustre la situation, où le coefficient de fiabilité pour une zone, appliqué seul, ne sera pas sélectif, car il ne fait pas de différence entre les contenus de zones, mais définit uniquement, si un objet peut y être détecté. L'évaluation de zones nécessite alors d'être affinée, afin de localiser les valeurs de coefficients de fiabilité aux objets concrets pour les évaluer individuellement, d'où l'intérêt d'estimer les coefficients de fiabilité sur les étapes suivantes et les combiner avec ceux des étapes précédentes.

L'évaluation a priori de zones sur la première phase de la solution dans le contexte de détection de petits objets dans une vidéo a plusieurs avantages :

- 1) Réduction importante du volume de données à traiter par la suite, grâce à la rétention uniquement de zones, avec de potentiels objets d'intérêt
- 2) Estimation de la fiabilité de détections d'objets dans ces zones, qui a un effet important sur le la suite de traitements
- 3) Possibilité de conserver les zones avec une faible probabilité de contenir des objets d'intérêt, qui permet de le confirmer et ajuster les traitements suivants ou supprimer finalement cette zone

Les limites principales de la première phase de la solution sont liées principalement à la segmentation d'images en zones pour l'organisation de la structure de quadtree :

- 1) Un petit objet à la limite de plusieurs zones est séparé en plusieurs parties et contribue moins à la variation d'intensité de couleurs de pixels de chacune. Ainsi, cette variation peut être insuffisante pour classer une zone comme hétérogène et contenant des objets d'intérêt. Aucune de zones ne sera dans ce cas caractérisée par un coefficient de fiabilité élevé, qui pourrait résulter en perte d'un objet.
- 2) Des petites parties d'autres objets dans une scène, qui apparaissent dans des zones en raison de segmentation peuvent être considérées comme des objets d'intérêt, si leurs formes apparentes sont similaires à ceux derniers (e.g. Figure 4.26 (f), (e))

Le coefficient de fiabilité d'évaluation de zones d'intérêt a pour l'objectif pas seulement de caractériser, à quel point les contenus des zones correspondent à potentiels objets d'intérêt. Il contribue aussi aux les étapes suivantes de la solution et à l'estimation de leurs coefficients de fiabilité correspondants. Au fur et à mesure d'exécution, les coefficients de fiabilité composent une combinaison de valeurs, où la valeur finale sera appliquée sur l'étape de dénombrement. Dans ce cas plusieurs situations sont possibles :

- La variation d'intensités de couleurs de pixels d'une zone peut être proche du seuil minimal admis, i.e. la fiabilité de détection d'objets d'intérêt dans cette zone sera basse. Mais localement, au niveau d'analyse de différences entre les pixels-voisins, leur contraste peut être suffisant pour repérer les contours d'objets. Si par la suite après la conversion en régions

continues sa forme est conforme au modèle théorique, cet objet sera attribué à l'ensemble de détections correctes.

- La variation d'intensités de couleurs des pixels d'une zone peut être largement supérieure au seuil de variation d'intensités de couleur, avec des contours nets. Dans ce cas, si les contours, ni la région englobée ne décrivent pas une forme similaire au modèle d'objet d'intérêt, le coefficient de fiabilité à l'étape correspondante sera bas, i.e. classifera l'objet concerné en tant qu'une fausse détection.

Dans ces deux situations un seul coefficient de fiabilité ne suffira pas pour construire un ensemble d'objets d'intérêt sans un nombre important de fausses détections. Des autres traitements locaux sont alors conçus pour compléter le coefficient de détections de zones potentielles avec des objets d'intérêt.

La composition de tous les coefficients, associés aux étapes de la solution, permet au fur et à mesure de confirmer la fiabilité de détection dans une zone au moyen de coefficients de fiabilité suivants, ou de la réfuter.

Les résultats des tests sur des vidéos de différentes complexités sont présentés dans le Tableau 4.1. Une baisse du nombre de fausses détections après l'application du coefficient de fiabilité sur les zones d'image justifie l'intérêt d'application de ces coefficients de fiabilité :

1. Sur 559 détections sans coefficients de fiabilité pour 68 objets réellement présents dans les vidéos du corpus 1, l'attribution de coefficients de fiabilité aux zones permet d'approcher le nombre de détections à la vérité terrain, et en retenir 66. Les objets, qui ont obtenu des coefficients de fiabilité dans l'intervalle (0, 1) donnent en somme 261.2. Le reste d'objets a été éliminé du dénombrement final par les coefficients égaux à 0.
2. Pour les vidéos du corpus 2, sur 816 détections sans coefficients de fiabilité pour 40 objets en vérité terrain, l'attribution de coefficients de fiabilité aux zones permet de garder 32 objets en tant que des objets d'intérêt. Parmi tous les 816 objets, détectés dans ces vidéos d'une complexité élevée, le nombre de 617.6 représentent la somme de coefficients de fiabilité compris entre 0 et 1

Les sections suivantes donnent également des exemples d'application de coefficients les uns après les autres, dans le contexte de leurs interactions consécutives.

Tableau 4.1 : Tous les coefficients de fiabilité appliqués séparément pour le calcul final de détections

Corpus	Vérité terrain	Détections avec un de coefficient coefficients								
		sans C(O)	HSV <sub>w</sub>		DR		MR <sub>1</sub>		MR <sub>2</sub>	
			TW	FW	TW	FW	TW	FW	TW	FW
1	68	559	65.8	261.2	59.2	130.7	44.2	294.6	48.3	260.5
2	40	816	32	617.6	16	326.4	12.5	346.8	8	285.6

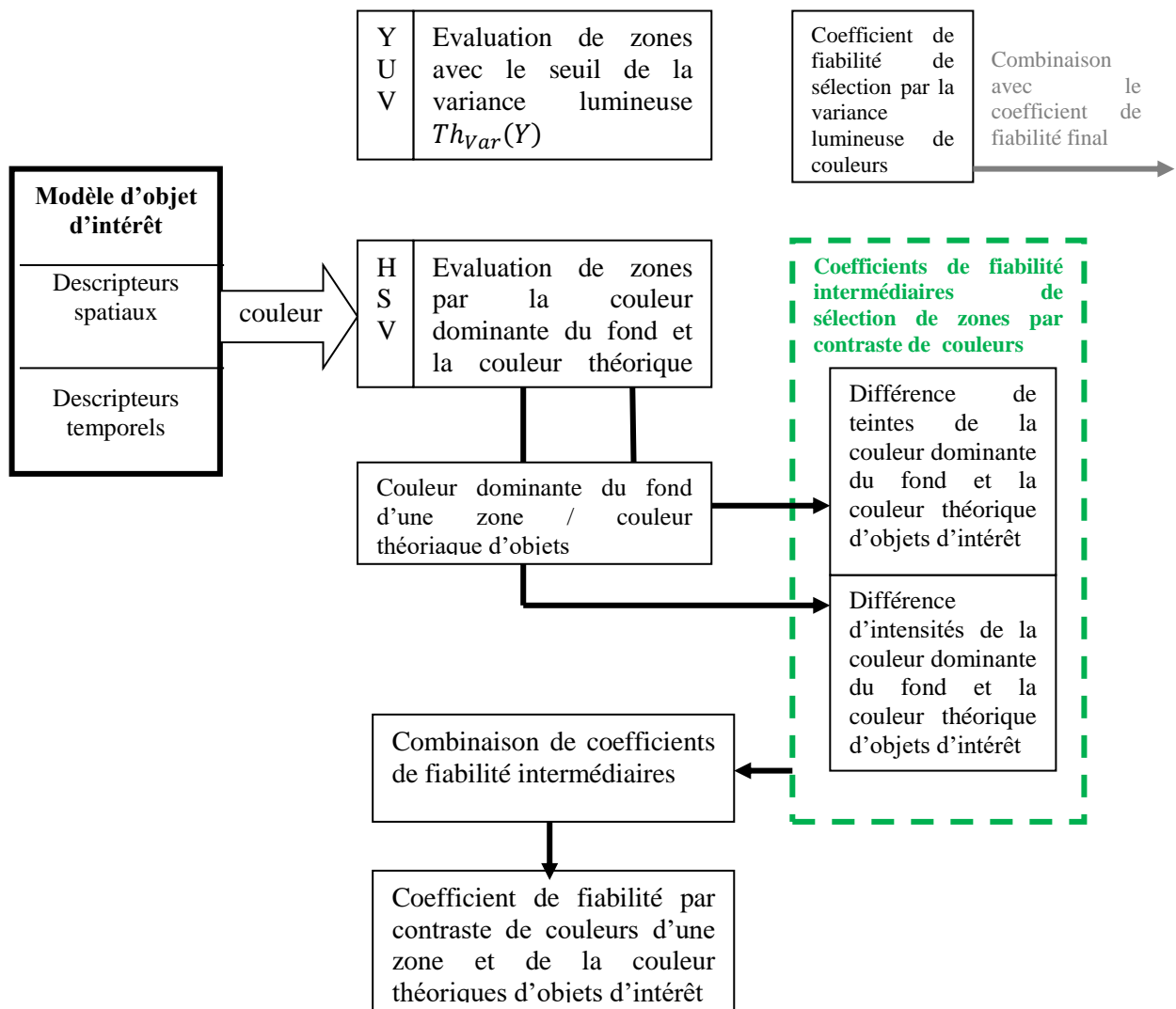


Figure 4.25: Principe d'estimation du coefficient d'intensité de contraste d'objets d'intérêt dans une zone

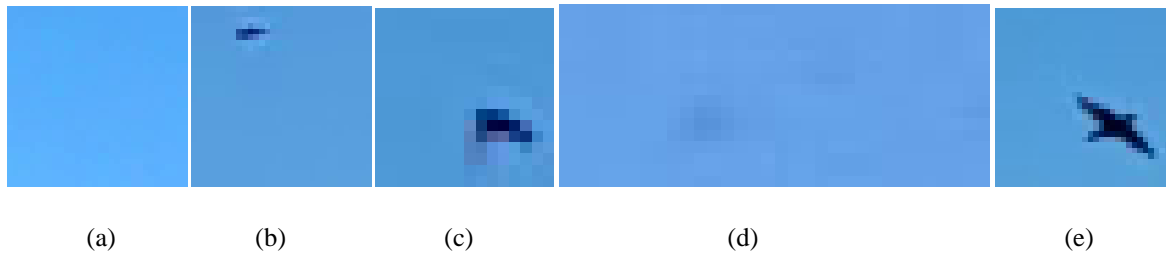


Figure 4.26 : (a) une zone vide avec le coefficient de fiabilité 0 ; (b) – (e) les types de zones fiables avec des coefficients de fiabilité 0.8-0.9. L'image (d) représente le type de zones, où globalement le contraste d'objets d'intérêt sur le fond est faible. Les coefficients de fiabilité sont estimés dans ce contexte et permettent de caractériser ces zones en tant que fiables pour les détections d'objets d'intérêt. La zone sur l'image (d) a le coefficient de fiabilité 0.87.

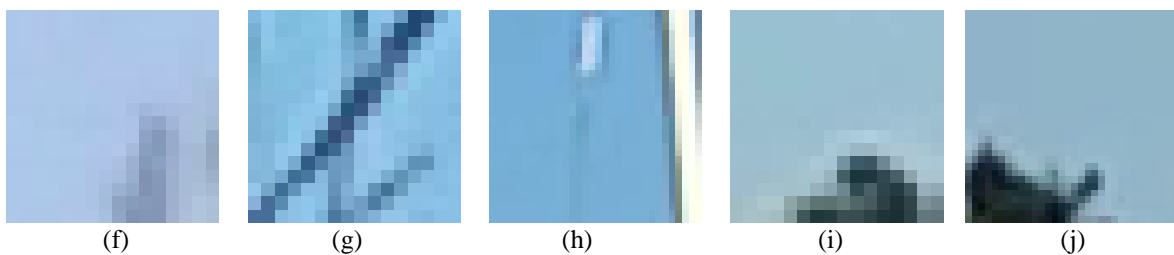


Figure 4.26 : des fausses détections qui génèrent des variations d'intensité dans les zones et contribuent à leur classification en tant que des zones avec des objets d'intérêt. Les formes des objets sur les images (g) et (h) ne correspondent pas aux formes apparentes d'objets, attendues pour les cas de suivi de migrations. Ces objets seront éliminés dans la suite de traitements. Les objets sur les images (a), (i) et (j) ont des formes apparentes similaires à celle des objets d'intérêt pour le cas de suivi de migrations. Ils seront traités sur les étapes suivantes, mais auront des différents coefficients de fiabilité. En tant que zones d'intérêt ces zones ont des coefficients 0.8 – 0.99.



Figure 4.26 : Les zones, qui obtiennent de coefficients de fiabilité bas pour la suite de traitements. Sur les images (k) – (m) la luminosité globale est proche de celle d'objets typiquement observés dans les vidéos (e.g. comme sur l'image (b)). Sur les images (n)-(o) les couleurs dominantes du fond ne correspondent pas à l'hypothèse que la couleur du fond devrait être celle du ciel. En fonction de la teinte de couleur, qui caractérise globalement une zone, ces coefficients de fiabilité varient de 0 pour les zones (n) et (o) à 0.1 pour la zone (m) et 0.2 – 0.3 pour les zones (l) et (k) respectivement.

#### 4.3.3.5 Conclusion sur l'application des coefficients de fiabilité pour les zones d'image

La classification de zones d'intérêt au moyen de coefficients d'intérêt est un premier exemple d'apport de ce principe : le nombre de fausses détections baisse d'une manière significative, si les objets détectés obtiennent une première classification des coefficients de fiabilité, attribués aux zones les contenant.

L'application du premier coefficient de fiabilité pour les zones d'intérêt diminue globalement le nombre de fausses détections, mais l'application d'un seul coefficient de fiabilité sur les zones n'est pas suffisante. Des différentes zones d'image potentiellement contiennent plusieurs petits objets et des



traitements locaux sont nécessaires pour analyser les objets un par un afin d'avoir un ensemble de détections proche à la vérité terrain.

#### 4.4.4 Deuxième phase : Détection d'objets dans les zones sélectionnées

Le traitement après la sélection de potentielles zones d'intérêt, où la présence d'objets d'intérêt est la plus probable, consiste en la sélection des objets eux-mêmes. Pour l'objectif de détection et de dénombrement de petits objets la détection de leurs contours est une étape importante, qui permet de les repérer sur le fond, les séparer les uns des autres et aura alors un effet direct sur le résultat final du dénombrement.

Dans le cas de suivi de migrations d'oiseaux depuis des postes d'observation, les vidéos ont des particularités, qui compliquent la tâche de détection d'objets dans une zone. Notamment ce sont des petites tailles d'objets, des localisations proches les uns à des autres et souvent un faible contraste sur le fond d'image. Dans ces conditions les méthodes existantes dans le domaine de détection de contours résultent en nombreuses fausses détections (la Figure 4.15 (g) est un exemple typique, où l'application du filtre Laplacien sur des petits objets avec un faible contraste sur le fond n'améliore pas suffisamment la netteté de représentation de ces derniers). Afin de prendre en considération ces difficultés identifiées la solution proposée applique une méthode, basée sur les notions de la statistique et du domaine de filtrage d'images. La présente section explique le déroulement et les bases théoriques de la méthode appliquée.

##### 4.4.4.1 Bases théoriques de la méthode de détection de contours

La deuxième phase de détection d'objets dans les zones sélectionnées débute par la détection de leurs contours. Typiquement, en raison des petites tailles d'objets une zone peut en contenir plusieurs. Il peut s'agir d'objets d'intérêt mais aussi d'objets de types différents, dont la détection est non seulement inutile, mais négative car générant des fausses détections. L'intensité de contraste d'objets sur le fond ainsi que leurs positions les uns par rapport aux autres ne peuvent pas être prédéfinies non plus en raison d'un impact considérable des changements dans une scène sur les formes apparentes d'objets dans une image. Dans ces conditions la correspondance de formes logiquement ne permettrait pas de sélectionner suffisamment d'objets d'intérêt à cause de possibles variations de leurs formes apparentes. La détection de contours de tous les objets, de toutes les formes et avec de contrastes plus et moins nets a été alors choisie comme une solution qui permettrait d'éviter les possibles pertes d'objets d'intérêt, comme dans le cas de leur sélection au moyen de correspondance de formes [10][11].

Des nombreuses techniques existent pour la détection de contours (voir la section 2.2.3 de l'état de l'art, qui les présente plus en détails) et des différents travaux de recherche sont dédiés à la comparaison de leurs avantages et inconvénances. Les travaux [25][26] en particulier souligne la sensibilité de l'opérateur Laplacien au bruit et aux faibles différences entre les pixels par rapport aux autres opérateurs de détection de contours. Un inconvénient pour des applications, où les objets sont nettement distinguables sur le fond d'image, cette particularité devient un avantage pour le cas de détection de petits objets peu contrastés sur le fond d'image (voir la figure 4.15). L'opérateur, ou le filtre, Laplacien permet de révéler des faibles différences entre les couleurs des pixels voisins, en les traitant sous forme de valeurs d'intensité de couleurs dans le contexte de la sélection de zones d'intérêt à l'étape précédente. En revanche, cela complique la sélection de contours : des contours désignant un objet ou le phénomène de numérisation de couleurs peuvent être similaires dans le masque de contours, généré par l'opérateur Laplacien, e.g. comme sur la figure [4.34]. De plus, moins les objets sont contrastés sur le fond d'image, plus leurs contours, définis par l'opérateur Laplacien seront proches de « faux » contours, qui apparaissent à cause des phénomènes suivants :

- des passages entre les couleurs dans une image
- des nuances de la couleur dominante dans une zone visuellement homogène, e.g. des légères variations de la composante de teinte de couleur, invisibles à l'œil

La présence dans une image de ces deux phénomènes génère des confusions dans les contours de petits objets et crée des faux contours, et donc des fausses détections dans le résultat final. Pour cette raison à l'étape de la détection de contours d'objet le filtre Laplacien est complété par l'analyse statistique de la distribution de pixels dans une zone traitée. La section suivante de ce chapitre introduit ces méthodes de la statistique.

#### 4.4.4.2 Notions de statistique descriptive et de paramètres de distribution

Le but de l'application de la statistique descriptive pour l'analyse de distribution d'intensités de couleurs de pixels dans une zone lors de la détection de contours par le filtre Laplacien consiste en la conception d'un moyen de son adaptation à une zone d'intérêt précise et le contraste sur le fond d'objets qu'elle contient. L'adaptation des traitements en fonction de zones est justifiée par l'impact sur la détection des paramètres du fond d'image, variables d'une zone à une autre. La Figure 4.30 est une illustration d'une situation, quand des légères hétérogénéités sur un fond du ciel bleu globalement uniforme – telles que les rayons et les reflets du soleil – peuvent perturber la détection de petits objets.

Par nature, une image représente un ensemble de pixels - d'échantillons dont la distribution peut être analysée par des méthodes statistiques. Les paramètres de distribution permettent de décrire un ensemble de données qu'elle représente et d'observer les phénomènes propres à cette distribution. Dans le contexte de détection de contours d'objets avec l'opérateur Laplacien ces paramètres de distribution servent à adapter son noyau standard de convolution en fonction de contraste d'objets sur le fond d'une zone.

La caractéristique d'asymétrie de distribution (« skewness » en anglais, [1]) est le paramètre de distribution principal, utilisé dans la méthode d'adaptation du filtre Laplacien développée ici. Il permet d'exprimer, à quel point la fonction d'une distribution est décalée de la moyenne. Le signe de la valeur d'asymétrie détermine la direction du décalage :

- positif si la fonction est décalée à gauche de la moyenne
- négatif dans le cas d'un décalage à droite.

La formule statistique suivante estime le degré d'asymétrie sur la base de la moyenne d'une distribution :

$$skewness = \frac{\sqrt{n(n-1)}}{n-2} \left[ \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}} \right]$$

L'opérateur Laplacien crée des masques de contours d'objets, et naturellement, sur un fond pratiquement homogène les petits objets ne génèrent que peu de contours. Dans l'ensemble des échantillons des pixels dans un masque de contours, les contours représentent le plus petit ensemble. Mais ils représentent aussi les maxima locaux du gradient, repérés par l'opérateur Laplacien, et possèdent alors les valeurs les plus élevées dans le masque de contours en noir et blanc. Ces deux phénomènes observés dans la majorité de zones analysées permettent de définir une forme attendue de la fonction de distribution d'intensités de pixels dans un masque de contours (voir la Figure 4.31 pour l'exemple d'un masque de contours pour une zone et de la forme de la fonction de distribution de ses pixels).

La modélisation de la fonction de distribution pour des zones d'intérêt est faite au moyen de construction de l'histogramme d'intensités de pixels dans un masque de contours. Influencée par les deux phénomènes, identifiés ci-dessus, l'histogramme aura le plus de pixels ayant des intensités proches du zéro, et sa fonction de distribution correspondante sera décalée à gauche. Ainsi l'importance de ce décalage peut être caractérisé par la valeur positive de l'asymétrie : plus elle et

élevée, plus il y a des pixels à des intensités proches du zéro et moins il y en a des ceux, qui représentent les contours d'objets. La valeur, qui caractérise l'asymétrie de la fonction de distribution est un paramètre essentiel dans les calculs et les formules conçus pour l'adaptation du noyau de l'opérateur Laplacien, présentés dans la section suivante.

#### 4.4.4.2 Méthodes de détection des contours

Pour prendre en considération les difficultés constatées dans le cas de détection de contours de petits objets, l'opérateur, proposé ici ne se limite pas uniquement à la détection de contours. Son objectif est de renforcer le contraste de contours de petits objets dans l'image et d'y atténuer les fausses détections au moyen d'une analyse statistique des zones d'intérêt. Cette solution innovante est conçue comme une technique de sélection de contours implicite, lors de l'application de l'opérateur. Cette approche permet de supprimer ou d'atténuer les faux contours et d'éviter une étape supplémentaire de sélection de contours d'objets. Le résultat de cette étape est important pour la suite des traitements, car il sert de point de départ pour la détection des objets d'intérêt.

La détection de contours dans la solution proposée s'exécute en plusieurs étapes :

1. *Application de l'opérateur Laplacien standard sur une zone d'intérêt* : permet de repérer les différences entre couleurs des pixels voisins. Ces différences sont reflétées sous forme d'un masque de contours en niveaux de gris. A cette étape tous les contours dans une zone d'image sont détectés.
2. *Construction de l'histogramme de distribution de pixels* avec des niveaux de gris dans le masque de contours (qui représente une image en niveaux de gris) et son approximation à la fonction de distribution Gaussien.
3. *Estimation de caractéristiques statistiques* (la moyenne, la distribution standard et l'asymétrie) de la fonction de distribution représentée par cet histogramme
4. *Estimation de paramètres pour la matrice de convolution* adaptée de l'opérateur Laplacien (le facteur de renforcement de contours  $F_{skewness}$  et du niveau de gris minimal  $\alpha$ , qui caractérise les contours d'objets) depuis les caractéristiques statistiques de la distribution, estimées à l'étape précédente.
5. *Construction de la matrice de convolution du filtre adapté*, dérivée des caractéristiques statistiques de la distribution de pixels dans le masque de contours standard.
6. *Application de la version adaptée de l'opérateur Laplacien sur la zone initiale* (i.e. une image en couleur).

La conception de cette méthode a été inspiré par le concept de « déplacement de contours », présenté dans l'article [26]. Dans le cas de difficultés liées à la détection de contours de petits objets, proches les uns aux autres et souvent peu contrastés sur le fond d'une image, le « déplacement » de contours détectés s'exécute sous forme de renforcement des plus nets entre eux – ceux qui sont le plus probablement attribués aux contours d'objets, et d'atténuation des moins contrastés – de ceux qui probablement correspondent aux faibles variations des teintes de couleurs, typiquement situées près de vrais contours d'objets (voir la figure 4.28 pour un exemple).

Pour « déplacer » les contours l'opérateur Laplacien adaptatif conçu applique une matrice de convolution calibrée par les paramètres statistiques estimés. Cela renforce le contraste des contours d'objets dans l'image et diminue l'intensité de faux contours représentant les plus faibles différences de pixels dans une image. La figure 4.27 présente la matrice de convolution de l'opérateur Laplacien standard et la matrice calibrée par les paramètres statistiques estimés. Le schéma 4.29 résume toute l'étape de détection de contours de petits objets, proposée dans la solution.

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & -\alpha/4 & 0 \\ -\alpha/4 & 4F_{skewness} & -\alpha/4 \\ 0 & -\alpha/4 & 0 \end{pmatrix}$$

a)    b)

Figure 4.27 : Matrice de convolution du filtre laplacien standard (a) et du filtre proposé (b)



Figure 4.28 : zones (zoomée) avec des objets d'intérêt et des faibles différences entre les couleurs de pixels (entouré de cercles rouges). Elles ne sont pas visibles à l'œil sans agrandissement, mais peuvent être détectés par des traitements numérique, tels que la détection de contours avec l'opérateur Laplacien. Ce phénomène peut générer des faux contours sur la deuxième phase de détection d'objets, et ainsi résulter en fausses détections par la suite de traitements.

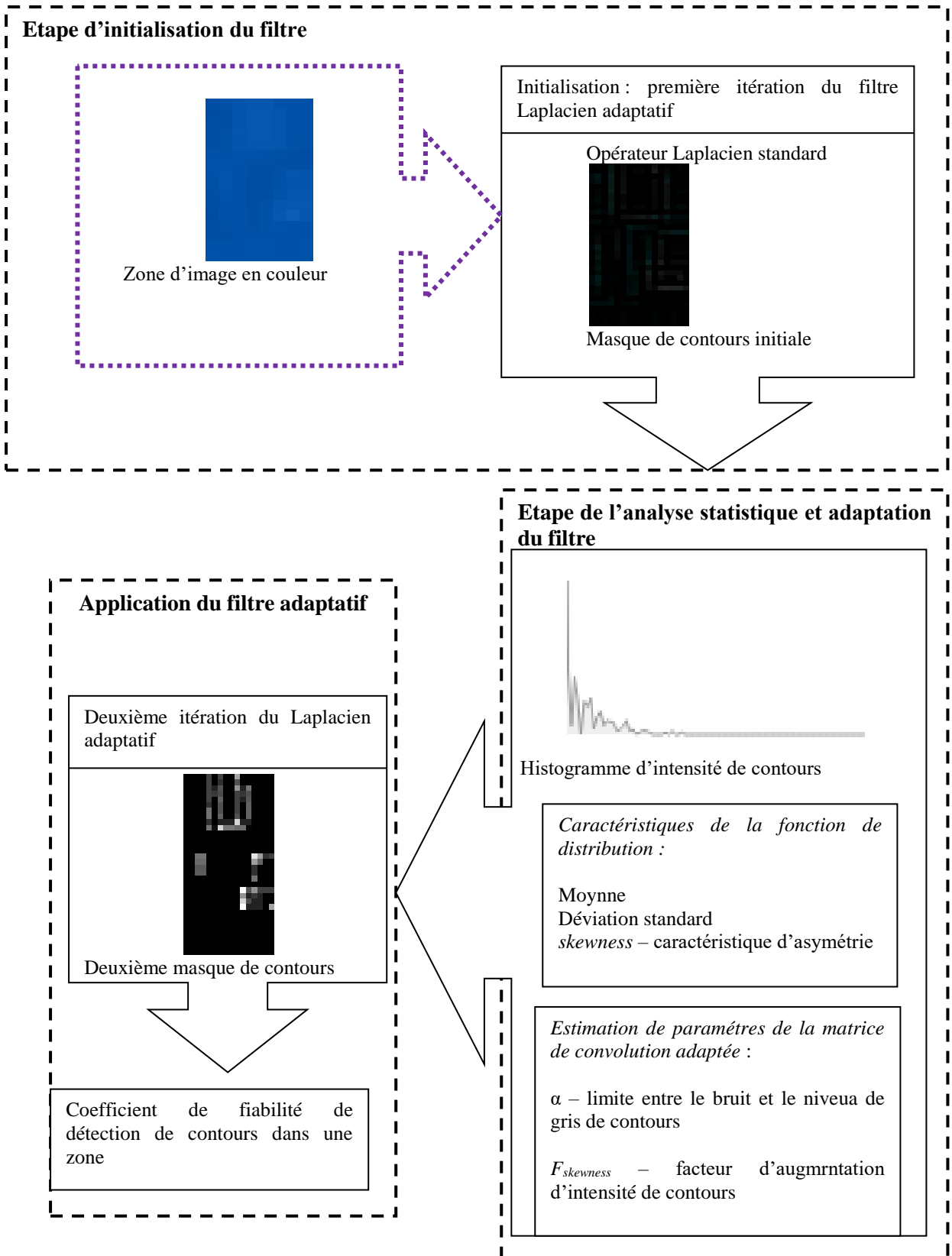


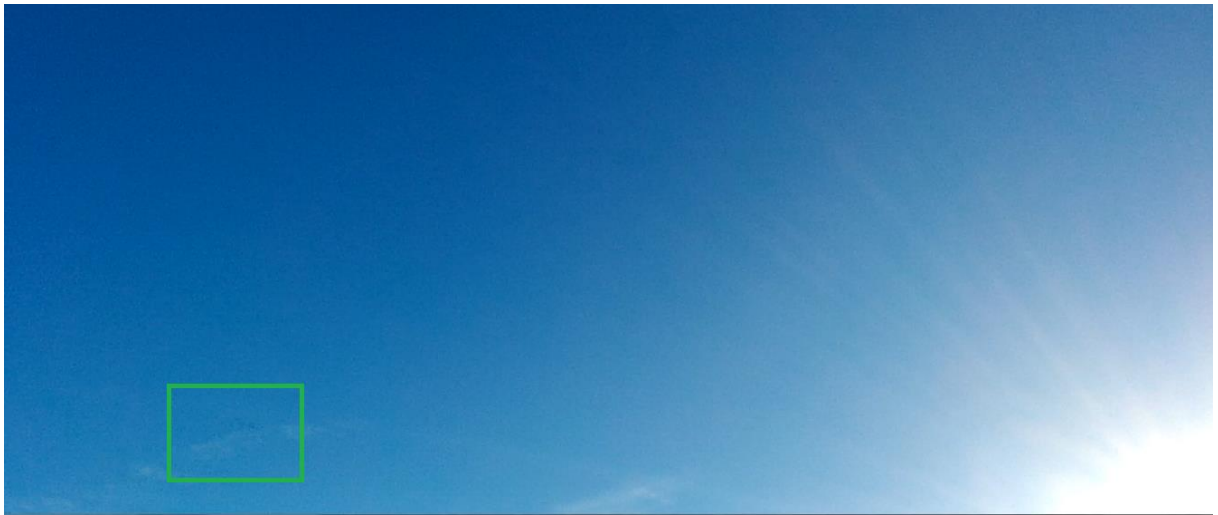
Figure 4.29 : Schéma général de la phase de détection de contours d'objets

#### 4.4.4.3 *Filtre adaptatif proposé : application du filtre Laplacien standard et analyse statistique*

L'exécution du filtre adaptatif se décompose en trois étapes logiques, schématiquement présentées dans la Figure 4.29:

1. *Initialisation du filtre adaptatif: Première itération du filtre adaptatif. Application du filtre Laplacien standard et création du premier masque de contours*
2. *Etape de l'analyse statistique et adaptation du filtre*
  - 2.1 *Création de l'histogramme d'intensité de contours dans le premier masque de contours*
  - 2.2 *Analyse statistique de l'histogramme d'intensité de contours dans le premier masque :*
    - *estimation de la moyenne et de la déviation standard d'intensité de contours*
    - *estimation de l'asymétrie de l'histogramme*
  - 2.3 *Estimation de paramètres pour la matrice de convolution adaptée à une zone*
3. *Application du filtre adaptatif : Application de la matrice de convolution adaptée sur la zone d'intérêt analysée*

Les sections suivantes décrivent en détails les étapes du filtre adaptatif proposé et expliquent leurs bases théoriques.



(a) Image initiale en couleur avec des très petits objets avec un faible contraste sur le fond

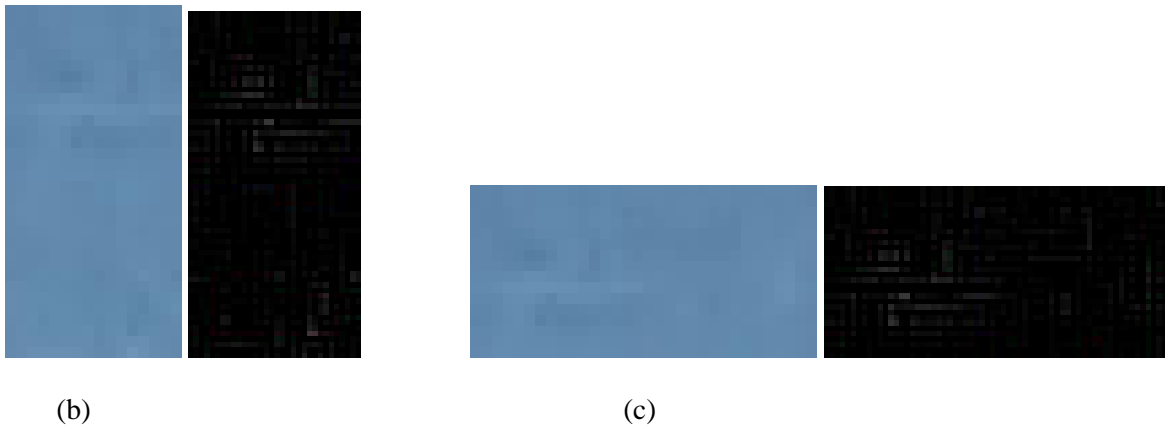


Figure 4.30 : Exemple de zones avec des objets à un très faible contraste sur le fond (a) et de leurs contours, détectés par le filtre Laplacien standard (b, c).

#### 4.4.4.4 Première itération et étape d'adaptation du filtre Laplacien adaptatif

La détection de contours de petits objets par le filtre Laplacien adaptatif conçu débute par l'application de sa version standard et la matrice de convolution suivante :

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

Il sert à créer un masque initial de contours qui permet de repérer toutes les différences entre les pixels voisins d'une zone (e.g. sur la figure 4.31). Sur cette étape toutes les variations locales entre les couleurs des pixels d'une zone (y compris les différences des teintes et des intensités lumineuses) sont reflétées en tant que contours.

Avant de procéder à l'étape d'analyse statistique du masque de contours initial, on introduit hypothèses supplémentaires sur son apparence. Ces hypothèses sont dérivées des phénomènes, observés dans des masques de contours de zones d'intérêt avec des petits objets :

1. Hypothèse sur le nombre de pixels, représentant les contours d'objets d'intérêt : Les zones d'intérêt avec de petits objets ont un fond homogène, d'où les masques de contours initiaux, créés par le filtre Laplacien standard, ont la majorité de pixels noirs. Ainsi, le nombre de

pixels, qui représente le fond d'une zone est considérablement plus important, que le nombre de pixels, représentant les contours, par rapport au nombre total de pixels dans une zone

2. Hypothèse sur le contraste des contours d'objets d'intérêt sur le fond : Des faibles niveaux de gris peuvent également définir les contours d'objets d'intérêt pour la raison de leur probablement faible contraste par rapport au fond (e.g. en raison d'une grande distance de la caméra).

En acceptant ces hypothèses pour l'ensemble des zones d'intérêt, on peut déduire sans perdre en généralité, une forme attendue de la fonction de distribution d'intensités de pixels dans le masque de contours initial.

L'objectif de cette étape est d'estimer les paramètres qui permettront d'adapter la matrice de convolution au contraste de petits objets dans d'une zone d'intérêt. En d'autres termes, sur cette étape on souhaite calculer les facteurs sous forme des valeurs numériques qui changeront le comportement du filtre Laplacien standard en fonction du contraste d'objets dans une zone d'intérêt. Pour répondre à ce besoin une analyse statistique est appliqué sur le masque de contours, créé lors de la première itération les notions et les méthodes statistiques sont appliquées:

- La fonction de distribution normale, associée à l'histogramme d'intensités des contours dans le masque standard
- La moyenne d'intensité de contours dans le masque initial et la déviati on standard
- L'asymétrie de la fonction de distribution d'intensités des contours.

D'autre part, l'histogramme du masque de contours initial met en évidence la distribution bimodale d'intensités de pixels dans l'ensemble de niveaux de gris présents dans le masque et ses deux maximas locaux (Figure 4.31) :

L'ensemble de pixels le plus nombreux : représentant les zones du fond d'image (où il n'a pas de variation, i.e. où les pixels sur le masque de contours ont le niveau de gris égal à 0) et les contours avec les plus faibles contrastes sur le fond (représentés par les niveaux de gris proches à 0)

L'ensemble beaucoup moins nombreux, mais suffisant pour générer une distribution bimodale des intensités de pixels. Ce deuxième maxima local représente les contours d'objets d'intérêt avec les contrastes sur le fond les plus forts.

Ces deux maximas repartissent implicitement les intensités de contours dans le masque initial en trois classes (Figure 4.31) caractérisées par :

- L'intensité du fond de l'image : désigné par les niveaux de gris les plus bas dans le masque de contours - le maxima local à gauche dans l'histogramme
- Les intensités des contours d'objets d'intérêt : désignés par les niveaux de gris les plus élevés dans le masque de contours et par le maxima local à droite dans l'histogramme
- Les intensités occasionnelles de faux contours (limites entre des couleurs, bruit numérique) : des valeurs des niveaux de gris distribuées entre les deux maximas

La caractéristique d'asymétrie, « *skewness* », donne une information qualitative supplémentaire sur la forme d'une fonction de distribution normale. Elle caractérise notamment le décalage de la fonction de distribution de la moyenne. Dans le cas d'une distribution bimodale la valeur, associée à l'asymétrie permet d'approximativement localiser le maxima local, représentant les pixels du fond du au masque initiale et leurs intensités.

La caractéristique d'asymétrie est rarement utilisée dans des travaux existants [36]. Dans le cadre de la méthode définie dans le cadre de cette thèse, elle représente une information quantitative sur la



distribution, ou les paramètres pour le filtre adaptif sont déduits et adaptés localement à une zone d'intérêt. Ces paramètres sont les suivants :

$\alpha$  – limite entre les niveaux de gris de pixels du fond et ceux de potentiels contours d'objets d'intérêt.

$F_{skewness}$  – facteur d'augmentation.

Le principe d'adaptation du filtre aux moyen de l'analyse statistique est décrit ci-dessous.

#### 4.4.4.5 Estimation du seuil de l'intensité de contours d'objets d'intérêt en fonction de la caractéristique d'asymétrie

La caractéristique d'asymétrie indique le décalage de la fonction de distribution de la moyenne et indique la plus forte concentration d'échantillons dans la distribution. En d'autres termes, la caractéristique d'asymétrie est appliquée comme un indicateur de la « densité » de l'histogramme, i.e. l'intensité (le niveau de gris) autour duquel les pixels sont le plus cumulé. Elle permet ainsi de localiser le premier maxima local sur l'histogramme, associé aux pixels du fond et à leurs intensités. De cette façon la caractéristique d'asymétrie permet de déterminer approximativement l'intensité attendue pour les pixels, représentant le fond sur d'une zone d'intérêt avec les contrastes les plus faibles.

Sachant l'intensité de contours maximale  $\max_{gray}(H)$ , présente parmi les pixels dans le masque initial (qui d'après les hypothèses admises est associée aux contours d'objets d'intérêt), le seuil  $\alpha$ , qui sépare les intensités de pixels du fond et des contours potentiels, est estimé par la formule suivante :

$$N_{sk} = \frac{\max_{gray}(H)}{skewness}$$

La valeur du seuil défini par la formule ci-dessus tend vers les niveaux de gris les plus bas dans l'histogramme – associés au fond d'une zone, où les échantillons sont les plus nombreux. Mais en pratique, même dans les zones visuellement homogènes le filtre Laplacien standard détecte aussi des faibles variations entre les teintes, non-visibles à l'œil. Afin de ne pas considérer ces faibles variations comme des contours potentielles d'objets d'intérêt avec une faible intensité, sa valeur est ajustée par la formule suivante :

$$\alpha_{sk} = \frac{\overline{Y(noise)} + \min(Y(p_j))}{2}$$

où  $|Y(p_j) - Y(p_i)| > \overline{Y(dist)}$ .

L'idée de cette formule est de réunir les caractéristiques numériques des classes de pixels - du fond d'image et de faux contours - en une seule, qui permettrait approximativement séparer tout l'ensemble en potentiels contours et en ceux qui ont des intensités trop basses. La figure 4.31 illustre d'une manière schématique, où se trouve la valeur recherchée du seuil ajusté  $\alpha_{sk}$ . Chaque terme de cette formule est expliqué plus loin dans cette section.

La caractéristique d'asymétrie permet de localiser le niveau de gris moyen pour les contours les moins intenses (mais les plus nombreux) sur l'histogramme d'intensités de contours. Ces contours sont attribués dans la première classe de trois, présentés sur la figure ci-dessus et par leur apparence ils sont les plus proches au fond d'image (i.e. les groupes de pixels le plus proches au cercle rouge). Afin d'éviter la sélection de ces faux contours dans le masque de contours d'objets, généré par la deuxième itération du filtre, la valeur de l'intensité minimale pour les contours d'objets d'intérêt - du seuil  $\alpha$  – est estimée en fonction des la dispersion des niveaux de gris de trois classes dans l'histogramme. Pour cette estimation deux paramètres supplémentaires sont introduits :

- $\overline{Y(dist)}$  : distance moyenne entre les intensités en niveaux de gris situés entre deux maxima locaux de l'histogramme du masque des contours initiaux.
- $\overline{Y(noise)}$  : niveau de gris approximatif attribué aux pixels du fond de l'image de contours

Concernant le terme  $\overline{Y(dist)}$  on admet, que les niveaux de gris entre les deux maxima locaux peuvent représenter des pixels dans les masques de contours analysés, dont l'intensité n'est pas suffisante pour les contours d'objets d'intérêt. Avec cette hypothèse on suppose, que la première valeur de niveaux de gris pour les contours potentiels (deuxième classe sur l'histogramme) peut se situer à une distance supérieure à la différence moyenne entre tous les niveaux de gris de contours potentiels. Ainsi on introduit la formule

$$|Y(p_j) - Y(p_i)| > \overline{Y(dist)}$$

où  $Y(p_i)$  et  $Y(p_j)$  sont les valeurs des niveaux de gris présentés respectivement par les colonnes  $p_i$  et  $p_j$  dans l'histogramme du masque de contours initial.  $Y(p_j)$  représente alors la valeur du premier niveau de gris dans l'histogramme, qui pourrait correspondre à des contours d'objets d'intérêt avec le plus faible contraste.  $\overline{Y(dist)}$  est la distance moyenne entre les niveaux des gris, répartis entre les deux maxima locaux est estimée comme :

$$\overline{Y(dist)} = \frac{1}{255 - N_{SK} - 1} \sum_{i=N_{sk}}^{255} |Y(p_{i+1}) - Y(p_i)|$$

La valeur moyenne pour le bruit  $\overline{Y(noise)}$  et les contours les moins contrastés est estimée comme :

$$\overline{Y(noise)} = \frac{1}{N_{SK}} \sum_{i=N_{sk}}^{255} Y(p_i)$$

La valeur de la caractéristique d'asymétrie *skewness* et la valeur estimée du seuil des intensités de potentiels contours d'objets d'intérêt

$$\alpha_{sk} = \frac{\overline{Y(noise)} + \min(Y(p_j))}{2}$$

construisent la matrice de convolution du filtre Laplacien adaptatif pour sa deuxième itération. Elle représente une combinaison linéaire de deux matrices, qui seront appliquées sur la même zone d'image à la place de la matrice de convolution standard lors de sa deuxième itération du filtre laplacien adapté :

$$\begin{aligned} & \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 * skewness & -1 \\ 0 & -1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -\alpha/4 & 0 \\ -\alpha/4 & 0 & -\alpha/4 \\ 0 & -\alpha/4 & 0 \end{pmatrix} \\ & = \begin{pmatrix} x_{00} & x_{01} & x_{02} \\ x_{10} & x_{11} & x_{12} \\ x_{20} & x_{21} & x_{22} \end{pmatrix} * \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 * skewness & -1 \\ 0 & -1 & 0 \end{pmatrix} - \alpha \end{aligned}$$

#### 4.4.4.6 Deuxième itération du filtre Laplacien adaptatif : application de la combinaison linéaire des matrices paramétrées

Les deux matrices de convolution du filtre Laplacien adaptatif exécutent en une action ses deux objectifs :

- Renforcer les contours d'objets d'intérêt
- Atténuer ou supprimer le bruit et les faux contours

La convolution avec la première matrice de la combinaison linéaire répond au premier objectif. Elle permet de renforcer le niveau de gris d'un pixel proportionnellement à la degré d'asymétrie de l'histogramme en plus du facteur 4 du noyau standard. Lors de la convolution les valeurs des voisins du pixel central, auquel la matrice est appliquée, y sont ajoutées avec les facteurs du noyau standard (0 ou -1). Ce processus n'affecte pas les pixels qui représentent le fond d'image. Si leurs valeurs valent 0 sur le masque standard de contours, elles restent à 0 après la multiplication par  $4 \cdot skewness$  et la soustraction de valeurs de pixels voisins. Si leurs valeurs des niveaux de gris sont légèrement supérieures de 0 (i.e. ces pixels représentent des faux contours ou du bruit), elles peuvent augmenter après la convolution avec la première matrice. Néanmoins, cette action n'augmente pas ces valeurs considérablement et n'affecte pas la suite de la deuxième itération du filtre adaptatif.

La deuxième matrice applique la soustraction de la valeur  $\alpha$  sur le pixel central traité afin de répondre au deuxième objectif de l'application du filtre adaptatif. Elle représente le niveau de gris minimal pour les contours d'objets d'intérêt et maximal pour les pixels du fond et de faux contours de transition de couleurs sur le fond. Cette soustraction atténue alors les contours les moins intenses, qui ont les niveaux de gris les plus bas sur le masque de contour. Les contours, dont les intensités sont supérieures ou proches à  $\alpha$ , restent quant à eux, présents dans le masque de contours. La précédente multiplication par le facteur d'asymétrie *skewness* permet de conserver les contours potentiels d'objets d'intérêt dans le masque de contours après la deuxième itération du filtre. Mais d'autre côté elle permet aussi de relever les niveaux de gris de vrais contours d'objets d'intérêt, dont les niveaux de gris de contours sont situés autour du deuxième maxima local dans la moitié de droite de l'histogramme d'intensités. La soustraction de la valeur  $\alpha$  n'a pas d'effet sur celles des niveaux de gris de ces contours, car après la multiplication par la valeur de *skewness* elles sont largement supérieures au seuil estimé. L'application de la combinaison linéaire de matrice se justifie ci-dessous sous forme des inéquations.

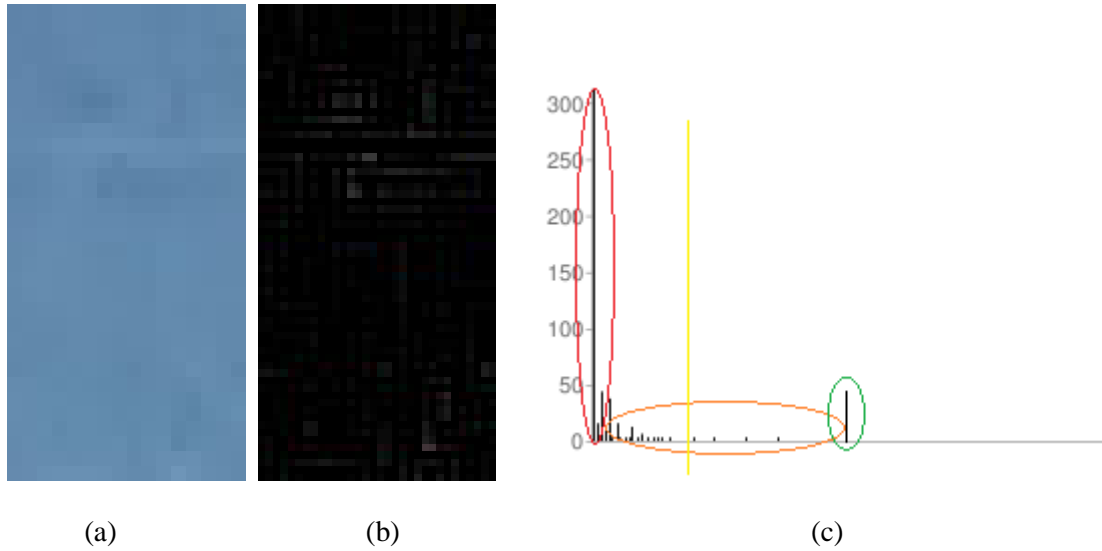


Figure 4.31 : Une zone avec des petits objets avec un faible contraste sur le fond, son masque de contours initial et l’histogramme d’intensités de ses pixels. L’histogramme représente les trois classes de pixels dans le masque initial de contours avec les intensités différentes. La classe de pixels, qui représentent le fond est entourée en rouge, la classe de pixels, associée à des contours d’objets – en vert, et la classe de pixels avec des intensités occasionnelles – en orange. La valeur recherchée du seuil  $\alpha_{sk}$  est indiquée par la ligne jaune.

#### 4.4.4.7 Justification analytique d’application du filtre adaptatif

L’objectif de cette justification est de prouver, qu’après la seconde itération l’intensité des contours d’objets d’intérêt reste supérieure ou égale au seuil minimal  $\alpha$ , si elle l’était initialement (après l’application du filtre standard), et les contours, dont l’intensité est inférieure à  $\alpha$ , sont supprimés. On admet deux hypothèses pour la preuve de l’efficacité du filtre proposé :

**Hypothèse 1:** si dans le masque de contours initial (créé en première itération du filtre standard) l’intensité du pixel  $p$  et les valeurs d’au moins deux de ces voisins sont supérieures au seuil  $\alpha$ , sa valeur après l’application de la deuxième itération du filtre adaptatif sera supérieure ou égale à  $\alpha$ . Dans la combinaison linéaire des matrices (section 4.3.4.5) ce pixel  $p$  correspond à l’élément central de matrice  $x_{11}$

**Hypothèse 2:** si dans le masque de contours initial l’intensité du pixel  $p$  et celles de ses voisins sont inférieures au seuil  $\alpha$ , sa valeur après l’application de la deuxième itération du filtre adaptatif sera inférieure du seuil  $\alpha$ . Ainsi le pixel  $p$  ne fera partie de l’ensemble des contours des objets d’intérêt.

En appliquant la formule de convolution, la valeur finale de l’élément central  $y$  est estimée comme :

$$y = 4 * skewness * x_{11} - (x_{01} + x_{10} + x_{12} + x_{21} + \alpha)$$

En appliquant l’hypothèse 1 :

$$si \begin{cases} x_{11} \geq \alpha \\ x_{01} \geq \alpha \\ x_{21} \geq \alpha \\ x_{10} < \alpha \\ x_{12} < \alpha \end{cases} \rightarrow \begin{cases} 4 * skewness * x_{11} \geq 4 * skewness * \alpha \\ x_{01} + x_{21} + x_{10} + x_{12} + \alpha \geq 3\alpha \\ 4 * skewness * x_{11} - 3\alpha > \alpha \\ x_{11} > x_{10} \\ x_{11} > x_{12} \end{cases} \rightarrow y \geq \alpha$$

L’hypothèse 2 permet de déduire les inéquations suivantes :

$$\text{si } \begin{cases} x_{11} < \alpha \\ x_{01} \geq \alpha \\ x_{21} \geq \alpha \\ x_{10} < \alpha \\ x_{12} < \alpha \end{cases} \rightarrow \begin{cases} 4 * \text{skewness} * x_{11} < 4 * \text{skewness} * \alpha \\ x_{01} + x_{21} + x_{10} + x_{12} + \alpha \geq 3\alpha \\ 4 * \text{skewness} * x_{11} - 3\alpha < \alpha \\ x_{11} \leq x_{10} \\ x_{11} \leq x_{12} \end{cases} \rightarrow y < \alpha$$

La figure 4.32 résume ces traitements. Après la deuxième itération du filtre adaptatif, le seuillage avec la valeur estimée  $\alpha$  est appliqué sur le nouveau masque de contours. Puis il sera converti en image binaire avec des régions continues, associées aux objets d'intérêt (voir la section 4.4.5 pour les détails). Le seuillage et l'estimation des coefficients de fiabilité lors de ces deux étapes de traitement sont effectués par la fonction de la logique floue – sigmoïde. Leurs principes sont détaillés dans la section suivante.

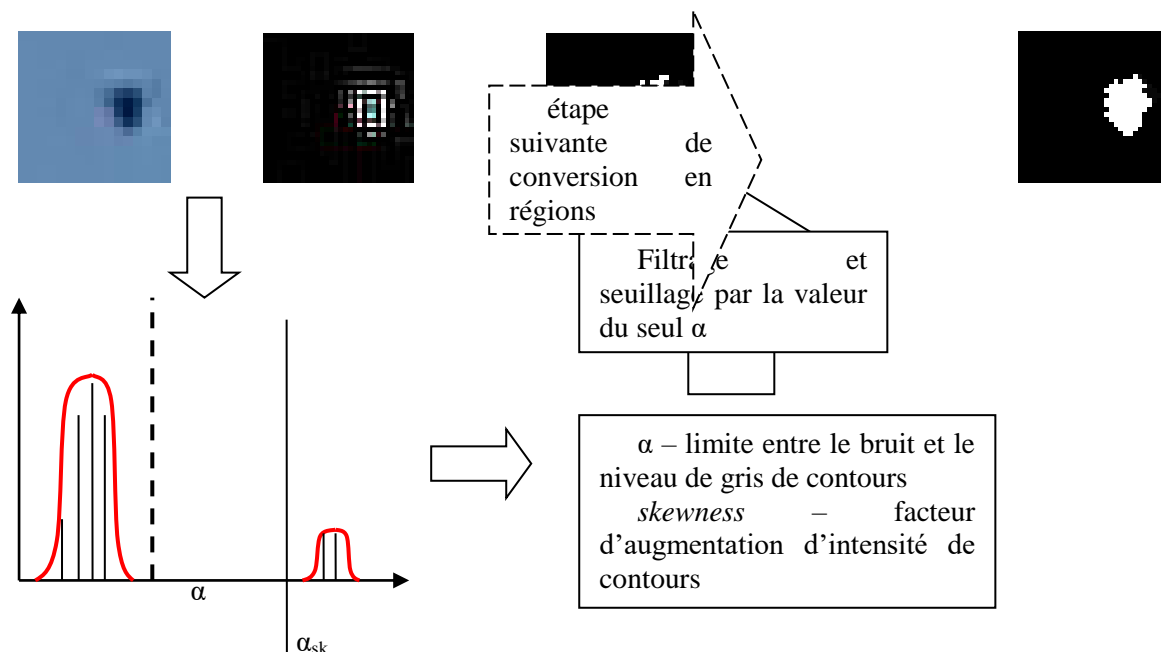


Figure 4.32 : Présentation schématique d'étapes de détection de contours par le filtre adaptatif pour le cas d'un petit objet. Le coefficient de fiabilité obtenu pour les contours de l'objet dans l'exemple 0.79, le coefficient de conversion de ces contours à l'étape suivante obtiendra la valeur de 0.97

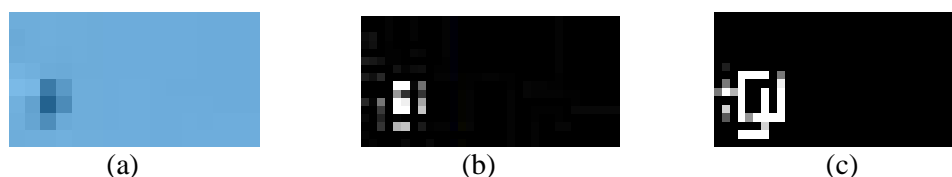


Figure 4.33 : (a) image initiale avec un très petit objet ; (b) contours de l'objet sur l'image (a) détectés par le filtre Laplacien standard, (c) contours du même objet détectés par le filtre Laplacien adaptatif

#### 4.4.4.8 Seuillage par la fonction sigmoïde du deuxième masque de contours

Le filtre adaptatif, proposé dans la solution, permet de calibrer sa sensibilité en fonction du contraste des contours d'objets dans une zone. Mais plus ce contraste est faible, plus il se rapproche des variations peu significatives des composantes de la couleur du fond. Dans ce cas il y a une probabilité que des faux contours, représentant ces variations, seront aussi présents dans le deuxième masque de contours. Un seuillage avec la valeur  $\alpha$  estimée précédemment est alors appliqué pour affiner la sélection de vrais contours d'objets.

Du point de vue analytique la nécessité de ce seuillage s'explique par le fait, que la valeur du seuil  $\alpha$  est dérivée depuis les valeurs observées dans une distribution et estimée par des méthodes statistiques. Dans le contexte de détection de petits objets avec un faible contraste sur le fond d'image, le niveau de gris estimé en tant que seuil pour les vrais contours d'objets d'intérêt ne peut être qu'approximatif. Des moindres valeurs supérieures à ce seuil peuvent correspondre aux faux contours, ou au contraire, des contours avec des niveaux de gris légèrement inférieurs peuvent toujours représenter des contours d'objets d'intérêt avec un contraste plus faible. Ce phénomène particulier pour le cas de détection de contours de petits objets est pris en considération dans l'étape de seuillage, détaillée dans cette section.

En appliquant le seuillage par la formule suivante, où  $p$  est un pixel sur le masque de contours après la deuxième itération

$$p \in \begin{cases} \text{ensemble de contours, si } Y(p) \geq \alpha \\ \text{fond, si } Y(p) < \alpha \end{cases}$$

on risque de supprimer des contours d'objets d'intérêt légèrement moins intenses. Pour résoudre cette difficulté et garder ces contours légèrement d'objets d'intérêt potentiels, le seuillage du deuxième masque des contours est effectué par la fonction d'appartenance de la logique floue – sigmoïde. Dans le contexte du concept fondamental de l'algorithme : de fiabilité de détections, cette fonction implémente une technique de seuillage, basé sur la fiabilité de chaque pixel de représenter les contours d'un objet d'intérêt. Son deuxième avantage est la possibilité d'estimer les coefficients de fiabilité d'exécution de l'étape de détection de contours tout au long de son exécution.

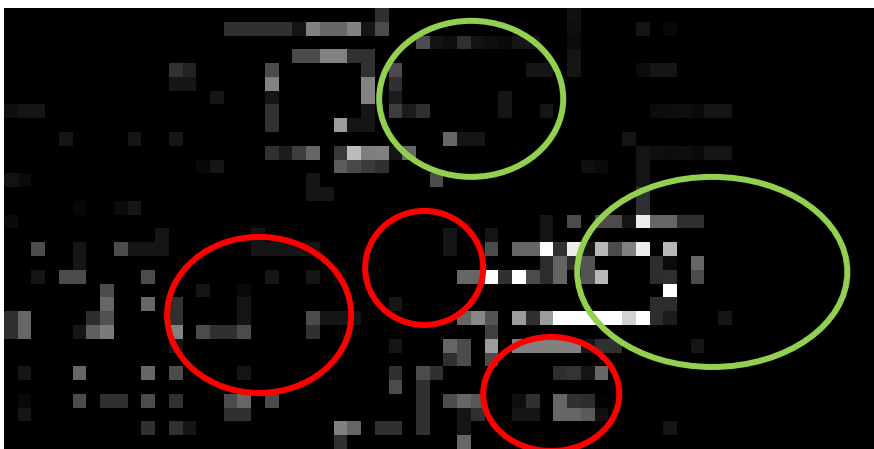
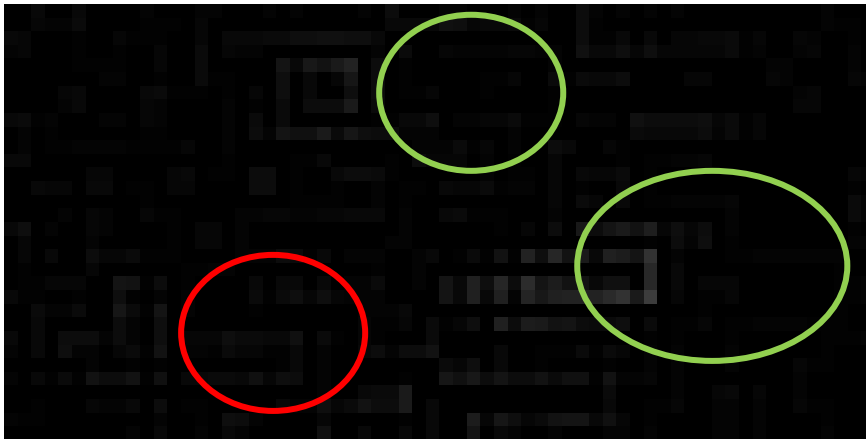
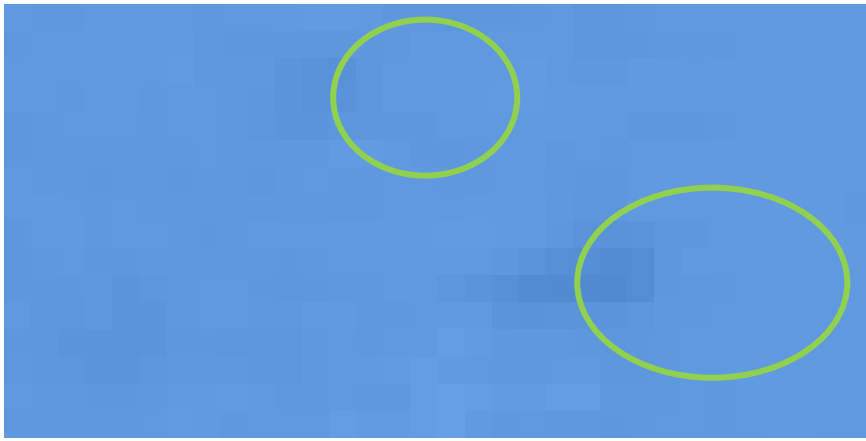


Figure 4.34 : Exemple de détection d'objets d'intérêt avec un faible contraste sur le fond. Les cercles verts entourent les objets d'intérêt (les oiseaux dans un grand vol très éloigné de la caméra) et leurs contours associés, détectés par le filtre standard (image de milieu) et le filtre adaptatif (image du bas). Les cercles rouges entourent les fausses détections de contours (e.g. des faibles variations des teintes de la couleur du fond). Les coefficients de fiabilité sont nécessaires pour diminuer l'effet de fausses détections.

#### 4.4.4.9 Formalisation de la technique de seuillage proposée

La technique proposée pour le seuillage du masque de contours après la deuxième itération du filtre se base sur la notion de fiabilité qu'un pixel fait partie de l'ensemble de contours d'objets d'intérêt. Tous les pixels  $p$ , dont le niveau de gris dans le deuxième masque de contours est différent du zéro (i.e. les pixels différents du fond d'une zone d'image) obtiennent leurs valeurs de fiabilité associées  $W(p)$ . Celles dernières sont calculées par la fonction sigmoïde dont la formule générale est :



$$f(x) = (1 + e^{-\alpha(x-c)})^{-1}$$

Les paramètres  $c$  et  $a$  dépendent du seuil  $\alpha$ , estimé pour une zone précise, et de la différence moyenne entre les niveaux de gris attribués aux contours. Pour la classification des pixels en tant que des contours et d'estimation de fiabilité de détection de contours, obtient la forme suivante :

$$W(p) = \begin{cases} 1, & \text{if } Y(p) \geq \alpha \\ (1 + e^{-\alpha(Y(p)-c)})^{-1}, & \text{if } Y(p) < \alpha \end{cases}$$

où  $c = \alpha$ , défini par le seuil estimé  $\alpha$ , représente la coordonnée X du point d'inflexion de la fonction (i.e. la valeur à partir laquelle les niveaux de gris de pixels seront considérés en tant que contours). Le paramètre  $a$  caractérise la forme de la courbe de fonction, et ainsi comment les valeurs de fiabilité qu'elle attribue aux pixels diminuent en s'éloignant du seuil (Figure 4.35) :

$$\alpha = \frac{\left[ \sum_{i=\alpha-\overline{Y(dist)}}^{\alpha+\overline{Y(dist)}} Y(p_i) \right] / N}{\max(Y)}$$

$N$  est le nombre de pixels avec des niveaux de gris proches à la valeur du seuil  $\alpha$  : dans l'intervalle  $[\alpha - \overline{Y(dist)}, \alpha + \overline{Y(dist)}]$  et  $\max(Y)$  est le niveau de gris maximal dans le masque de contours après la deuxième itération.

Le concept de ce seuillage fonctionnel est illustré sur la figure 4.35. En principe, un seuillage standard est appliqué par la fonction de Dirac, qui attribue les mêmes valeurs aux éléments, supérieurs à un seuil, et la valeur 0 aux autres. Dans le contexte de sélection de contours dans le deuxième masque un tel seuillage fera disparaître une partie de contours, qui peuvent aussi faire partie de l'ensemble de contours d'objets d'intérêt. A la différence du seuillage standard, le seuillage fonctionnel proposé permet d'analyser les niveaux de gris proches à celui estimé des contours d'objets d'intérêt aux moyens de la fiabilité qu'ils représentent des contours d'objets, implémentée par la fonction sigmoïde. Pour les contours des objets d'intérêt avec les niveaux de gris supérieures du seuil  $\alpha$ , la fonction donne la valeur 1, qui signifie qu'après le seuillage ils obtiennent la valeur 255 de la couleur blanche. Sinon, le niveau de gris est atténué proportionnellement à la fiabilité, estime par la fonction. Les contours dans le masque ainsi obtenu sont transformés dans des régions continues, représentant des objets, à l'étape suivant de la solution.

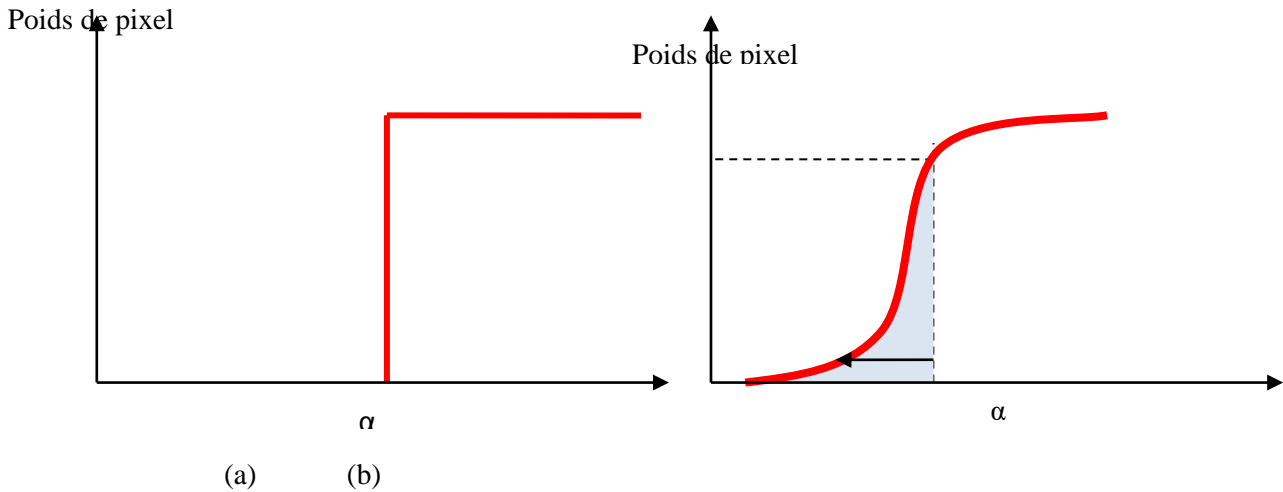


Figure 4.35: Illustration de fonctions, appliqués pour les seuillages : la fonction de Dirac pour le seuillage classique (a) et la fonction sigmoïde pour le seuillage fonctionnel. La zone bleue sur l'image (b) est éliminée lors de l'application de la fonction de Dirac pour le seuillage standard. C'est-à-dire que les pixels avec ces niveaux de gris sur le deuxième masque de contours sont classés comme des faux contours et obtiennent alors les coefficients de fiabilité de 0. La fonction sigmoïde, quant à elle, inclut ces pixels avec des coefficients de fiabilité dans l'intervalle (0, 1] et les fait apparaître après le seuillage du masque. Le poids de pixel est l'intervalle [0, 1].

#### 4.4.4.10 Coefficients de fiabilité de détection de contours

Le coefficient de fiabilité de détections de contours est estimé en même temps que le seuillage au moyen de la fonction sigmoïde. La valeur estimée par cette fonction d'appartenance et comprise dans l'intervalle [0, 1] définit la fiabilité que ce pixel fait partie de l'ensemble de contours d'objets d'intérêt. Le coefficient de fiabilité est calculé pour une zone d'image avec des potentiels objets d'intérêt par la formule suivante :

$$DR(B) = \frac{1}{N_c} \sum_c^{N_c} W(p_c)$$

où le niveau de gris du pixel  $p_c$  sur le masque de contours (i.e. son intensité  $Y(p_c)$ ) satisfait la condition  $Y(p_c) \in (0, \alpha]$ , et son poids est estimé par la même fonction:

$$W(p_c) = \begin{cases} 1, & \text{si } Y(p_c) \geq \alpha \\ (1 + e^{-\alpha(Y(p)-c)})^{-1}, & \text{si } Y(p_c) < \alpha \end{cases}, c = \alpha \text{ et } \alpha = \frac{[\sum_{i=\alpha-\overline{Y}(\text{dist})}^{\alpha+\overline{Y}(\text{dist})} Y(p_i)] / N}{\max(Y)}$$

Le coefficient de fiabilité de détection de contours est estimé alors pour tous les pixels différents du fond et pour la zone entière, car à cette étape de traitement le nombre d'objets dans une zone est inconnu. De plus, la forme d'objets d'intérêt n'est pas définie dans le contexte de détection de petits objets, qui complique la séparation d'objets dans une zone à cette étape de traitement.

#### 4.3.4.11 Analyse de résultats d'application du coefficient de détection des contours

Dans le contexte du concept de la solution proposée la somme des coefficients doit graduellement converger vers le nombre final d'objets d'intérêt détectés dans une vidéo. En d'autres termes, en commençant par l'analyse de sélection des entités globales – des zones d'images, les étapes de la

solution appliquent une analyse visée sur les entités locales d'objets d'intérêt. La sélection de ces derniers est confirmée ou réfutée par l'estimation de coefficients de fiabilité aux étapes suivantes et au moyen de leurs compositions.

Les résultats expérimentaux du tableau 2 témoignent que le principe de composition de coefficients de fiabilité permet en effet d'approcher le nombre d'objets détectés à la vérité terrain. Contrairement à l'application d'un des deux coefficients séparément, qui à cette étape représente les nombres finaux d'objets détectés, leur composition donne de meilleurs résultats. Appliqué tout seul sur le même ensemble de vidéos, le coefficient de fiabilité de détection de contours permet de supprimer des fausses détections au moyen de la suppression des contours les moins fiables par la deuxième itération du filtre adaptatif. Sinon, sans estimer l'intensité de niveaux de gris de ces contours dans l'image générée, ils sont comptés comme des objets d'intérêt, en impliquant un nombre important de fausses détections.

Les résultats d'application de coefficients de fiabilité séparément dépendent des vidéos traitées et ne permet pas de faire une conclusion sur les avantages de l'un ou de l'autre. Par contre, quand les coefficients sont combinés ensemble, le nombre total de détection converge plus vers la vérité terrain.

**Tableau 2 : Résultats d'application des coefficients de fiabilité de sélection de zone et de détection de contours**

Corpus	Détections avec un de coefficient coefficients						
	sans R(O)	HSV <sub>w</sub>		DR		HR + DR	
		TW	FW	TW	FW	TW	FW
1	559	65.8	261.2	59.2	130.7	57.4	54.8
2	816	32	617.6	16	326.4	20.8	173.6

#### 4.4.5 Conversion de contours de petits objets d'intérêt en régions continues

Les contours d'objets, obtenus à l'étape précédente ne donnent qu'une apparence approximative d'objets d'intérêt. Les polygones, ou des courbes qui représentent les contours d'objets sont des formes très génériques pour pouvoir comparer la similarité d'objets détectés au modèle efficacement. L'objectif de cette troisième étape est alors de générer des régions, dont la forme peut être utilisée pour l'analyse de conformité d'objets détectés au modèle d'objet d'intérêt.

Naturellement, l'apparence d'objets dans une vidéo sera différente en fonction de leur distance de la caméra : plus elle est courte, plus détaillés sera la forme d'objets et objets mieux leur apparence pourra être analysée. En pratique, dans le contexte de suivi de vol d'oiseaux migratoires une distance relativement courte signifie qu'une forme dans une vidéo peut être approximée à une forme générique (e.g. une ellipse ou un cercle, des polygones non convexes ou une courbe). C'était le cas pour le corpus 1 de vidéos, fournies pour la première phase du développement et expérimentations. Leurs contours pouvaient être approximés par la plus petite ellipse englobante, dont les semi-axes et la superficie définissaient le modèle théorique d'objet d'intérêt. Pour le corpus 2 d'une complexité supérieure les objets peuvent visuellement ressembler à des primitives géométriques, comme dans le cas précédent, mais la comparaison numérique de leurs paramètres demandera la comparaison de très petites valeurs et d'une précision élevée, qui est techniquement limitée. Pour ce cas particulier, au lieu de la comparaison directe des objets des petites tailles la notion de coefficients de fiabilité est appliquée. Les contours des très petits objets d'intérêt sont caractérisés par leur continuité et par le coefficient de fiabilité associé. En même temps, les contours d'objets d'intérêt sont transformés en région fermées d'objets et dans un masque binaire d'une image traitée, où ils sont calculés sur la dernière étape de traitement. Dans ce chapitre détaille le processus de conversion de contours d'objets dans les région continues.

#### 4.3.5.1 Présentation de l'approche de conversion de contours de petits objets en régions continuées

Techniquement le processus de conversion de contours d'objets en régions continues implique la fermeture de contours et le remplissage avec la région fermée avec des pixels en niveaux de gris. Ce processus se base sur une hypothèse supplémentaire que dans une zone d'intérêt les contours de petits objets sont suffisamment proches les uns à des autres pour analyser leurs formes localement :

**Hypothèse sur la taille minimale d'une zone avec un objet d'intérêt :** sachant qu'un objet d'intérêt dans une zone doit occuper une région supérieure à deux pixels (voir la section 4.3.5 pour l'estimation de la taille minimale), un bloc carré 3x3 – appelé par la suite « un patch » - contiendra un objet entier ou au moins sa partie.

**Hypothèse sur la continuité de contours d'objet d'intérêt :** si un patch 3x3 contient une partie de contours d'objets d'intérêt, ils continuent dans un ou plusieurs patches voisins.

**Hypothèse sur la configuration de contours d'objet d'intérêt dans un patch 3x3 :** les contours dans les patches 3x3 sont localisés d'une manière, qui permet le plus probablement remplir l'intérieur de la région représentant un objet d'intérêt.

Un point important dans le traitement de contours par patches et de s'assurer que les pixels dans les patches avec des niveaux de gris supérieurs au seuil  $\alpha$  appartient à des contours continus dans des patches voisins et non à des pixels isolés dans le même patch. Le traitement conçu passe alors par des patches plus grands – de la taille 5x5 – composés de patches 3x3 superposés. L'analyse de ces patches 3x3 est conçu pour pondérer chacun de leurs 9 pixels au moyen d'analyse de leurs 8-voisinages. Chacun de 8 pixels-voisins est caractérisé et pondéré à son tour par ses 8-voisins.

Le principe de cette conversion est de suivre la progression et la continuité de contours des petits objets localement. Les patches 5x5 sont centrés dans les pixels classifiés en tant que contours fiables dans le masque créé par le filtre adaptatif. Les voisinages de ces contours fiables servent alors à confirmer qu'un pixel avec un niveau de gris supérieur au seuil  $\alpha$  est bien une partie de contours d'un objet d'intérêt et non un pixel isolé du bruit.

La pondération locale de contours dans des plus petits patches 3x3, associés aux 8 voisins d'un pixel analysé à l'intérieur du patch 5x5 (figure 4.36) permet de définir la direction de conversion et la valeur du niveau de gris de remplissage de l'intérieur de l'objet. Ce principe est implémenté par la séquence suivante, illustré dans le schéma 4.36 :

Pour chaque pixel  $p$ , classifié comme un contour potentiel d'objet (i.e. pour lequel le niveau de gris dans le masque de contours généré par la seconde itération répond à l'équation  $Y(p) \geq \alpha$ ) :

1. associer 4 patches  $P_1, P_2, P_3, P_4$  centrés dans les pixels voisins  $p_1, p_2, p_3, p_4$  respectivement est constitués de leurs 8 voisins
2. calculer les moyens de niveaux de gris  $M_{p_1}, M_{p_2}, M_{p_3}, M_{p_4}$  pour chaque patch  $P_1, P_2, P_3, P_4$
3. estimer les rapports des moyennes au seuil  $\alpha$  et en déduire la direction de remplissage de contours
4. Privilégier la direction de remplissage avec le poids maximal en choisissant le patch avec le rapport maximal:  $\frac{M_{p_i}}{\alpha}$ 
  - Si  $\frac{M_{p_i}}{\alpha} \geq \frac{M_p}{\alpha}$  remplir le patch  $i$  par la valeur  $\alpha$ , définie localement pour une zone traitée
  - Sinon, remplir le patch correspondant au pixel  $p$  (composé par ses 8 voisins) par des valeurs pondérées par la combinaison linéaire :

$$Y_p = Y(p) * C(p) + \alpha * (1 - C(p)),$$

où le poids  $C(p)$  du niveau de gris attribué avant au pixel  $p$  est estimé en fonction de son voisinage:

$$C(p) = \frac{1}{8} \sum_{i=1}^8 \frac{M_{p_i}}{\alpha} (1 - w) + \frac{M_p}{\alpha} w$$

où  $w \in [0, 1]$  est un paramètre, qui définit l'importance du poids du 8-voisinage du pixel  $p$ , ou ceux de ses voisins (dans l'implémentation d'algorithme  $w=0.5$ )

Cette séquence de traitement résulte ainsi en une valeur de niveaux de gris pondérée par ceux du voisinage du pixel traité, qui peut être comparée avec celle du niveau de gris minimal pour les contours d'objets d'intérêt. C'est-à-dire que la proportion de cette valeur et celle du seuil représente implicitement le coefficient de fiabilité, qu'un pixel appartient à l'ensemble de contours d'objets d'intérêt. Grâce à cette forme modifiée de représentation de coefficient de fiabilité les faux contours peuvent être supprimé par un seuillage avec la même valeur  $\alpha$  estimée pour les objets d'intérêt.

Finalement, les pixels dans le masque dont les niveaux de gris sont supérieurs au seuil obtiennent la valeur de la couleur blanche 255, et les autres – la valeur du fond 0. Le dernier traitement pour obtenir un masque binaire d'objets d'intérêt dans une zone est l'opérateur morphologique d'ouverture par l'élément structurant du cercle au rayon 1. Il sert à séparer des objets probablement fusionnés lors de la conversion et pour supprimer le bruit occasionnel d'une taille inférieure à la taille minimale acceptée pour un objet d'intérêt.

La conversion des contours en régions continues également obtient un coefficient de fiabilité. Il est calculé par la formule du coefficient de corrélation de Mattieu (Mattieu Correlation Coefficient). L'explication et le principe d'application de ce coefficient de fiabilité sont donnés dans la Section 4.4.5.2 suivante.

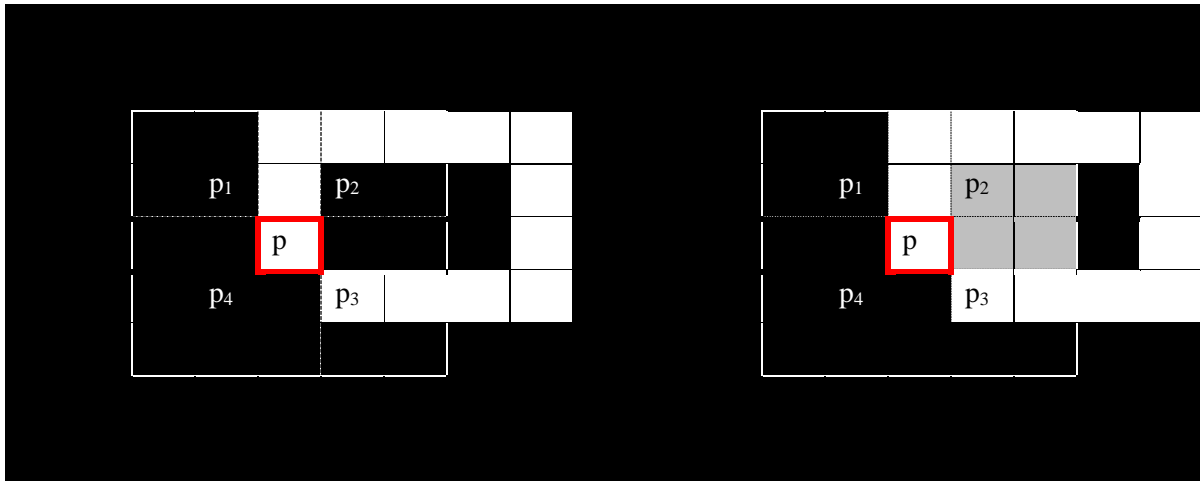


Figure 4.36 : exemple de conversion de contours dans une région continue. Le rapport entre la valeur du seuil  $\alpha$  et des niveaux de gris moyens pour les patches 3x3 centrés dans chaque pixel  $p_1$ ,  $p_2$ ,  $p_3$ ,  $p_4$  classe les patches 3x3 correspondants à 8-voisins du pixel central  $p$ . Ce rapport définit une direction de remplissage de contours dans le patch 5x5 qui contient le pixel  $p$  traité et ses 8 voisins, ainsi que 8 voisins de ces pixels. Pour le petit objet sur cet exemple le patch avec le pixel central  $p_2$  a la moyenne de voisinage la plus élevée et sera choisie pour le remplissage de contours (la zone grise sur l'image de droite).

#### 4.4.5.2 Coefficients de fiabilité de conversion des contours détectés en régions continues

À l'étape de conversion de contours un coefficient de fiabilité est associé à une zone traitée, et donc à chaque objet détecté qu'elle contient. Le principe d'estimation de ce coefficient de fiabilité est d'analyser chaque pixel qui constitue la zone avec des objets reconstruits depuis des contours détectés précédemment. L'idée consiste à répartir tous les pixels en classes de correctes et de fausses détections en fonction de leur nature. Dans un masque de contours, les classes suivantes de pixels peuvent être présentes :

- Contours d'objets d'intérêt
- Pixels du fond d'images
- Fausses détections (des faux contours ou du bruit)

La formule qui sert à estimer le coefficient de fiabilité analyse l'appartenance de pixels à toutes les trois classes et d'ajuster la valeur de coefficient de fiabilité à la classe la plus nombreuse. Les zones, où il y a beaucoup de fausses détections ou de pixels du fond, obtiendront des coefficients bas. Les zones avec des pixels, représentant des contours d'objets d'intérêt et ceux du fond obtiendront des coefficients de fiabilité plus élevés. La formule du coefficient de corrélation de Matthew (Matthew correlation coefficient, MCC [41]) permet de prendre en considération plusieurs classes d'objets et d'estimer la qualité de classification. Dans le contexte de l'étape de conversion de contours cette classification représente l'estimation de coefficients de fiabilité individuellement aux pixels, ce qui les attribue dans une des 3 classes (contours, pixels du fond ou de fausses détections).

**Le coefficient de corrélation de Matthew (Matthew correlation coefficient, MCC)** est un outil issu des techniques de machine learning. Il permet d'évaluer la précision de classification binaire pour des applications en biochimie. Le principe de ce coefficient est d'évaluer le rapport entre les nombres de classifications correctes et fausses.

Dans le cadre de l'étape de conversion de contours de petits objets en régions continues, le coefficient MCC est interprété comme des classifications de pixels dans l'ensemble de pixels, qui composent un objet (ses contours ou son intérieur), dans l'ensemble des faux contours et du bruit, ou dans l'ensemble des pixels qui composent le fond de l'image. En fonction de la valeur du niveau de

gris d'un pixel, estimée lors de la conversion de contours en région contenue, et celle de son voisinage le pixel sera classé en termes du MCC comme :

- « *True positif (TP)* » : les pixels dans une zone qui sont associés aux contours ou aux régions qu'ils englobent (i.e. les pixels de l'intérieur d'un objet)
- « *True négatif (TF)* » : les pixels qui désignent le fond d'image
- « *False positif (FP)* » : les pixels qui représentent le fond de l'image, mais qui ont obtenu des valeurs de niveau des gris supérieures à celui du fond de l'image
- « *False négatif (FN)* » : les pixels, qui représentent les contours ou l'intérieur d'un objet, mais qui ont obtenu des niveaux des gris inférieurs au seuil  $\alpha$ , estimé pour la zone, qui les contient

A la différence du coefficient de corrélation MCC classique, la classification faite sur l'étape de conversion de contours n'est pas binaire. Dans le principe d'estimation de fiabilité des détections elle est pondérée par le coefficient  $C(p)$ , associé aux pixels d'un masque de contours. C'est-à-dire, qu'au lieu de classer un pixel simplement comme un pixel associée ou non à un objet d'intérêt ou au fond de la zone d'intérêt un pixel est attribué dans une de quatre classes avec un poids d'appartenance à cette classe. Ces poids définiront les niveaux de fiabilité que les pixels désignent le fond d'image, ou un objet. Les conditions d'attribution aux pixels des classes et de poids sont présentés dans le tableau 3.

#### 4.4.5.3 Principe de classification de pixels d'une région continue

Pour la classification de pixels  $p_{\text{converted}}$  dans une zone d'intérêt en termes de MCC, ils obtiennent des poids  $W(p_{\text{converted}})$ , en fonction du poids du patch, auquel le pixel  $p_{\text{converted}}$  appartient :  $W(\text{patch})$ . Ces poids sont définis par les rapports suivantes :

$$W(p_{\text{contour}}) = \frac{Y(p_{\text{contour}})}{\alpha} \text{ et } W(\text{patch}_i) = \frac{\text{mean}(Y(\text{patch}_i))}{\alpha}$$

Les poids, estimés par ces formules, donneront ensuite une caractéristique au pixel, qui déterminera son effet dans l'estimation du coefficient de fiabilité. Avant d'expliquer ces caractéristiques, la formule qui estime le coefficient et son principe seront présentés.

Une classe est attribuée à un pixel  $p$  en fonction d'une paire de valeurs : son niveau de gris  $Y(p)$  et le niveau de gris de ces voisins dans le patch 3x3, qui est associé à ce pixel :  $\text{mean}(\text{patch}_i)$ , où  $i$  est le numéro du patch 3x3 avec la moyenne de niveau de gris la plus élevée dans le patch 5x5 qui le contient.

Le seuil  $\alpha$  de sélection des contours d'objet est une valeur minimale, à partir laquelle les pixels d'une zone d'intérêt sont considérés comme les contours d'objets. Ainsi une zone peut contenir également des pixels avec des valeurs de niveau de gris supérieures à celle de  $\alpha$ .

Pour cette raison, tous les pixels de l'ensemble, qui construisent une région continue après la conversion des contours d'un objet ont des effets différents sur les valeurs des moyennes des patches 3x3 dans leurs patches 5x5 associées.

L'approche, conçue pour la classification et la pondération des pixels d'une région consiste à comparer les proportions de la valeur du seuil  $\alpha$  d'une zone d'intérêt donnée avec de moyennes de cinq patches 3x3 qui se superposent dans le patch supérieur 5x5, associée au pixel  $p$ . Ces proportions définissent la direction et le patch 3x3 pour remplissage pour la conversion des contours en objet et le

pois du pixel  $p$  pour sa classification. Les classes de pixels et les règles d'appartenance sont présentées dans le tableau ci-dessous :

**Tableau 3 : Paramètres et conditions de conversion de très petits objets**

Classe attribuée	Type de pixels	Poids attribué	Niveau de gris $Y(p)$ attribué	Conditions et principe d'estimation
TP («true positif») : détections correctes	un objet dans une zone d'intérêt : ses contours ou son intérieur	1	$\text{Max}(\alpha, Y(p))$	$Y(p) \geq \alpha$ $W(\text{mean}(\text{patch}_i)) > W(\text{patch}),$  $W[(\text{mean}(\text{patch}_i)) \geq 1]$ && $W(\text{patch}) \geq 1] \parallel [(\text{mean}(\text{patch}_i)) \geq 1]$ && $W(\text{patch}) < 1]$
		$\frac{\text{mean}(\text{patch}_i)}{\alpha}$		$Y(p) \geq \alpha,$  $W(\text{mean}(\text{patch}_i)) > W(\text{patch}),$  $[(\text{mean}(\text{patch}_i)) < 1 \ \&\& \ W(\text{patch}) < 1]$
TN («true négatif») : détections correctes	le fond d'image	1	0	$Y(p) = 0$
FP («false positif») : fausses détections d'objets d'intérêt	du bruit ou des faux contours de transmission de couleurs, mais qui sont considérés comme des parties d'objet	$\left( \frac{\text{mean}(\text{patch}_i)}{\alpha} + \frac{Y(p)}{\alpha} \right) * \frac{1}{2}$	$\frac{\text{mean}(\text{patch}_i)}{\alpha} * 255$	$Y(p) \geq \alpha,$  $W(\text{mean}(\text{patch}_i)) < W(\text{patch}),$  $[(\text{mean}(\text{patch}_i)) < 1 \ \&\& \ W(\text{patch}) < 1]$
FN («false négatif») : fausses détections du fond d'image	des objets d'intérêt dans l'image, mais qui sont considérés comme le fond d'image	$\left( \frac{\text{mean}(\text{patch}_i)}{\alpha} + \frac{Y(p)}{\alpha} \right) * \frac{1}{2}$	$\frac{Y(p)}{\alpha} * 255$	$Y(p) < \alpha$  $W(\text{mean}(\text{patch}_i)) > W(\text{patch})$  $[(\text{mean}(\text{patch}_i)) < 1 \ \&\& \ W(\text{patch}) < 1]$

Le principe de cette pondération de pixels dans le contexte des régions, générées après la conversion des contours est de les intégrer dans la formule standard du coefficient de corrélation de Matthew en fonction de fiabilité de leurs classifications dans ces 4 classes. Si la classification d'un pixel n'est pas assez fiable pour l'attribuer définitivement dans l'ensemble de pixels d'un objet d'intérêt, ou à la classe du fond d'image, ce pixel, la fiabilité de sa propre conversion sera reflétée dans la formule. De la même manière, en fonction de la complexité d'une scène, un pixel peut correspondre à un objet, ainsi qu'aux autres phénomènes dans une image. Dans ce cas un poids sera attribué à un pixel en tant qu'à une fausse détection et compris dans la formule de MCC proportionnellement à son intensité dans une zone d'intérêt.

#### 4.4.5.4 Principe de l'estimation du coefficient de fiabilité et son effet sur le résultat final

En fonction de classification de pixels en quatre classes, leurs poids contribuent à un de quatre termes de la formule de MCC, correspondantes à ces quatre classes.

$$DR_{MCC}(B) = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TP + FP)(TP + FN)}}$$



Le deuxième coefficient  $DR_{MCC}(O)$  est alors estimé en fonction d'importance individuelle de chaque pixel après la conversion de contours en régions, et de celui de son voisinage au niveau local dans une zone d'intérêt. Cette approche permet d'attribuer des coefficients de fiabilité conformément aux contenus des zones, et prendre en considération la qualité de détection de chaque pixel dans sa classe associée.

Le coefficient  $DR_{MCC}(B)$  est estimé globalement pour une zone, car il analyse l'intégralité des pixels de cette zone : ceux, considérés comme les parties d'objets d'intérêt et ceux, qui représentent le fond d'image. Cette méthode reflète la fiabilité de détection des objets dans une zone en fonction de la qualité de ces contours et de sa conversion.

$DR_{MCC}(B)$  est le dernier coefficient de fiabilité, qui est appliqué sur l'intégralité d'objets dans une zone d'intérêt. Il est combiné avec les autres coefficients de ce niveau et converge l'estimation de fiabilité de détections vers le niveau local d'un objet d'intérêt.

Les figures 4.37 et 4.38 illustrent l'effet de pondération des pixels au cours de la conversion de contours d'objets en régions continues par patches. La première image est l'image d'origine, la deuxième est un masque de contours obtenu par un simple seuillage, où tous les pixels dont l'intensité est supérieure à  $\alpha$ , obtiennent la valeur de 255. Les autres conservent leurs valeurs de niveaux de gris estimées lors de la conversion. La troisième image est un masque de contours binarisé au moyen d'application de poids de contours.

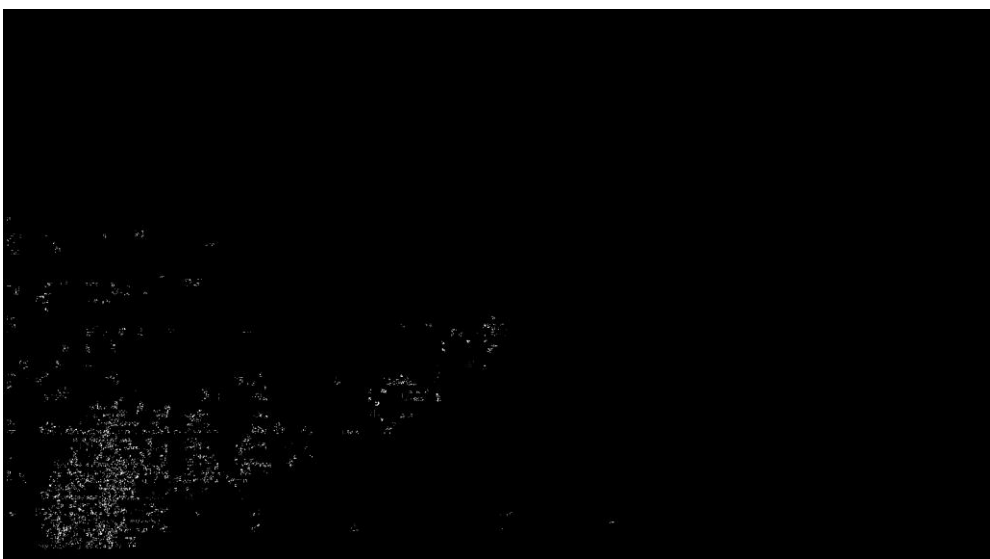
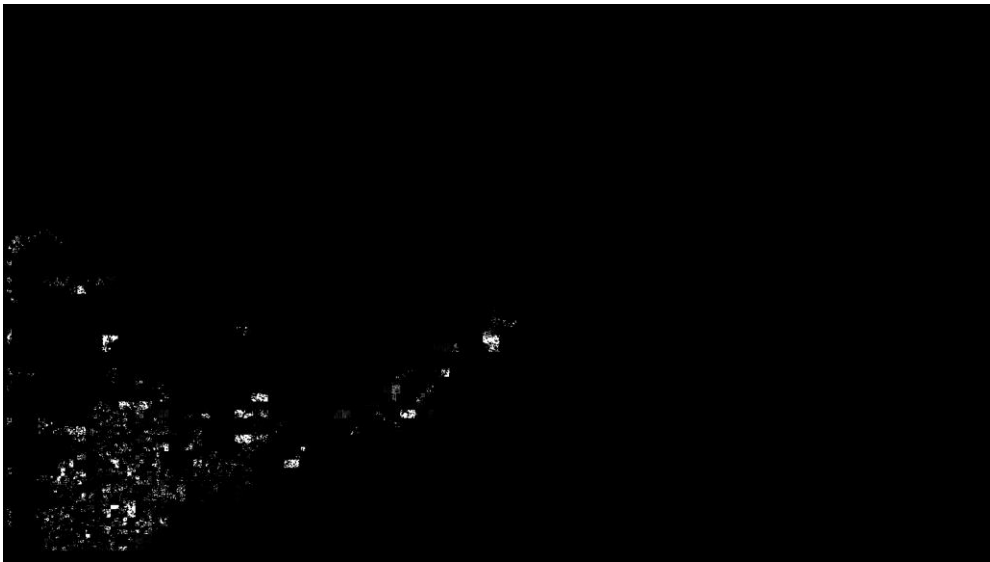


Figure 4.37 : Exemple 1 de conversion de contours d'objets d'intérêt (image du milieu) et de son amélioration en fonction des poids de pixels dans le contexte de MCC (image du bas).

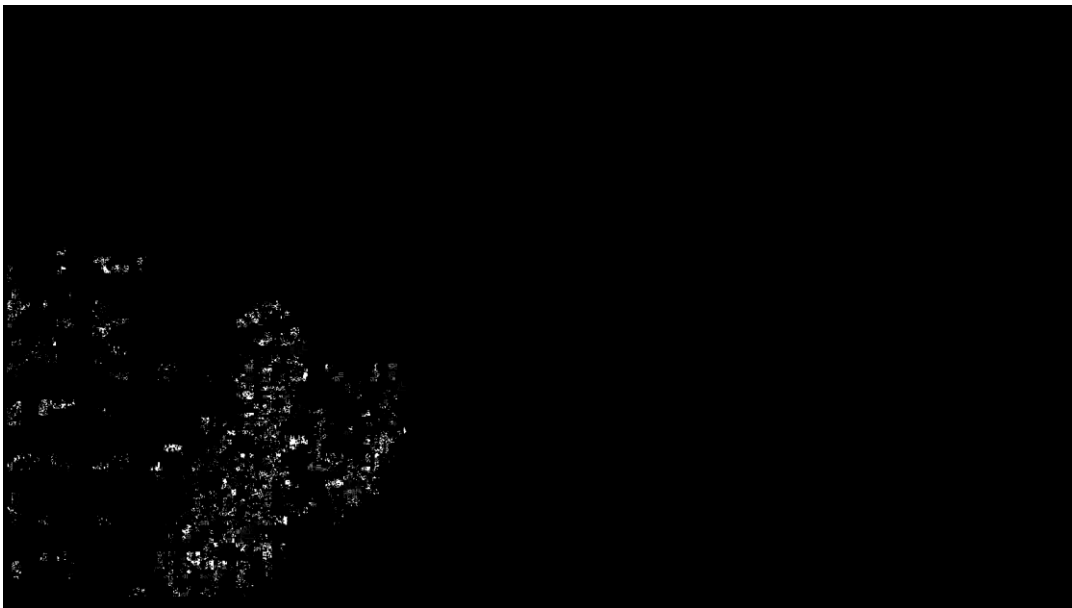


Figure 4.38 : Exemple 2 de conversion de contours d'objets d'intérêt (image du milieu) et de son amélioration en fonction des poids de pixels dans le contexte de MCC (image du bas).

#### 4.4.5.5. Exemples d'estimation du coefficient de fiabilité

Sur cette étape intermédiaire de détection d'objets d'intérêt, les coefficients de fiabilité caractérisent la fiabilité du processus de leur reconstruction depuis des données, dont la nature n'est pas connue (e.g. les contours peuvent représenter autant des contours d'objets avec un faible contraste, que des variations de couleur sur le fond d'image d'une basse qualité). Les objets reconstruits sur cette étape serviront également à l'estimation des paramètres du modèle d'objet d'intérêt à l'étape suivante, d'où l'importance d'y fournir les objets, dont la forme apparente et la plus nette et fiable.

La figure 4.39 ci-dessous récapitule toutes les étapes du traitement précédentes qui permettent d'arriver d'une image, extraite d'une vidéo d'entrée, à son masque binaire d'objets détectés. A cette étape les objets associés au bruit peuvent aussi obtenir des coefficients de fiabilité élevés lors de la conversion, car le filtre Laplacien conçu s'adapte aux différents contrastes sur une image, afin de ne pas perdre des petits objets avec un faible contraste. A ce stade on considère qu'il n'y a pas encore de modèle de référence pour des objets d'intérêt et que tous les objets peuvent être importants dans les traitements suivants.

Le tableau 4 récapitule les résultats de combinaison des coefficients de fiabilité sur les étapes précédentes de l'algorithme et présente les résultats d'application du coefficient de fiabilité de conversion seul et en combinaison. Ces résultats montrent que le fait de combiner graduellement les coefficients de fiabilité réduit au fur et à mesure le nombre de fausses détections et approche le résultat de dénombrement à la vérité terrain.

**Tableau 4. Résultats d'application du coefficient  $DR_{MCC}$  et de sa combinaison avec les coefficients de fiabilité des étapes précédentes.**

Corpus	Vérité terrain	Détections avec des combinaisons de coefficients				
		sans R(O)	$DR_{MCC}$		HR + DR + $DR_{MCC}$	
			TW	FW	TW	FW
1	68	559	447.2	33.5	52.5	11.6
2	40	816	660.9	62.4	15.7	15.6

La figure 4.39 montre un exemple, où, en fonction de la qualité d'une vidéo et la complexité d'une scène les contours de certaines régions sont convertis en « objets-parasites » ou en bruit. Ce fait s'explique par le phénomène de numérisation d'images et des traitements numériques, effectués. La modélisation d'une forme apparente de petits objets peut être alors affectée par les résultats des étapes intermédiaires. Dans ce cas, l'estimation de paramètres du modèle de petits objets d'intérêt peut être affectée par ces « objets-parasites », visuellement différentes, mais numériquement (par les valeurs de leurs paramètres) proches aux objets d'intérêt.

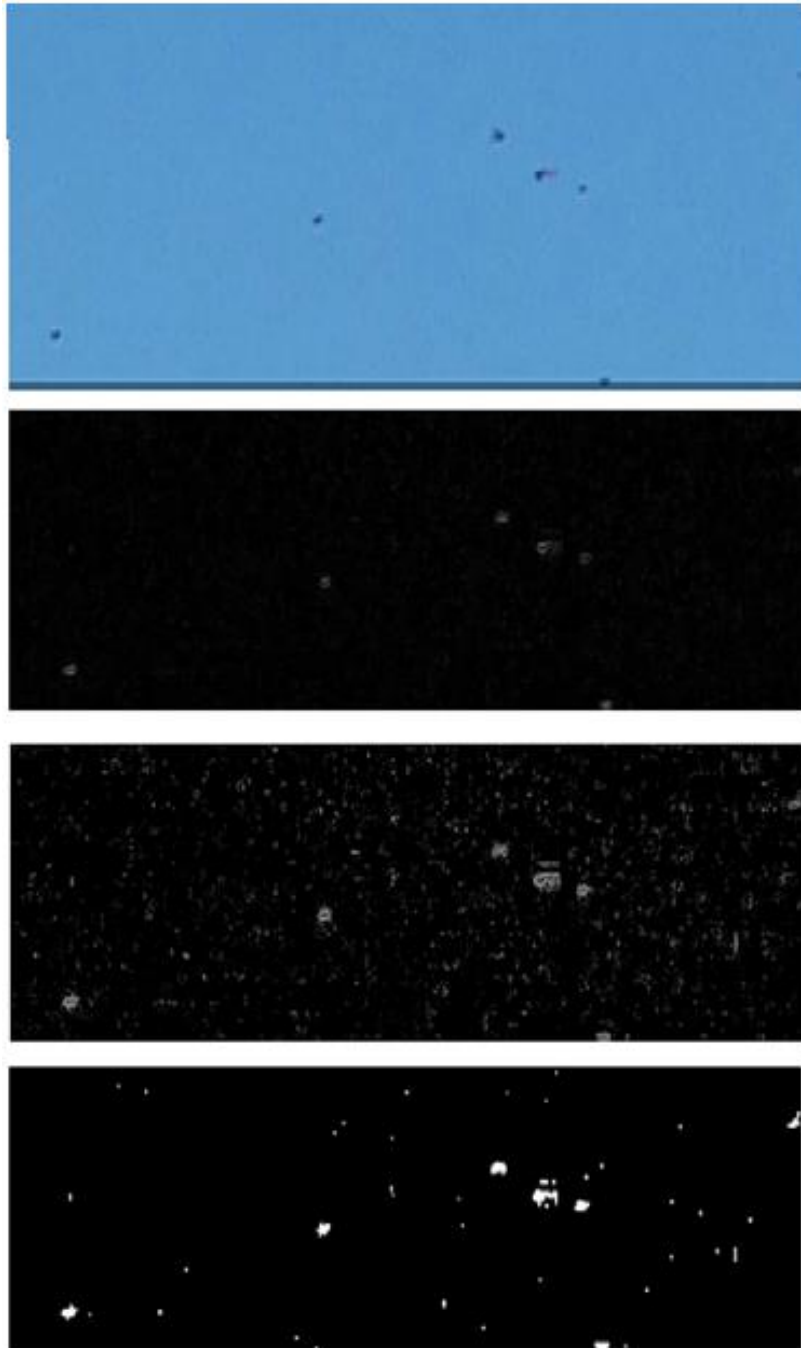


Figure 4.39 : Etapes de détection d'objet d'intérêt dans une image d'entrée sur une vidéo du corpus 1.

#### **4.5 Aspect temporel et sa contribution dans l'aspect spatial**

Les chapitres précédents se sont concentrés sur les aspect spatiaux de la méthode de détection et de comptage d'objets d'intérêt proposé dans le cadre de cette thèse. Par nature les objets d'intérêt visés ici (les oiseaux) ont une dimension temporelle et pas simplement spatiale. De ce fait, une phase d'analyse temporelle doit logiquement compléter la phase d'analyse spatiale.

##### **4.5.1 Introduction de l'aspect temporel**

L'introduction de l'aspect temporel dans la solution a plusieurs objectifs :

1. Analyse du contexte dynamique de la scène d'une vidéo.
2. Elimination des fausses détections présentes à l'issue de la phase d'analyse spatiale.
3. Construction d'un ensemble de données pour l'estimation des paramètres dynamiques du modèle de petits objets.

L'aspect temporel se base sur les hypothèses introduites précédemment (voire la section 4.1) :

- Régularité de détections des objets d'intérêt d'une même nature.
- Déplacements continus et réguliers d'objets tout au long d'une vidéo.

Dans le contexte d'application de l'algorithme pour le dénombrement des oiseaux migratoires, le suivi d'objets un par un dans des vidéos contenant des grands vols d'oiseaux n'est guère possible pour les raisons suivantes :

- L'apparence d'un très petit objet (une taille de quelques pixels, une couleur souvent proche du fond d'images, des contours irréguliers) ne permet pas de l'identifier dans un groupe et de suivre exactement le même objet dans la séquence d'images.
- Présence d'un groupe d'objets d'intérêt à forte densité spatiale et similaires par leurs formes apparentes et leurs couleurs ne permettant pas de les suivre individuellement.

Pour résoudre les difficultés d'identification et de suivi de mêmes objets, et en supposant que les objets se déplacent par groupes avec une même vitesse constante, ces vols sont suivis par « clusters » caractérisés par les barycentres de coordonnées d'objets qu'ils contiennent.

Le suivi de clusters dans une vidéo est réalisé à l'aide d'un algorithme de partition d'un ensemble (« union-find »). Cet algorithme trouve une application pour le suivi de composantes connexes dans un graphe non-orienté, dont la structure permet de modéliser les déplacements de clusters d'un frame à l'autre (les clusters d'objets sont les sommets de graphe et leurs déplacements – les connexions entre les mêmes groupes d'un frame à l'autre représentent les arêtes de ce graphe). Les connexions entre les clusters sont établies en fonction de la vitesse de déplacements de clusters.

Dans ce contexte, une méthode d'analyse temporelle a été développée par ailleurs dans le cadre du projet Octobre Bleu, et intégrée dans le cadre de cette thèse pour obtenir les résultats finaux issus de l'analyse spatio-temporelle. Cette méthode se décompose selon es phases algorithmiques suivantes :

1. Création de clusters d'objets dans chaque frame : des groupes d'objets avec une caractéristique commune. Pour les vidéos avec des oiseaux migratoires, qui se déplacent dans des vols groupés la distance entre les objets d'un même groupe est utilisée en tant que cette caractéristique.
2. Calcul du barycentre des objets dans chaque cluster.
3. Liens temporels entre les clusters des frames successives : Ces liens sont créés en fonction de la vitesse et de la direction de chaque cluster basé sur le déplacement de leur barycentre.
4. Sélection de groupes d'objets en mouvements : L'établissement des liens temporels entre clusters d'images successives, permet d'obtenir des clusters spatio-temporels sur une durée temporelle pouvant atteindre toute la vidéo, ou à l'inverse, avoir des clusters isolés sur une seule frame.

L'application de la régularité de déplacements par l'aspect temporel permet de compléter l'aspect spatial (apparence d'objets d'intérêt) et de sélectionner des groupes d'objets dont les déplacements peuvent être suivis dans une vidéo.

Ce regroupement en clusters d'objets permet alors :

- 1) de « filtrer » les objets immobiles de l'ensemble de potentiels objets d'intérêt
- 2) Il permet l'estimation des paramètres dynamiques du modèle d'objets d'intérêt
- 3) Il permet d'affiner les coefficients de fiabilité associé à chaque objet.

Après ce filtrage temporel on obtient un ou des ensembles d'objets d'intérêt en mouvement, caractérisés par les coefficients de fiabilités de la détection spatiale et regroupés par la proximité de leurs coordonnées. Ils pourraient servir directement pour la modélisation théorique de petit objet, mais en raison de difficultés de détection de petits objets ces ensembles contiennent en pratique des faux objets d'intérêt, dû aux phénomènes aléatoires d'une image à l'autre (des reflets de lumière, variations de teinte de couleur du ciel...). Si un tel phénomène se produit près d'un groupe d'objets d'intérêt en mouvement, il sera considéré comme un objet d'intérêt par l'aspect spatial (la fiabilité de détection) et temporel (car ses coordonnées seront comprises dans un groupe d'objets en mouvement). Pour éviter ces fausses détections l'estimation de paramètres du modèle d'objets d'intérêt, décrite dans la section suivante 4.5.2, prend en considération la régularité de détections par l'aspect spatio-temporel tout au long du traitement d'une vidéo. La notion de la régression linéaire du domaine de statistique permet un moyen pour intégrer l'hypothèse de la régularité de détection d'objets d'intérêt en mouvement. La notion et l'application de la régression linéaire sont présentées dans la section 4.5.3.1.

#### 4.5.2 Prise en compte des coefficients de fiabilité liés à l'aspect temporel

Les aspects spatiaux et temporels doivent être appliqués conjointement afin d'optimiser le processus de détection et de réduire au mieux les fausses détections. Cela suppose d'intégrer l'aspect temporel aux coefficients de fiabilité.

Le coefficient de fiabilité  $C_{temporel}(O)$ , associé au déplacement d'un objet est calculé à l'issue de la méthode qui implémente l'analyse par l'aspect spatiale. Ce coefficient consiste en une évaluation du suivi de la trajectoire pour un groupe d'objets d'intérêt et sa conformité à l'hypothèse de régularité des déplacements, i.e. de leurs vitesses et leurs directions. En d'autres termes, si plusieurs groupes d'objets d'intérêt sont détectés dans une vidéo, la méthode implémentée pour l'aspect spatial peut classifier ses trajectoires en tant que plus ou moins fiables et leur attribuer des coefficients de fiabilité temporels associées :

- Si pour un groupe d'objets sa trajectoire est linéaire et régulière (i.e. sans déplacements erratiques et le groupe est présent dans tous les frames analysés, respectivement), le coefficient de fiabilité temporel se rapproche de 1
- Si, au contraire, un groupe d'objets n'est pas régulièrement présent dans des frames consécutives et/ou ses déplacements sont erratiques, alors le coefficient de fiabilité temporel se rapproche de 0.

Ainsi, avant d'évaluer la qualité d'estimation de paramètres du modèle d'objet d'intérêt (voir la section 4.5.3), l'aspect temporel caractérise son côté dynamique et attribue en même temps un coefficient de fiabilité de déplacement  $C_{temporel}(O)$  aux objets qui ont servis à la construction du modèle.

Etant complémentaires pour l'estimation de fiabilité de détection de petits objets d'intérêt, ces deux aspects intègrent les coefficients de fiabilité finaux sur la dernière phase d'analyse. Cette dernière phase consiste en une combinaison de coefficients, estimés sur toutes les étapes de la solution, elle est détaillée dans la section 4.6.1.

### 4.5.3. Troisième phase : Estimation des paramètres du modèle théorique de petits objets et évaluation de conformité

Pour permettre l'estimation de paramètres du modèle d'objet d'intérêt, on suppose que la majorité des objets inclus dans les clusters sélectionnés par l'aspect temporel sont des objets d'intérêt et que leurs déplacements sont réguliers d'un frame à l'autre. Ces clusters sélectionnés permettent de modéliser la forme apparente de petits objets d'intérêt, qui servira de référence pour l'estimation de coefficients de conformité au modèle d'objet.

En fonction de la distance entre les objets d'intérêt et la caméra, la taille d'objet sera plus ou moins petite et sera estimée par une des deux notions suivantes :

- Les paramètres de la plus petite ellipse englobant un objet d'intérêt : ses deux semi-axes et sa taille (exprimée en nombre de pixels dans la région englobée).
- Le nombre de pixels dans une région associée à un objet d'intérêt, si les longueurs de ses axes rapprochent la limite de visibilité estimée (voir la section 4.3.5). Dans ce cas, le plus fréquent dans le contexte applicatif visé ici, les régions associées aux objets détectées ne peuvent pas être approximées à une primitive géométrique.

Sans prérequis concernant la nature et les types d'objets dans une vidéo, on suppose, qu'il peut y avoir plusieurs types d'objets d'intérêt, visuellement distinctes (e.g. plusieurs espèces d'oiseaux, des types variés de véhicules sur une autoroute etc.). Afin de vérifier cette hypothèse, on applique la condition suivant, basée sur la propriété de la distribution normale :

- Si dans une distribution normale  $STD \geq M$ , l'ensemble de données contient des échantillons variés appartenant à des distributions différentes.
- Si dans une distribution normale  $STD < M$ , les échantillons de l'ensemble appartiennent à la même distribution

Avec ces conditions l'estimation de paramètres du modèle ou des modèles d'objets d'intérêt se fait en plusieurs étapes :

1. La distribution normale de chaque paramètre  $p$  est caractérisée par la moyenne  $M_p$  et la déviation standard  $STD_p$
2. Si pour un paramètre  $p$   $STD_p > M_p$ , il est considéré qu'au même moment d'exécution  $t$  il y a des distributions différentes des paramètres, et donc des objets avec de formes différents. Si  $STD_p < M_p$  les objets détectés sont attribués à la même distribution.
3. Si les valeurs d'un de paramètres appartiennent à des distributions différentes, les objets, détectés dans la frame sont repartis en plusieurs classes tels que :
  - a. la classe 1 d'objets d'intérêt contient les objets, pour lesquels tous les paramètres de forme sont inclus dans l'intervalle  $[M_p - STD_p, M_p + STD_p]$
  - b. l'autre classe contient le reste de détections. Sa segmentation continue jusqu'au moment où tous les objets soient répartis dans les classes, pour lesquelles la condition ci-dessus est satisfaite

Après ce traitement d'un intervalle de  $T$  frames les classes successivement construites sont analysées pour leurs régularités par l'équation de régression linéaire (voir la figure 4.42). Les hypothèses pour cette analyse sont dérivées des hypothèses générales de la solution (section 4.1) :



- Les mêmes classes ont le nombre constant des objets attribués dans chaque frame (des légères différences sont acceptées).
- Les classes régulièrement présentes dans les frames successifs ont un nombre de détections cumulés en augmentation avec chaque nouveau frame traité.

#### 4.5.3.1. Application de la régression linéaire pour la modélisation de petits objets d'intérêt

Dans le domaine de la statistique l'analyse de paramètres de l'équation de régression linéaire [48] est une des techniques utilisées pour la modélisation et l'explication des relations entre la variable à expliquer  $Y$  et les variables explicatives :  $\{x_i\}$ .

En appliquant ces notions pour le cas de détection de petits objets, la variable à expliquer  $Y$  est le nombre de détections dans chaque frame cumulé avec les détections dans les frames précédents, et les variables explicatives  $\{x_i\}$  sont les nombres de frames traités.

L'équation théorique suivant définit le modèle de régression linéaire :

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

où,  $\beta_0$  et  $\beta_1$  sont les estimateurs des Moindres Carrés Ordinaires et  $\varepsilon$  signifie les erreurs centrées, non-corrélées entre elles et ayant la même variance que celle de la distribution de variables correspondantes.

D'après l'hypothèse que le nombre d'objets d'intérêt est constant tout au long d'une vidéo, (section 4.1) on suppose que dans le cas idéal on détecte le même nombre d'objets à chaque instant  $t$ . Cela signifie que le nombre cumulé d'objets d'intérêt du même type augmente à la même valeur avec chaque nouveau frame traité. En présentant l'ensemble de détections dans un intervalle de  $T$  frames sous forme d'histogramme cumulé, l'évolution du nombre de détections peut être modélisée par une fonction linéairement croissante (dont l'équation générale  $y = b_0 + b_1 x$  décrit la droite de la régression linéaire). Le modèle de la régression linéaire permet d'estimer les coefficients  $b_0$  et  $b_1$  aux moyens de la statistique et la méthode de Moindres Carrés Ordinaires (MCO).

En termes d'équation de droite, le coefficient  $b_1$  définira sa pente et sa direction, et le coefficient  $b_0$  – son décalage par rapport au centre du système de coordonnées. Graphiquement, cette droite sera centrée dans le nuage de points, qui représente le nombre de détections à chaque instant  $t$ . Le coefficient  $b_1$  définira alors le nombre moyen de détections à chaque instant  $t$ , et le coefficient  $b_0$  – le premier moment  $t_0$ , où les objets ont été détectés.

La section 4.5.3.2 suivante explique le principe de l'application linéaire basée sur la méthode de Moindres Carrés Ordinaires (MCO) pour la modélisation de petit objet d'intérêt.

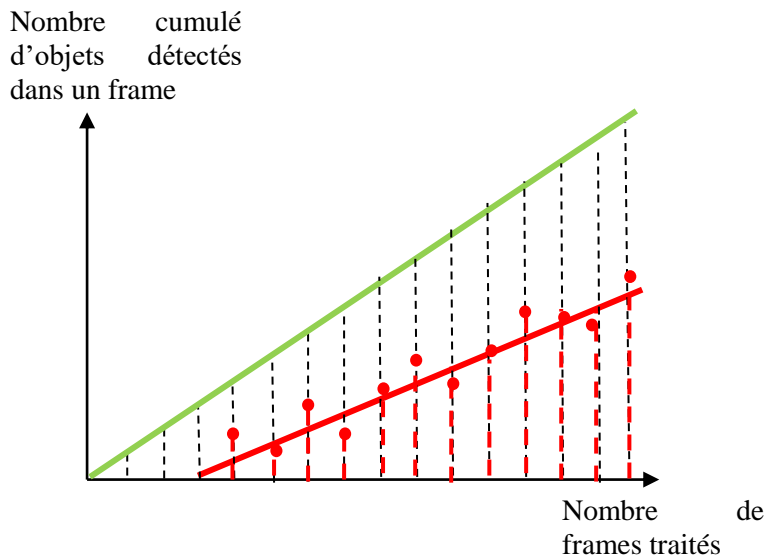


Figure 4.42 : principe de la régression linéaire et son application pour l'estimation de régularités dans les paramètres d'objets détectés. La droite verte : le cas idéal d'après les hypothèses de l'algorithme – le même nombre d'objets est détecté dans chaque frame. La droite rouge : approximation de la droite de régression linéaire dans le cas réel – des variations de nombre de détections dans les frames consécutifs. Les lignes pointillées rouges désignent l'évolution du nombre cumulé d'objets détectés, qui varie en cours de traitement d'un frame suivant.

#### 4.5.3.2. Composition de la méthode de MCO et de la régression linéaire pour la modélisation de petits objets d'intérêt

Sans perdre en généralité pour la preuve d'utilisation de la méthode de MCO, on suppose que dans une vidéo il n'y a qu'une classe d'objets détectés dans chaque frame. A chaque frame  $x_i$  de l'intervalle de T frames le nombre d'objets détectés  $y_i$  est associé. Pour modéliser la fonction de régression linéaire les estimateurs sont estimés depuis la moyenne de détections dans un frame  $\bar{y}$  et la moyenne des nombres de frames  $\bar{x}$ :

$$\beta_1 = \frac{\sum_t^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_t^T (x_t - \bar{x})^2}; \beta_0 = \bar{y} - \beta_1 \bar{x}$$

L'exactitude du modèle de régression linéaire avec les coefficients  $\beta_0$  et  $\beta_1$  (i.e. à quel point il correspond aux données réelles) est estimée par le coefficient de corrélation  $C^2$ :

$$C^2 = \left( \frac{\sum_t^T (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_t^T (x_t - \bar{x})^2} \sqrt{\sum_t^T (y_t - \bar{y})^2}} \right)^2$$

La valeur de  $C^2 \in [0, 1]$  et permet d'évaluer la qualité du modèle statistique construit :

- Si  $C^2=0$ , le modèle estimé n'est pas adapté aux données et ne les modélise autour de la moyenne
- Si  $C^2=1$ , le modèle explique tout et les données-échantillons sont alignés sur la droite décrite par l'équation.

Généralement,  $C^2 \in (0, 1)$  et sert à caractériser la qualité du modèle de régression linéaire et sa conformité aux données réelles.

Par sa définition, le modèle de la régression linéaire devrait minimiser les distances entre la droite et les points représentant les données par une équation, adaptée aux données. Mais en pratique, et dans le cas d'application, quand l'ensemble d'objets d'intérêt contient également de fausses détections, la droite n'est pas alignée parfaitement aux données. Cette erreur – la variance résiduelle – est également utilisée pour évaluer la qualité du modèle, et estimée par la solution. La section suivante 4.5.3.3 présente sa définition théorique et son application dans le cas de détection et dénombrement d'objets d'intérêt dans les vidéos.

#### 4.5.3.3. Analyse de la variance résiduelles pour l'estimation de qualité du modèle

Les résidus sont les différences entre les données réelles et des valeurs, estimées par le modèle. Ils sont calculés par la formule suivante :

$$\varepsilon_i = y_i - \hat{y}_t = y_i - \beta_0 - \beta_1 x_i$$

où  $y_i$  représente la valeur réelle des données pour la variable  $x_i$  et  $\hat{y}_t$  - la valeur correspondante estimée par le modèle de la régression linéaire pour la même variable  $x_i$  (voir la figure 4.42).

Pour évaluer la qualité du modèle construit, plusieurs caractéristiques suivantes sont estimées depuis ces données :

- la variance de résidus
- la somme de carrés expliquée par le modèle (SCE)
- la somme de carrés total (SCT)
- le coefficient  $R^2$  de détermination du modèle sous forme du rapport de SCE et SCT. En d'autres termes le coefficient de détermination  $R^2$  estime à quel point le modèle estimé est proche aux données réelles.

Le coefficient de détermination  $R^2$  est calculé par la formule suivante :

$$R^2 = \frac{SCE}{SCT} = \frac{\sum_t^T (\hat{Y} - \bar{y}_t^1)}{\sum_t^T (\hat{Y} - \bar{y}_t^1) + \sum_t^T \varepsilon^2} = 1 - \frac{\sum_t^T \varepsilon^2}{\sum_t^T (Y - \bar{y}_t^1)}$$

Par cette estimation on peut faire la conclusion suivante sur le modèle de régression linéaire, construit dans le contexte de la modélisation de la forme apparente des petits objets dans la vidéo :

- Plus le rapport  $\frac{\sum_t^T \varepsilon^2}{\sum_t^T (Y - \bar{y}_t^1)}$  est proche de 1, plus le coefficient de détermination  $R^2$  est proche de 0. Cela signifie que la variation de résidus est proche de celle des données expliquées par le modèle. Dans ce cas, les objets dans la classe de détections décrite par ce modèle sont hétérogènes et leurs valeurs sont éloignées de la moyenne (i.e. variations de leurs paramètres sont trop élevées par rapport à leurs moyennes). Ce modèle représentera alors des objets-parasites d'une nature et apparence variées, plutôt qu'un ensemble d'objets d'intérêt.

- Plus le rapport  $\frac{\sum_t \varepsilon^2}{\sum_t (y - \bar{y}_t^1)}$  est proche de 0, plus le coefficient de détermination  $R^2$  est proche du 1. Cela signifie, dans ce cas le modèle de la régression linéaire construit décrit les données avec une précision suffisante, pour les associer aux paramètres du modèle d'objets d'intérêt

En fonction du nombre de types d'objets d'intérêt attendus dans une vidéo, un ou plusieurs ensembles, dont la variation de résidus et le coefficient de variation sont les plus petits, sont sélectionnés pour estimer les paramètres des modèles d'intérêt.

Ainsi les méthodes statistiques appliquées pour l'estimation de paramètres du modèle d'objet d'intérêt permettent de construire ce modèle en respectant l'hypothèse sur la régularité de détection d'objets d'intérêt tout au long d'une vidéo (par le nombre de ces détections linéairement croissant) et aussi évaluer la qualité du modèle construit avec les caractéristiques suivantes :

- le coefficient de corrélation  $C^2$  maximal
- la moindre erreur résiduelle

Les moyennes de paramètres d'objets dans l'ensemble sélectionné sont attribuées aux paramètres du modèle théorique d'objets qui servira à l'estimation de coefficient de fiabilité de conformité au modèle. Le principe d'estimation du coefficient de conformité de la forme d'un objet à celle du modèle est présenté dans la section 4.5.4.1.

#### **4.5.4. Estimation et rectification des coefficients de fiabilité de conformité au modèle d'objet d'intérêt**

Les situations avec des objets en mouvement, qui gardent leurs formes apparentes tout au long d'une vidéo arrivent rarement dans des vidéos réelles. Un objet en mouvement change naturellement sa forme et position qui résulte en modifications de sa forme apparente. Dans le cas de petits objets cela implique des variations dans le nombre de pixels qui le construisent, i.e. de sa taille. Dans cette application même des faibles variations à quelques pixels près ne sont pas négligeables, car ces différences peuvent entraîner des valeurs basses pour les coefficients de fiabilité de conformité au modèle même pour des objets d'intérêt. En fonction du niveau hiérarchique de la solution, les coefficients de fiabilité peuvent avoir effet sur un ensemble d'objets d'intérêt et la suite des traitements :

- La sélection des zones d'intérêt influence la fiabilité globale de la détection d'objets d'intérêt dans ces zones.
- Le nombre des contours détectés dépend de la complexité des textures.
- Les nombre de régions continues générées dépendent des contours sélectionnés à l'étape précédente

Il est alors nécessaire d'avoir des « valeurs de référence » pour les coefficients de fiabilité (souvent inférieur à 1), qui définiront des bonnes détections dans les conditions d'une scène filmée. Cela permet d'éviter les pertes des bonnes détections ou une sélection excessive d'objets-parasites, du bruit ou des faux contours.

#### 4.5.4.1 Estimation du coefficient de conformité au modèle

Tout d'abord, il faut préciser, que le coefficient de conformité au modèle d'objet reflète ses deux caractéristiques – sa forme apparente (ou uniquement la taille, en fonction du nombre de pixels qui compose un objet d'intérêt) et sa couleur. Les sections 4.5.4.1 et 4.5.4.2 expliquent l'estimation du coefficient de conformité à la forme, associée au modèle théorique, et la section 4.5.4.3 – l'estimation de conformité à la couleur théorique, associée au modèle.

Dans les sections, qui présentent le cas d'application sur suivi d'oiseaux migratoires pour la détection de petits objets d'intérêt (et dans la section 4.5.3 du présent chapitre), leurs formes apparentes peuvent être décrites par plusieurs paramètres. Le nombre de paramètres dépend de la distance entre les objets et la caméra, qui a un effet important sur la taille d'objets dans une vidéo. En conséquence, le premier coefficient de conformité dépend d'un seul paramètre si les objets sont trop petits pour approximer une forme à leurs apparences (la taille – le nombre de pixels, qui composent un objet). S'il est possible de trouver une forme approximative, commune pour les objets d'intérêt (dans le cas d'application la forme la plus proche est la primitive géométrique d'ellipse, sous forme de la plus petite ellipse, englobante l'objet), la similarité de ses paramètres est utilisée pour définir un coefficient de conformité à la forme du modèle théorique.

Si la taille d'objets permet une approximation à une ellipse, ses paramètres suivants sont utilisés pour estimer la conformité à la forme du modèle théorique d'objets d'intérêt :

- Sa taille, représentée par le nombre de pixels, qui composent un objet (il faut noter, que ce nombre n'est pas égale à la superficie de l'ellipse, calculée par la formule correspondante en raison de formes irrégulières d'objets)
- Ses paramètres : les longueurs de semi-axes, présentées sous forme d'une somme. Sans perdre en généralité, ses deux valeurs sont réduites en une seule, car on ne s'intéresse pas aux orientations d'objets.

Dans le cas, où les paramètres de la plus petite ellipse englobante décrivent un objet d'intérêt, son coefficient de conformité à la forme du modèle d'objets d'intérêt  $MR_1(O)$  est estimé par la fonction de densité de la distribution normale bimodale (Figure 4.44 (c)). C'est une fonction avec une équation plus complexe car elle agit en dimension 2D :

$$MR_1(O) = f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right]\right)$$

où  $x$  et  $y$  sont les valeurs de deux paramètres, qui décrivent la forme d'objets d'intérêt et  $\mu_x$  et  $\mu_y$  sont les valeurs des mêmes paramètres estimés pour le modèle d'objet. L'intérêt de son utilisation est la possibilité d'estimer le coefficient de fiabilité, en appliquant simultanément deux caractéristiques.

Si un objet est trop petit, uniquement sa taille est utilisée pour estimer sa conformité au modèle, car il devient compliqué de comparer les longueurs de ses semi-axes avec celles du modèle d'objet. L'ajustement de ces paramètres en fonction de la conformité globale des détections d'un moment à l'autre, où les objets sont comparés les uns avec les autres (voir la section 4.5.4.2) est aussi difficile, car les notions de la statistique de la moyenne et l'écart-type, dont elle se sert, résultent en très petites valeurs. Dans ce cas, l'estimation de sa conformité au modèle d'objet se réfère à la taille, estimée pour le modèle d'objet à l'étape de sa construction depuis les classes régulières (voir la section 4.5.3). La proportion suivante permet de comparer la différence entre la taille du modèle et celle d'un objet par rapport aux différences entre les objets, qui ont servis à construire le modèle, et d'associer à un objet son coefficient de conformité au modèle :

$$MR_1(O) = \frac{M_{size} - O_{size}}{STD_{size}}$$

#### 4.5.4.2 Approche proposée pour l'ajustement des coefficients de fiabilité de conformité au modèle

Le principe d'estimation de conformité d'objets au modèle prend en considération que des changements des formes apparentes théoriques d'objets sont possibles au cours du traitement des frames consécutives. Dans le cas de petits objets même de faibles variations de ces formes apparentes résultent en des variations de paramètres suffisantes pour qu'un objet ne corresponde plus au modèle théorique.

Dans ce contexte, le concept de coefficients de fiabilité permet d'analyser la conformité de chaque objet au modèle et d'ajuster les paramètres du modèle en cas de changements graduels et réguliers (e.g. si les objets s'éloignent lentement de la caméra, voir les hypothèses dans la section 4.1). Cette méthode permet ainsi de mettre à jour les paramètres du modèle, si besoin. Deux approches de rectification des coefficients de fiabilité ont été analysées :

- en termes de probabilité conditionnelle
- en termes d'approche fonctionnelle

L'approche basée sur la probabilité conditionnelle s'exprime en condition, qu'un objet détecté correspond à un objet d'intérêt, si les étapes précédentes du traitement ont été effectuées sur une zone, des contours et une région continue, dont les coefficients de fiabilité correspondent à un objet d'intérêt. Par définition, la probabilité conditionnelle procède par la multiplication de probabilités. Sur les coefficients de fiabilité, inclus dans l'intervalle  $[0, 1]$  elle résulte en valeurs de coefficients finaux bas même pour des détections fiables.

Une approche fonctionnelle a été envisagée pour éviter une diminution de coefficients de fiabilité. Ils sont présentés sous forme de fonctions, plutôt que de facteurs d'un produit. Cette solution permet de conserver les valeurs de coefficients et de les rectifier proportionnellement à la fiabilité d'exécution d'une étape de l'algorithme.

L'approche fonctionnelle d'estimation du coefficient de conformité au modèle d'objet se base sur les notions de la logique floue. En fonction des paramètres du modèle et des paramètres estimés lors des phases de traitement précédentes, chaque nouvelle frame  $t$  est traité en deux étapes :

1. Génération d'une fonction trapézoïde de la logique floue pour représenter la fiabilité de détections aux instants précédents (i.e. la conformité des objets précédemment détectés aux paramètres du modèle générés)
2. Ajustement par la fonction trapézoïde générée du coefficient de conformité au modèle pour un objet détecté à l'instant  $t$

La fonction trapézoïde sert à ajuster le coefficient de fiabilité d'un objet à l'instant  $t$ . Elle est générée depuis la fiabilité moyenne de détections à l'instant précédent  $t-1$  et la déviation standard dans cet ensemble de coefficients. L'objectif de cet ajustement est d'attribuer à un objet un coefficient de fiabilité de conformité au modèle proprement au moment  $t$ , et en fonction de la fiabilité moyen de détections d'objets conformes au modèle, détectés aux instants précédents et leurs différences du modèle (se sont paramètres  $c$  et  $a$  respectivement dans la fonction d'estimation du coefficient de conformité au modèle ajusté :  $\overline{MR}_1(O_{i,t})$ ). Formellement, l'estimation du coefficient de conformité au modèle est présentée par la fonction :

$$\overline{MR}_1(O_{i,t}) = f(MR_1(O_{i,t}), c, a)$$

où  $c$  est le coefficient de conformité au modèle moyen pour les objets détectés à l'instant  $t-1$ .

Le paramètre  $a$  représente un estimateur qui caractérise à quel point les objets détectés à deux instants consécutifs,  $t-1$  et  $t$ , sont similaires entre eux et à quel point cette similarité répond à la conformité au modèle, estimée pour les objets détectés à l'instant  $t-1$ . Le facteur  $a$  permet de limiter les déviations des paramètres spatiaux des objets détectés et ne sélectionner à l'instant  $t$  que des objets similaires à ceux détectés à l'instant  $t-1$ , et au modèle théorique d'objets d'intérêt.

Le paramètre  $c$  est défini comme

$$c = \frac{1}{Nb_{t-1}} \sum_{O_{j,t-1}}^{Nb_{t-1}} MR_1(O_{j,t-1})$$

avec  $Nb_{t-1}$  - le nombre d'objets détectés à l'instant  $t-1$ . Le paramètre  $a$  est estimé comme :

$$a = |c - S_o|; S_o = \frac{1}{Nb_{t-1}} \sum_{O_{j,t-1}}^{Nb_{t-1}} S(O_{i,t}, O_{j,t-1})$$

où

$$S(O_{i,t}, O_{j,t-1}) = \begin{cases} 0, & \text{si } \min_j Dist(O_{i,t}, O_{j,t-1}) \geq \sigma_p \\ \frac{\min_j Dist(O_{i,t}, O_{j,t-1})}{\sigma_p}, & \text{sinon} \end{cases}$$

$Dist$  définit la distance Euclidienne entre les ensembles de paramètres de chaque objet  $O_{i,t}$ , détecté à l'instant  $t$ , avec les paramètres de chaque objet  $O_{j,t-1}$  détectés à l'instant  $t-1$ . La différence  $\sigma_p$  entre les paramètres acceptés pour les considérer comme des objets similaires est définie comme :

$$\sigma_p = \sqrt{\frac{1}{Nb_{t-1}} \sum_i^{Nb-1} (V(M) - V(O_i))^2}$$

où  $V(M)$  et  $V(O_i)$  représentent respectivement les vecteurs de paramètres du modèle d'objet estimé et de l'objet  $O_i$ . Dans le cas de très petits objets  $V(M)$  et  $V(O)$  sont les valeurs des tailles, associées au modèle théorique et à un objet.

Pour l'estimation de conformité d'objets détectés au modèle théorique, on admet que ce dernier est suffisamment proche des apparences réelles des objets d'intérêt. On introduit au moyen d'une fonction de logique floue une flexibilité pour ce modèle, afin d'attribuer des coefficients de fiabilité appropriés aux objets avec des paramètres dans les intervalles de variations acceptés de  $t-1$  à  $t$ . Cette solution permet d'actualiser le modèle au cours du traitement, sans éloigner ses paramètres des paramètres du modèle initial.

Dans la solution proposée, on suppose, que si les paramètres d'un objet  $O_{i,t}$  ont une déviation du modèle de plus de 50% (i.e. le coefficient de fiabilité de conformité au modèle  $MR_1(O_{i,t})$  est inférieur de 0.5), cet objet ne correspond pas au modèle (son coefficient ajusté  $\overline{MR}_1(O_{i,t})$  se réduit alors à 0), et

la détection n'est pas fiable. D'autre part, si à l'instant  $t$  l'objet  $O_{i,t}$  reste similaire au modèle et aux autres objets d'intérêt, qui ont les coefficients de fiabilité élevée, alors  $\widehat{MR}_1(O_{i,t})$  est élevé à 1. Dans les autres cas, la nature de l'objet  $O_{i,t}$  n'est pas définie précisément en tant qu'une bonne ou fasse détection, et son coefficient  $\widehat{MR}_1(O_{i,t})$  est calculé par l'équation de la fonction de trapézoïde. Sa forme est définie par quatre paramètres, qui sont mis à jour au cours de l'exécution :

- $Mean(Model)_t$ : le coefficient de fiabilité minimal attendu au moment  $t$  en fonction des paramètres du modèle actualisé
- $c$  : le coefficient de conformité au modèle moyen pour les objets détectés au moment  $t-1$
- l'estimateur  $a$  : qui représente la différence acceptée entre les objets détectés à l'instant  $t-1$
- et le modèle d'objet théorique correspondant
- le coefficient  $MR_1(O_{i,t})$ , qui caractérise le conformité de l'objet  $O_{i,t}$  au moment  $t$

Dans les section suivantes le coefficient  $\widehat{MR}_1(O_{i,t})$  est définie comme  $C_{forme}(O)$ . Son estimation est illustrée graphiquement sur la figure 4.43.

$$\widehat{MR}_1(O_{i,t}) = f(MR_1(O_{i,t}), c, a) = \begin{cases} 0, & MR_1(O_{i,t}) < c - a \\ 1, & MR_1(O_{i,t}) > c \\ \frac{MR_1(O_{i,t}) - (c - a)}{Mean(Model)_t - (c - a)}, & \text{sinon} \end{cases}$$

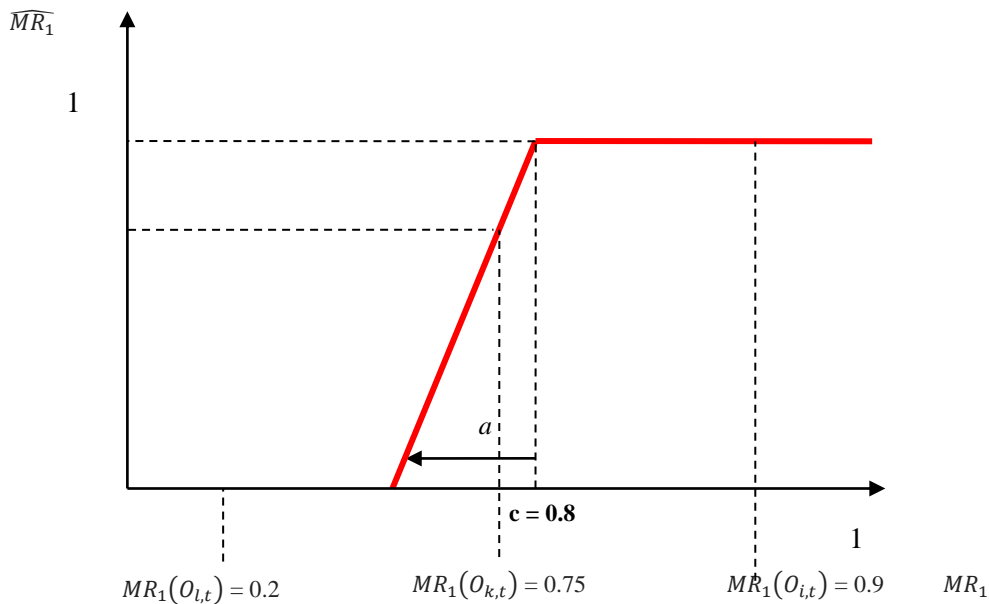


Figure 4.43 : Présentation graphique d'ajustement du coefficient de fiabilité de conformité d'un objet détecté à l'instant  $t$   $O_{i,t}$  au modèle théorique, actualisé à l'instant  $t-1$ . Cette figure illustre les trois situations, où les coefficients de conformités au modèle théorique sont ajustés en fonction des paramètres actualisé du modèle (objet  $O_l$  avec le coefficient initial à l'instant  $t$  obtient le coefficient égal à 0, ajusté par la fonction trapézoïde, représentant les paramètres du modèle d'objet à l'instant  $t-1$  (car sa conformité au modèle à l'instant  $t-1$  est inférieure à la conformité estimée pour l'ensemble d'objets à l'instant  $t-1$ ). Objet  $O_i$  avec le coefficient initial 0.9 obtient le coefficient égal à 1, i.e. il est considéré comme conforme au modèle à l'instant  $t$ . Le coefficient de l'objet  $O_k$  égal à 0.75 à l'instant  $t$  sera diminué et compris dans l'intervalle  $(c - a, c)$ , car sa conformité au modèle est différente de celle de l'ensemble d'objets conformes au modèle).



Ainsi, le concept de comparaison d'objets détectés avec le modèle théorique permet de prendre en considération les limites de petites tailles d'objets d'intérêt et d'éviter l'application direct de leurs paramètres spatiaux (en plus, dans le cas général d'application, leurs formes sont irrégulières et des contours peu contrastés, qui complique davantage l'analyse). Le concept conçu répond également au concept d'enchaînement de phases de la solution, où les traitements ont des effets sur le déroulement de la phase suivante.

Une approche pour la rectification du coefficient de fiabilité par une fonction limitée, permet de prendre en considération des légères évolutions du modèle au cours du temps, sans le rendre général, quand les faibles valeurs de coefficients signifient un objet conforme au modèle.

#### 4.5.4.3 Sélection d'objets conformes au modèle par la couleur théorique

D'après les hypothèses admises, un objet conforme au modèle théorique a une forme et une couleur proches de ce modèle. Un objet d'une forme proche de celle du modèle, mais d'une couleur différente ne sera pas conforme au modèle. Par le même principe, un objet de la bonne couleur, mais avec une forme différente de celle du modèle ne sera pas considéré comme une détection fiable et ne sera pas inclus dans le résultat final.

L'estimation de conformité de la couleur d'un objet à la couleur théorique du modèle est faite en utilisant la texture de la région correspondante dans le masque binaire. En pratique, la région dans le masque binaire ne correspond généralement pas exactement à l'objet détecté dans l'image (la région peut être plus petite ou plus grande que le vrai objet). Cette situation ce produit, par exemple, si certaines contours d'un objet sont moins contrastés sur le fond que les autres. Les pixels, qui les construisent, peuvent alors être associés au fond d'image et réduire aussi la superficie de l'objet dans le masque binaire. A l'inverse, des pixels du fond de la scène peuvent être intégrés à l'objet.

En pratique, ces différences ne font que quelques pixels, mais par rapport à une petite taille d'objet elles peuvent avoir un effet considérable sur l'estimation de sa couleur, ainsi que sur l'évaluation du coefficient associé de conformité par la couleur. Il est alors important d'analyser les étapes précédentes de détection de contours et de leur conversion en région continue afin de caractériser l'importance de chaque pixel individuellement. Dans ce contexte, l'estimation du coefficient de conformité à la couleur théorique  $MR_2(O)$  d'objet  $O$  se fait par le calcul de la distance pondérée  $CIEDE2000(O)$  entre la couleur de l'objet et la couleur théorique :

$$MR_2(O) = \begin{cases} 1 - \frac{CIEDE2000(O)}{\max(Dist_{color})}, & \text{si } CIEDE2000(O) < \max(Dist_{color}) \\ 0, & \text{sinon} \end{cases}$$

où la distance pondérée  $CIEDE2000(O)$  entre la couleur d'objet et la couleur théorique, associée au modèle est estimée comme

$$CIEDE2000(O) = \frac{1}{n} \sum_i^n w_{contour}(p_i) CIE2000(p_i)$$

Les distances entre les couleurs des pixels  $CIEDE2000(p_i)$  de la région considérée est la couleur théorique sont calculées par la formule CIEDE2000 [54] dans l'espace  $La^*b^*$ . Ensuite, ces distances individuelles entre les couleurs de pixels et de couleurs théoriques sont pondérées individuellement en fonction de l'importance de ce pixel dans la région reconstruite. Le poids  $w_{contour}(p_i)$  est ensuite estimé en fonction de l'intensité de contours, générés par le filtre adaptatif et le poids de ces pixels dans le masque de régions continues, associées aux objets, par le système suivant :

$$w_{contour}(p_i) = \begin{cases} 1, & \text{si } \frac{Y_{contour}(p_i)}{\alpha} \geq 1 \text{ et } w(Y_{region}(p_i)) = 1 \\ \max\left(\frac{Y_{contour}(p_i)}{\alpha}, w(Y_{region}(p_i))\right), & \text{si } w(Y_{region}(p_i)) > 0 \text{ et } Y_{contour}(p_i) < 1 \end{cases}$$

où  $w(Y_{region}(p_i)) = \frac{Y_{region}(p_i)}{255}$  et  $p_i$  fait partie de l'ensemble de pixels de la région associée à un objet (i.e. son niveau de gris est supérieur à 0).

$Y_{contour}(p_i)$  représente l'intensité d'un pixel de l'objet dans le masque de contours, généré par le filtre Laplacien adaptatif.  $Y_{region}(p_i)$  est l'intensité de ce pixel dans le masque de régions après la conversion de contours. Le poids  $w_{contour}(p_i)$  individuel au pixel  $p_i$  est déterminée en utilisant les conditions suivantes :

- si  $p_i$  appartient initialement à l'ensemble de contours (i.e. son niveau de gris est supérieur au seuil  $\alpha$ , qui donne la condition  $\frac{Y_{contour}(p_i)}{\alpha} \geq 1$ ) et est considéré comme un pixel de l'intérieur d'un objet lors de la conversion de contours (i.e. son poids  $w(Y_{region}(p_i)) = 0$ ), alors  $p_i$  fait partie de l'ensemble de pixels qui compose un objet d'intérêt
- si  $p_i$  ne fait pas partie de l'ensemble de contours ou un contour avec un niveau de gris inférieur au seuil  $\alpha$ , alors son poids estimé lors de la conversion de contours  $w(Y_{region}(p_i))$ , caractérisera le pixel  $p_i$ .

La valeur obtenue de  $CIEDE2000(O)$  - la moyenne pondérée de distances de la couleur de chaque pixel avec la couleur théorique - représente alors la couleur dominante d'un objet  $O$ . Elle est comparée ensuite avec la distance  $max(Dist_{color})$  - la différence entre les couleurs, acceptée pour considérer, qu'un objet  $O$  a une couleur proche de la couleur théorique.

#### 4.5.5 Coefficient final de conformité d'objet au modèle d'objet d'intérêt

Le coefficient de fiabilité  $C_{model}(O)$  est destiné à décrire, à quel point la détection d'un objet est fiable en fonction de trois caractéristiques, associées également au modèle d'objet :

- Les paramètres de forme apparente
- La couleur
- Les déplacements linéaires et réguliers d'un instant à l'autre

Le coefficient final  $C_{model}(O)$  de conformité au modèle d'objet reflète alors les aspects spatial et temporel. C'est-à-dire, qu'un objet doit satisfaire les deux conditions imposées pour être considéré comme un objets d'intérêt :

- la conformité au modèle d'objet d'intérêt par son apparence toute au long du traitement d'une vidéo
- les déplacements (i.e. l'évolution de ces coordonnées) linéaires et régulières d'une frame de la vidéo à l'autre

En d'autres termes, si l'apparence d'un objet n'est pas suffisamment similaire au modèle, ou ses déplacements s'éloignent de la trajectoire estimée pour les objets d'intérêt, son coefficient final de conformité au modèle sera diminué proportionnellement à ces écarts.

Le coefficient final de conformité au modèle est formalisé par l'expression suivante :

$$C_{Model}(O) = \min\{spatial[C_M(O)], temporel[C_M(O)]\}$$

où le coefficient de fiabilité par l'aspect temporel est composé de la même manière de sa fiabilité par la forme et par la couleur :

$$spatial[C_M(O)] = \min\{C_{colour}(O), C_{forme}(O)\}$$

## 4.6 Phase de comptage en fonction des coefficients de fiabilité

La phase finale de la méthode proposée consiste en un comptage de tous les objets détectés en fonction de leurs coefficients de fiabilité. Dans ce contexte l'objectif de traitements précédents était d'associer à chaque objet détecté un coefficient permettant de l'inclure dans le calcul du nombre d'objet en tenant compte de la fiabilité avec laquelle il a été détecté. La formule suivante rappelle le principe de comptage. ( $C(O_{i,t})$  est le coefficient de fiabilité final pour l'objet  $O_i$  dans le frame  $t$ ,  $T$  est le nombre de frames analysés dans une vidéo):

$$Nb_t = \sum_{O_{i,t}} C(O_{i,t}) Nb = \frac{1}{T} \sum_T Nb_t$$

Cette section explique, comment les coefficients, estimés tout au long d'exécution de la solution sont combiné en une seule valeur et quelles techniques sont appliquées pour atteindre cet objectif.

### 4.6.1 Principe de combinaison de coefficients de fiabilité d'objet d'intérêt en un coefficient final

Avant d'expliquer le principe de combinaison de coefficients de fiabilité en une seule valeur, on rappellera les étapes hiérarchiques de la solution proposée dans le cadre de cette thèse (voir la figure 4.16) :

- Etape 1 (niveau hiérarchique de zone d'image) : fiabilité de sélection d'une zone d'intérêt, probablement contenant des objets d'intérêt – estimation du contraste d'objets potentiels sur le fond d'une (section 4.4.3)
- Etape 2 (niveau hiérarchique de zone d'image) : fiabilité de détection de contours dans une zone donnée (section 4.4.4) et estimation de fiabilité de conversion des contours dans une région continue associée à un objet (section 4.4.5)
- Etape 3 (niveau hiérarchique d'un objet) : comparaison d'objets détectés avec le modèle théorique d'objet (section 4.5.3)

Chaque étape définit un coefficient de fiabilité caractérisant le niveau de certitude ou de fiabilité de la détection. Les coefficients de fiabilité sont alors liés les uns aux autres dans la structure hiérarchique de la solution.

En termes de logique floue, dont les fonctions sont à la base de l'estimation proposée des coefficients de fiabilité, par nature, ils caractérisent la certitude d'appartenance d'un élément à un ensemble flou. En fonction des étapes de traitements successifs, ce sont les appartenances :

- d'une zone à l'ensemble de zones d'intérêt
- des contours détectés à l'ensemble de contours d'objets
- de pixels à la région associée à un objet d'intérêt
- des objets détectés à l'ensemble d'objets d'intérêt

Dans ce contexte les notions et les techniques de la logique floue, permettent d'implémenter le principe de fiabilité de détections. Les coefficients de fiabilité sont associés à des fonctions

d'appartenance de la logique floue et réunis dans un ensemble de fonctions d'où le coefficient final est estimé en fonction de l'importance de chaque coefficient dans l'ensemble.

Avant de décrire le principe d'application des fonctions de la logique floue dans la solution, la section 4.6.2 présente ces fonctions, ainsi que les raisons de leurs choix et les modifications apportées dans le cas de détection de petits objets.

#### 4.6.2 Fonctions d'appartenance aux ensembles flous utilisés

Deux fonctions principales servent à estimer les coefficients de fiabilité tout au long de la solution [42][53] :

- 1) Fonction de sigmoïde (Figure 4.44 (a)). Cette fonction interprète le principe de «soft thresholding», le seuillage fonctionnel, où les valeurs au-dessous d'un seuil ne sont pas supprimées, comme en cas du seuillage classique. Cette fonction est appliquée plus tôt dans la solution – en deuxième étape, après le filtrage par l'opérateur Laplacien adaptatif (section 4.4.4). Elle sert à évaluer les valeurs de niveau de gris de pixels des contours sélectionnés par rapport au niveau de gris du seuil estimé pour de contours fiables, i.e. assez nets pour délimiter un objet d'intérêt sur le fond d'image. A la place, leurs proportions par rapport au seuil sont analysées pour calculer au final le coefficient de fiabilité de détection de contours (section 4.4.4). L'équation de la fonction sigmoïde est définie par deux paramètres ( $a$  et  $c$ ) le suivant :

$$f(x, a, c) = \frac{1}{1 + e^{-a(x-c)}}$$

- 2) Fonction trapézoïde (Figure 4.44 (b)). Cette fonction est définie par quatre paramètres et par le système suivant :

$$f(x, a, b, c, d) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & x \geq d \end{cases}$$

Cette fonction sert à deux objectifs :

- interprète le coefficient de fiabilité attribué à un objet sur une étape de la solution en fonction de la fiabilité globale de détection sur cette étape et dans le contexte du niveau hiérarchique.
- permet de combiner les coefficients de fiabilité de toutes les étapes du traitement sous forme d'une fonction, qui pourra prendre en considération les influences de chaque étape de séquence et les convertir en coefficient de fiabilité final de détection d'un objet.

La fonction trapézoïde est utilisée à la dernière étape de l'algorithme pour l'ajustement de coefficients en fonction de la fiabilité de détections tout au long du traitement d'une vidéo et la combinaison en valeur finale de la fiabilité de détection.

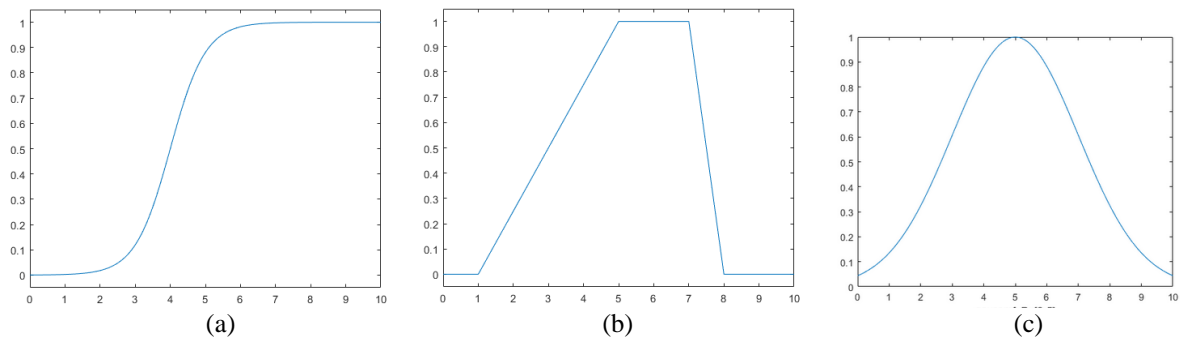


Figure 4.44 : fonctions d'appartenance aux ensembles flous de gauche à droite : fonction sigmoïde, fonction trapézoïde, fonction gaussienne 1D (source : documentation MatLab)

### 4.6.3 Méthodes de combinaison de coefficients

L'objectif de la combinaison de coefficients de fiabilité est d'en composer un seul, qui exprimera l'ensemble de fiabilité de détection sur toutes les étapes de traitement en une seule valeur. Les coefficients de fiabilité sont estimés à chaque phase algorithmique en fonction des méthodes appliquées. En pratique, ses phases sont liées les unes avec les autres car la fiabilité de détection à une étape précédente a un effet direct sur l'étape suivante. La méthode de combinaison de coefficients de fiabilité est alors très importante dans la phase finale de comptage.

Plusieurs opérations peuvent a priori être utilisées:

1. Produit de coefficients. Dans ce cas, la composition de tous les coefficients est présentée dans les termes de la probabilité conditionnelle. En théorie, la probabilité conditionnelle représente un produit de probabilités, associées à une observation, et dans le cadre de la solution pour la détection et dénombrement de très petits objets, ces probabilités représentant les fiabilités de détection, estimées à chaque étape. Or la multiplication de valeurs élevées en termes de probabilité (0.8, 0.9, qui représentent des bonnes fiabilités et sont logiquement associées à des bonnes détections) diminue considérablement le coefficient final (e.g. un coefficient de 0.9 obtenu à chacun de cinq critères d'une bonne détection – pour la zone, la détection de contours d'objets, la conversion de contours dans une région continue, la conformité au modèle par la forme et la couleur - donnera au final d'une faible fiabilité de 0.59049). Ainsi cette approche ne permet pas d'associer lors de la combinaison de coefficients de fiabilité une valeur, qui se traduira en une détection fiable, même si tous les coefficients ont des valeurs proches de 1.
2. Moyenne de coefficients. Le calcul de la moyenne aura un moindre effet sur la diminution du coefficient final lors de la combinaison de coefficients de fiabilité. L'inconvénient de cette approche est que toutes les fiabilités contribuent aux à égales dans le coefficient final. Par exemple, un coefficient de fiabilité bas sur une étape peut se compenser par les autres, plus élevés. Le coefficient final définira alors une fausse détection en tant qu'un objet d'intérêt.
3. Composition de fonctions de la logique floue. Pour résoudre les deux inconvénients des méthodes, basées sur la probabilité conditionnelle et la moyenne, la solution proposée applique une approche fonctionnelle qui se base sur les méthodes de la logique floue. Cette approche consiste à associer les coefficients de fiabilité, interprétés sous forme de fonctions d'appartenance à des ensembles flous. La méthode d'inférence de la logique floue sert à combiner les conditions d'appartenance à plusieurs ensembles flous dans une fonction complexe. Le procédé suivant - la défuzzification – permet ainsi d'en déduire une moyenne « pondérée » par chaque fonction, représentante le coefficient de fiabilité et par l'importance

d'un coefficient dans la séquence de traitements. C'est le mécanisme d'inférence de Mamdani, souvent appliqué dans la logique floue.

La solution se base sur la notion de fiabilité (de détection et de traitements), qui se projette logiquement dans la notion de certitude. En termes de logique floue la notion de fiabilité se traduit dans la notion de degré d'appartenance aux ensembles flous. Cela permet alors d'utiliser les coefficients de fiabilité comme des degrés d'appartenance d'objets détectés aux ensembles flous – de détections à chaque phase de la solution et en fonction de niveaux hiérarchiques (voir la figure 4.16). Les techniques de la logique floue permettent de manipuler des éléments d'ensembles flous, i.e. des éléments « incertains ». C'est le cas de détection d'objets au moyen de coefficients de fiabilité qui explique le choix de méthodes de la logique floue pour la combinaison de coefficients de fiabilité.

La section 4.6.3.1 suivante décrit en détails le principe d'application du mécanisme d'inférence de Mamdani [43] pour obtenir le coefficient de fiabilité final, qui représentera un objet lors du dénombrement.

#### 4.6.3.1 Application du mécanisme d'inférence de Mamdani

L'interprétation de coefficients de fiabilité sous forme de fonctions d'appartenance de la logique floue permet d'appliquer le mécanisme d'inférence de Mamdani, couramment utilisé dans les applications de prise de décisions et dans l'intelligence artificielle. Son déroulement et son association avec l'application des coefficients de fiabilité sont schématiquement expliqués dans la figure 4.45.

Le mécanisme d'inférence de Mamdani débute par une variable nette (une valeur d'un élément). Dans le contexte de la détection d'objets dans les vidéos avec les coefficients de fiabilité cette variable nette est une caractéristique, associée à une étape de la solution (e.g. la variance d'intensité lumineuse de couleurs dans la première phase d'évaluation à priori de la fiabilité de détections, l'intensité d'un pixel dans les contours etc...). Le mécanisme d'inférence se déroule ensuite en plusieurs étapes :

1. **Fuzzification** : En termes de logique floue cette étape consiste en une composition d'un ensemble de variables floues. Une variable linguistique (floue) est ainsi associée à une variable nette (arrivée en entrée du mécanisme), en appliquant une fonction d'appartenance à un ensemble flou. Dans la méthode proposée, il s'agit d'une étape d'estimation du coefficient de fiabilité depuis une caractéristique et une valeur de référence avec une fonction choisie (e.g. l'intensité de contraste de contours détectés au niveau de contraste minimal, exprimé en niveaux de gris en utilisant la fonction sigmoïde).
2. **Moteur d'inférence (implication, ou activation de règles « if... then... » de la logique floue)** : En termes de logique floue, sur cette étape une variable floue linguistique est associée à une variable floue. En fonction de règles, admises pour un cas d'application, il est possible d'estimer pour une variable ses degrés d'appartenance à plusieurs ensembles flous. Dans la solution cette étape consiste en ajustement du coefficient de fiabilité estimé en fonction de fiabilité de détections à l'instant précédent. Cela est fait par la fonction de la logique floue : le trapézoïde. Elle représente l'évolution d'un coefficient de fiabilité de l'instant  $t-1$  à l'instant  $t$  et construite par les paramètres suivants :
  - Le sommet de la fonction représente la moyenne de fiabilité des détections à l'instant  $t-1$ , c'est-à-dire la valeur du coefficient à partir laquelle les détections à l'instant  $t$  seront fiables (leurs coefficients estimés à l'instant  $t$  seront alors mis à 1).
  - La pente du côté gauche de la fonction est définie par la différence entre la moyenne et la déviation standard à l'instant  $t-1$ . Elle permet d'attribuer une valeur proche de 1 aux détections avec des coefficients proches de la moyenne. Ainsi il est possible de prendre en considération des variations de fiabilités de bonnes détections et l'évolution d'une modèle. La moyenne et la déviation standard sont

recalculées pour chaque frame en permettant l'adaptation d'estimation de coefficients de fiabilité au cours du traitement d'une vidéo. La valeur minimale acceptée pour une potentielle détection d'un objet d'intérêt (la différence  $C_{\text{mean}} - C_{\text{std}}$ ) est limité à 0.5, qui définit, que si un coefficient de fiabilité rapproche une différence de 50% avec une caractéristique d'une bonne détection, la détection, à laquelle il est associé, n'est pas fiable. Graphiquement ce processus est illustré sur la figure 4.45.

3. **Agrégation d'ensembles flous** : Sur cette étape les fonctions, qui ont été associées aux coefficients de fiabilité après les ajustements lors de l'application du moteur d'inférence sont réunies dans un ensemble. Dans les termes de la logique floue il s'agit alors de la génération d'une union de fonctions en une fonction complexe (voir les figures 4.45 et 4.48 pour des exemples de fonctions composées), et dans le contexte de la solution - d'une combinaison de coefficients, associés à un objet, afin d'en déduire une valeur de coefficient final.
4. **Défuzzification** : Par cette étape appliquée dans les termes de la logique floue, on obtient une variable nette depuis une variable floue. C'est-à-dire, que la défuzzification définit depuis une degré d'appartenance, à quel ensemble net elle se réfère. Ces ensembles finaux peuvent être différents d'ensembles pour les variables nettes à l'entrée du mécanisme. Dans le contexte de la solution cette étape est différente de cette définition classique. Dans la méthode proposée ici, l'étape de défuzzification n'attribue pas les objets détectés dans deux ensembles – de bonnes et de fausses détections, mais définit la certitude (un coefficient de fiabilité), qu'un objet détecté est un objet d'intérêt, i.e. une bonne détection. C'est notamment ce principe, qui permet d'appliquer par la suite simple addition de coefficients de fiabilité dans chaque frame. L'avantage d'utilisation de la méthode de défuzzification, par rapport à celle de la probabilité conditionnelle (voir la section 4.5.4.2) est la possibilité de cumuler les coefficients de fiabilité proportionnellement à la fiabilité de détection sur chaque étape. La défuzzification permet alors de garder les valeurs de coefficients qui signifient des bonnes fiabilités (proches à 1) et ne pas les baisser, comme dans le cas du produit (la probabilité conditionnelle). Pendant le processus de défuzzification la valeur finale peut être calculé par deux méthodes :
  - la méthode du centre de masse (Figure 4.47). Cette méthode reflète la degré d'appartenance à plusieurs ensembles flous d'une manière proportionnelle à leurs superficies, mais elle est assez complexe et coûteux en temps de calcul
  - la méthode de moyenne de maximums. C'est une alternative à la méthode du centre de masses, qui donne des résultats similaires, mais nécessite moins de temps de calcul. Le coefficient final est estimé alors par cette méthode, illustrée sur la figure 4.48

Nous utilisons ici la méthode de la moyenne pondérée. Elle estime une « moyenne » pour chaque fonction trapézoïde représentant un coefficient de fiabilité dans leur union (la figure 4.48). La somme de ces moyennes donne le coefficient final par la formule suivante :

$$C_{\text{final}} = \frac{\sum_i^n (\mu_2 - \mu_1)_i}{n}$$

où  $n$  est le nombre de fonctions, qui représentent les coefficients de fiabilité,  $\mu_1$   $\mu_2$  sont leurs deux valeurs maximales (voir la figure 4.48 pour la présentation graphique).



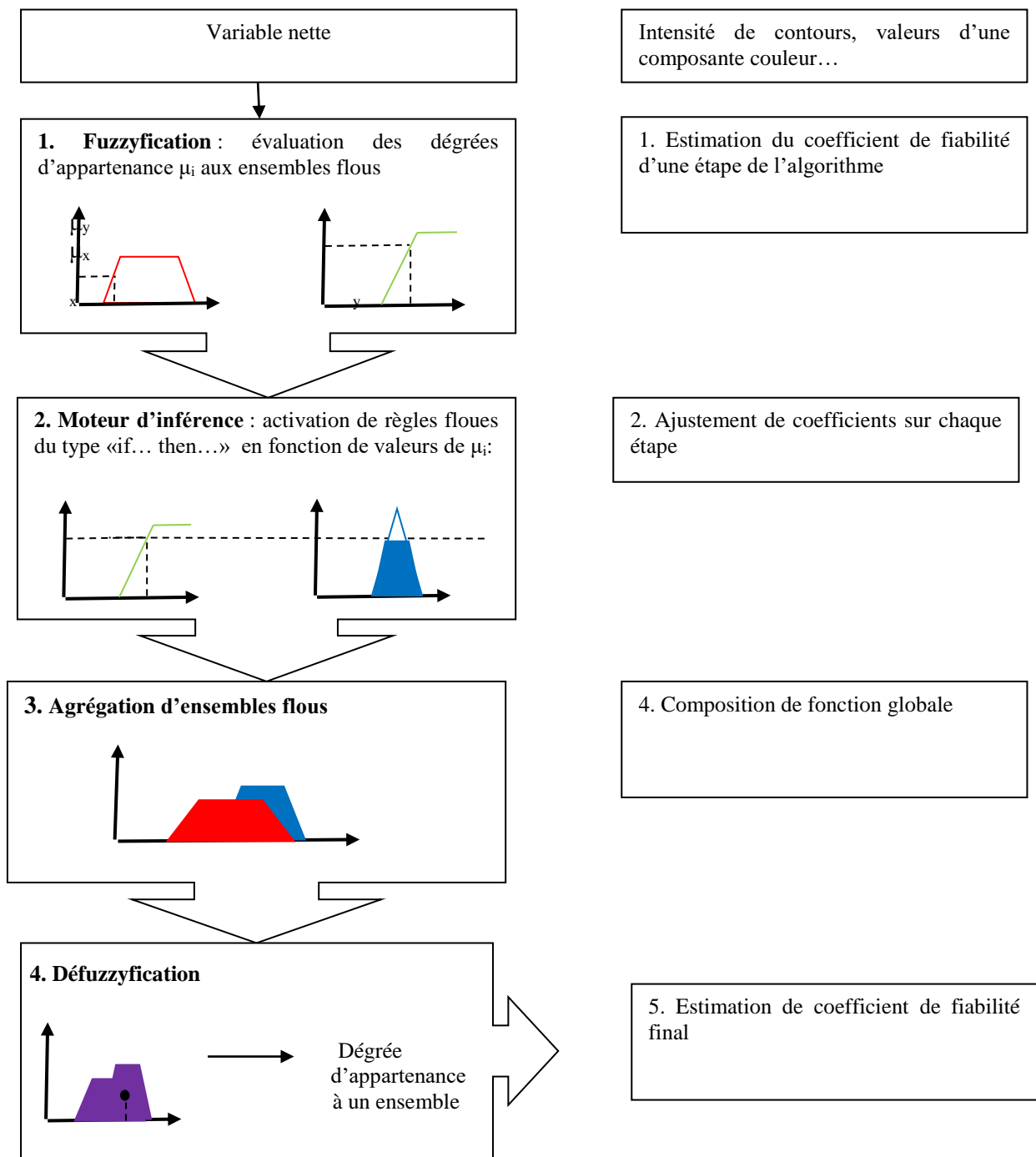


Figure 4.45 : Association de l'inférence de Mamdani avec les phases de la solution implémentée

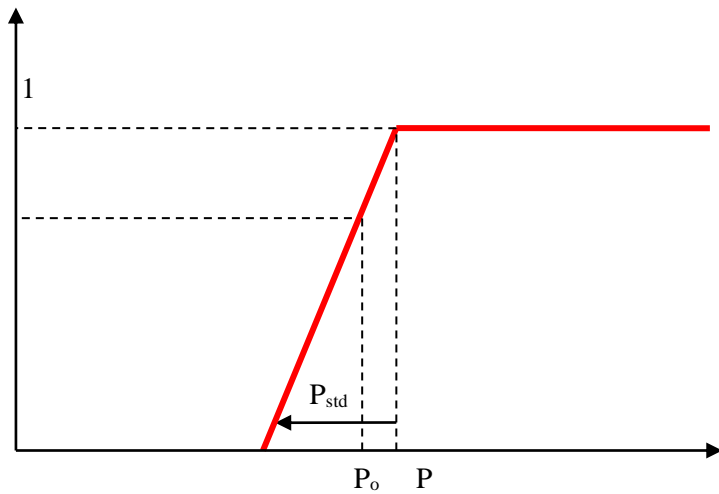


Figure 4.45 : Principe d'estimation de coefficient de fiabilité sur une étape depuis son paramètre  $P_0$ , celui défini pour des potentiels objets d'intérêt et sa déviation acceptée.

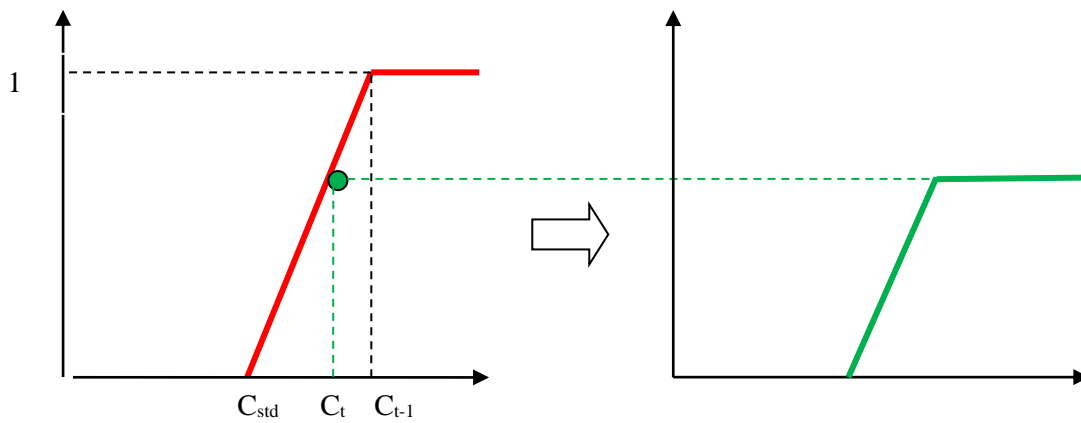


Figure 4.46: Illustration de la méthode d'implication pour générer la fonction d'un coefficient de fiabilité  $C_t$  depuis la moyenne et la déviation de coefficients de fiabilité, estimés à l'instant  $C_{t-1}$  pour l'union en une ensemble de fonctions.

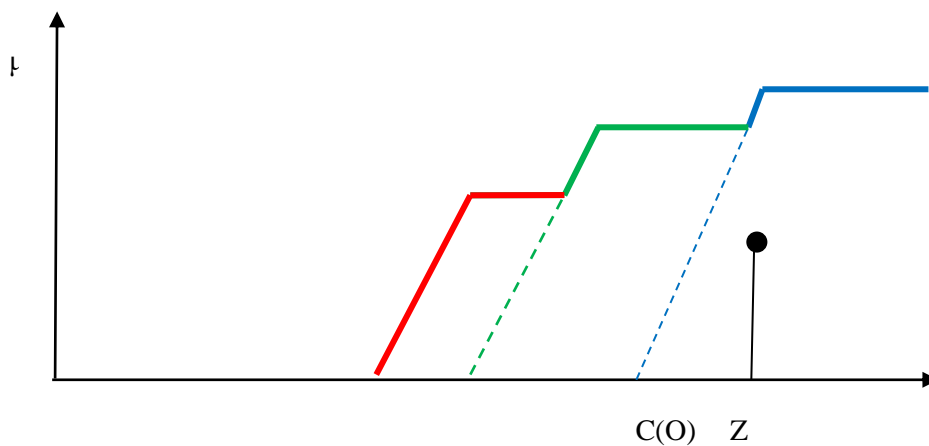


Figure 4.47 Illustration d'estimation du coefficient de fiabilité final par la méthode du centre des masses

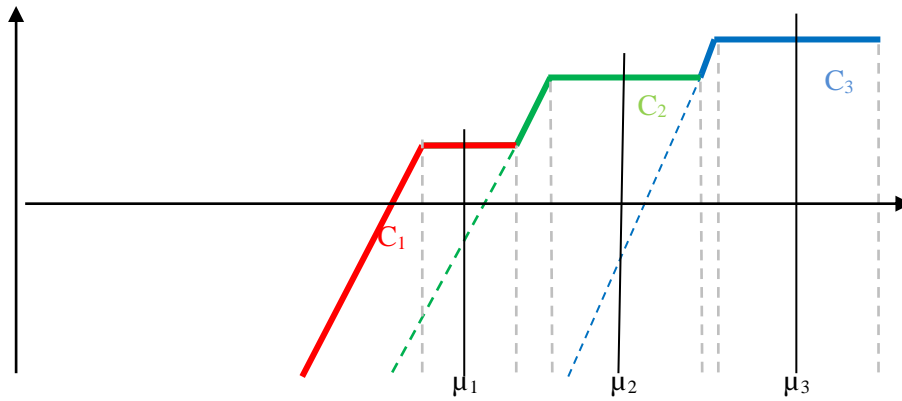


Figure 4.48 Illustration d'estimation du coefficient de fiabilité final par la méthode de la moyenne de maximums

#### 4.6.3.2 Dénombrement d'objets détectés

Après la phase de d'estimation de coefficients de fiabilité le comptage consiste en sommation de coefficients de fiabilité de tous les objets détectés  $C(O_{i,t})$  pour chaque instant  $t$ . Le résultat pour une vidéo sera alors estimé comme la moyenne de ces détections pondérées  $Nb_t$  sur toute la séquence de  $T$  frames :

$$Nb_t = \sum_{O_{i,t}} C(O_{i,t}) Nb = \frac{1}{T} \sum_T Nb_t$$

## 5. Expérimentations et résultats

Pour les expérimentations différentes vidéos, contenant des flux migratoires de palombes ont été fournies. Les vidéos étaient réparties en deux corpus de complexités différentes (section 3.3), et les expérimentations se sont séparées logiquement en deux parties :

- *Première partie* : application de la solution proposée sur les vidéos du corpus 1. Dans les vidéos de ce corpus les objets d'intérêt correspondent à la définition « objets de petits dimensions », donnée dans la section 2.4.1 (objets d'une taille suffisamment grande pour l'approximation de la forme apparente à une primitive géométrique)
- *Seconde partie* : application de la solution proposée sur les vidéos du corpus 2. Dans les vidéos de ce corpus les objets d'intérêt correspondent à la définition « objets de très petits dimensions », donnée dans la section 2.4.1 (les objets ont une petite taille, qui rend impossible l'approximation de la forme apparente à une primitive géométrique d'ellipse et / ou les objets ont un faible contraste sur le fond d'une zone d'intérêt)

L'ensemble de vidéos de la première partie d'expérimentations a été fait avec des outils mobiles, portés dans la main, et pas toujours depuis des postes d'observations dédiés. Leur caractéristique commune était la netteté d'oiseaux sur le fond et une taille suffisamment grande pour modéliser leurs formes apparentes en tant qu'une primitive géométrique. L'analyse de résultats d'exécution sur cette première partie d'expérimentations a montré, que certains cas nécessitent un modèle de « très petit objet » (« objets de très petits dimensions »), basée sur la taille de très petits objets, où l'estimation de longueurs de semi-axes de la plus petite ellipse englobante était compliquée.

Les vidéos, utilisées pour la deuxième partie d'expérimentations était enregistrées lors de la campagne de migrations de palombes en automne 2016 et 2017 depuis les postes d'observation à 500-750 mètres des vols d'oiseaux. Les distances importantes entre les oiseaux et la caméra ont significativement réduit leur taille et leur contraste apparents dans ces vidéos, ce qui a limité les moyens d'analyse des formes apparentes de ces objets et a motivé l'introduction du modèle de « très petit objets » avec des méthodes d'analyse appropriées.

### 5.1 Résultats expérimentaux sur les vidéos du corpus 1

Cette section présente les résultats obtenus sur des vidéos du corpus 1. Les objets d'intérêt dans ces vidéos correspondent à la notion de « petits objets », dont la forme apparente peut être approximée par la plus petite ellipse englobante, avec des longueurs de semi-axes suffisantes pour distinguer les objets d'intérêt de la plupart de fausses détections. Ces objets ont été aussi suffisamment contrastés sur le fond d'image pour générer des variances d'intensités lumineuses supérieures à celles dans les zones avec des légères différences entre les couleurs de la même teinte de ces pixels, mais visuellement homogènes.

La figure 5.1 illustre un exemple de détections dans une vidéo du corpus 1 avec ces caractéristiques. Les régions dans les ellipses vertes correspondent aux détections fiables - les objets d'intérêt qui ont des coefficients de fiabilité proches à 1. Les ellipses bleues désignent les objets d'intérêt avec de coefficients de fiabilité proches de 0.8. Ces objets sont également considérés comme des objets d'intérêt, mais moins conformes au modèle. Dans la première partie d'expérimentations on considère, que les conditions dans les vidéos sont favorables pour avoir des détections fiables, alors que les coefficients de fiabilité proches de 0.8 étaient associés à des détections correctes, et mis à 1 dans le dénombrement final. Les coefficients de fiabilité inférieurs à cette valeur signifient des détections moins fiables et contribuent dans le résultat final par leurs valeurs initiales.

Les points rouges sur le masque d'objets détectés dans la figure 5.1, correspondent au bruit ou aux zones de légères variations d'intensité lumineuse de couleurs similaires. Les coefficients de fiabilité de ces régions sont proches à 0, qui les suppriment implicitement du résultat final.

Les tableaux 5.1 et 5.2 récapitulent les résultats d'application de la solution sur l'ensemble de vidéos du corpus 1, avec le modèle de petit objet d'intérêt. Le coefficient de conformité au modèle d'objet se compose ici sous forme de deux coefficients distincts :  $MR_1$  pour la conformité à la forme du modèle et  $MR_2$  pour la conformité à sa couleur.

Le coefficient  $HSV_w$  est le coefficient, appliqué sur toute la zone et tous les objets qu'elle contient. Il caractérise l'évaluation à priori de fiabilité de détection dans cette zone (i.e. l'intensité du contraste d'objets dans cette zone et la probabilité qu'une zone contient des objets d'intérêt). Le coefficient DR est également appliqué à toute une zone et caractérise l'intensité des contours détectés dans une zone (la probabilité que ces contours appartiennent aux objets d'intérêt). L'application de coefficients de fiabilité séparément et en combinaison pour le comptage des objets détectés dans la première partie d'expérimentation a mené aux conclusions suivantes :

- Le coefficient  $HSV_w$  appliqué seul permet de rapprocher le résultat de dénombrement à la vérité terrain, si les objets d'intérêt sont suffisamment contrastés sur le fond. La variance d'intensité de couleurs dans les zones avec des objets d'intérêt est supérieure de celle dans des zones avec du bruit ou des transitions des couleurs. Cela permet de supprimer des zones d'image sans perdre les objets d'intérêt
- Le coefficient DR appliqué seul est moins efficace pour la sélection d'objets d'intérêt, car il ne fait pas de distinction entre les zones contenant des objets d'intérêt ou du bruit. I.e. une région ou un contour à un faible contraste plus faible d'un contour peut correspondre à un objet d'intérêt, à une variation d'intensité des contours au sein d'un même objet ou des limites entre deux couleurs.
- Les coefficients de conformité au modèle d'objets appliqués seuls permettent de les sélectionner correctement, si lors de la deuxième phase les régions continues générées depuis les contours détectés, correspondent aux objets d'intérêt dans des images initiales sur toutes les étapes de traitement (i.e. les coefficients  $HSV_w$  et DR étaient élevés grâce à un bon contraste d'objets d'intérêt sur le fond. Leur remplacement par 1 pour séparer les coefficients de fiabilité ne changera pas la classification d'une région en tant que détection fiable)
- Une combinaison consécutive de coefficients de fiabilité diminue significativement le nombre de fausses détections, mais aussi le nombre d'objets d'intérêt sélectionnés, si leurs valeurs de coefficients de fiabilité ne sont pas ajustées aux fiabilités des étapes précédentes.
- Dans la combinaison de tous les coefficients  $MR_1$  et  $MR_2$  ont des poids les plus importants, car un objet sera exclu de l'ensemble de bonnes détections, s'il ne correspond pas au modèle par sa forme ou par sa couleur. Les deux dernières colonnes du tableau 5.1 confirment, que les ensembles d'objets sélectionnés par un de ces deux coefficients sont proches, mais la combinaison de deux permet de diminuer davantage le nombre de fausses détections.

Les vidéos du corpus 2, utilisées sur la deuxième partie des expérimentations avaient une complexité supérieure pour les raisons suivantes :

- les objets d'intérêt avaient des tailles plus petites et des contrastes plus faibles à cause d'une plus grande distance de la caméra.
- l'effet de ces caractéristiques s'est exprimé en valeurs plus faibles de coefficients de fiabilité sur toutes les phases de la solution proposée
- une complexité supplémentaire dans certaines de ces vidéos était la scène de l'arrière-plan avec des objets fixes, mais mouvants avec les tremblements de la caméra, qui avaient des paramètres proches aux paramètres du modèle d'objet d'intérêt

Les résultats présentés dans les tableaux 5.1 et 5.2 montrent qu'avec un niveau de complexité des vidéos croissant le nombre de zones d'intérêt à priori « fiables » diminue, parce que les couleurs d'objets d'intérêt et du fond deviennent similaires avec un agrandissement de la distance entre la caméra et les objets. Sans prendre en considération ce phénomène et sans pondérer les zones d'image, des zones avec des objets d'intérêt obtiennent les coefficients de fiabilité égaux à 0 et ne sont pas analysées dans la suite de séquence de la solution proposée. C'est une des raisons, qui explique un faible nombre de bonnes détections par rapport à la vérité terrain, si l'approche conçue pour le modèle du « petit objet » de la première partie d'expérimentations est appliquée sur l'ensemble des vidéos de corpus 2.

Pour récupérer un certain nombre d'objets d'intérêt il est possible de diminuer le seuil minimal de variance d'intensité lumineuse pour sélectionner plus de zones d'intérêt. Mais avec la sélection de plus d'objets d'intérêt, cette approche implique aussi un taux croissant de fausses détections.

Le traitement des vidéos du corpus 2 a mis en évidence la nécessité d'interprétation de coefficients de fiabilité en fonction de fiabilité des traitements des étapes précédentes de la solution. La conception d'un modèle de « très petits objets » s'est avérée nécessaire pour résoudre le problème de similarité d'objets d'intérêt avec des fausses détections et le bruit.

Les tableaux 5.1 et 5.2 montrent les résultats de dénombrement avec et sans application de coefficients de fiabilité et les complètent par les résultats d'application de coefficients de conformité au modèle d'objets (séparément et en combinaison).

A noter que sur la première phase d'expérimentations, la méthode d'inférence de Mamdani n'était pas appliquée pour la composition de coefficients de fiabilité. Les coefficients de fiabilité, obtenus sur chaque phase de la solution ont été combinés dans un coefficient final sous forme de moyenne. Avec les caractéristiques de vidéos plus simples, la moyenne de coefficients permet d'avoir un coefficient de fiabilité correcte pour les objets d'intérêt et les fausses détections.

La suite des recherches était alors orientée vers la conception d'une méthode pour traiter des vidéos d'une complexité supérieure, où la fiabilité de classification d'objets détectés en tant qu'un objet d'intérêt ou une fausse détection dépend de l'interprétation de fiabilité de chaque étape de la solution proposée.

Tableau 5.1. Application de coefficients de fiabilité séparément, e.g. la colonne HSV<sub>w</sub> signifie qu'uniquement le coefficient HR est appliqué et les autres ont la valeur fixe de 1.

Corpus	Vérité terrain	Détections avec un de coefficient coefficients								
		sans C(O)	HSV <sub>w</sub>		DR		MR <sub>1</sub>		MR <sub>2</sub>	
			TW	FW	TW	FW	TW	FW	TW	FW
1	68	559	65.8	261.2	59.2	130.7	44.2	294.6	48.3	260.5
2	40	816	32	617.6	16	326.4	12.5	346.8	8	285.6

Tableau 5.2. Application de combinaisons de coefficients de fiabilité. La colonne HSV<sub>w</sub>+DR signifie que deux coefficients DR et HSV<sub>w</sub> sont utilisés et les autres ont les valeurs fixes de 1.

Corpus	Vérité terrain	Détections avec des combinaisons de coefficients						
		sans C(O)	HSV <sub>w</sub> + DR		HSV <sub>w</sub> + DR + MR <sub>1</sub>		HSV <sub>w</sub> + DR + MR <sub>1</sub> + MR <sub>2</sub>	
			TW	FW	TW	FW	TW	FW
1	68	559	57.4	54.8	54.8	13.3	52.5	11.6
2	40	816	20.8	173.6	18.7	52.8	15.7	15.6

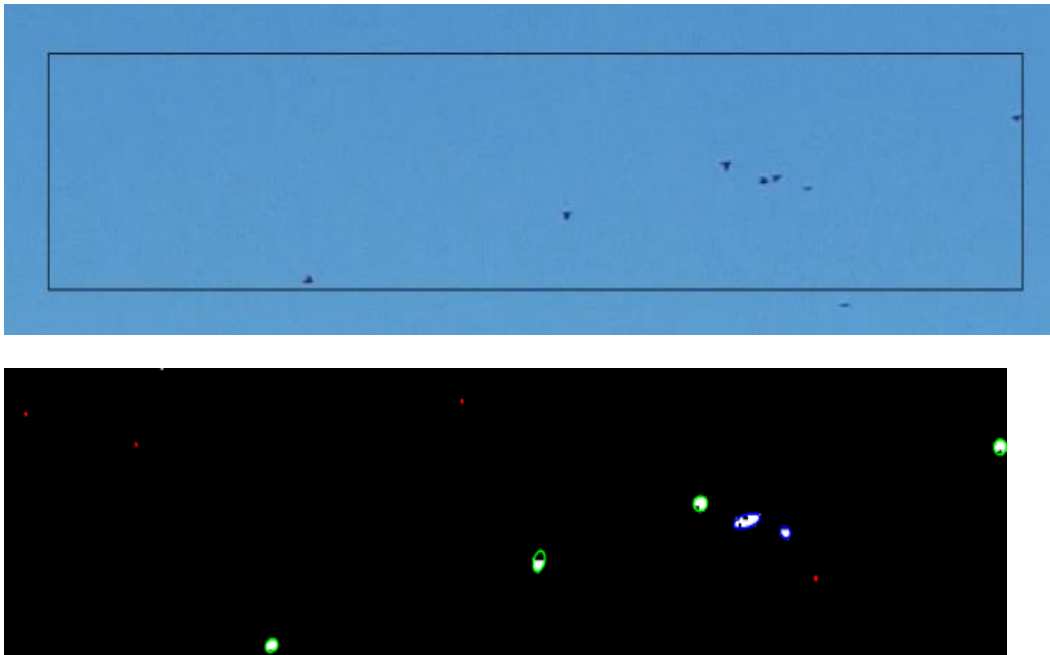


Figure 5.1 : Exemple idéal d'application de la solution sur une vidéo du corpus 1. Les détections considérées comme correctes sont entourées d'ellipses vertes et bleues (les objets, pour lesquels les coefficients de fiabilité sont compris dans l'intervalle  $[0.8, 1]$ ). Les fausses détections, qui n'entrent pas dans le résultat final est le bruit marqué par les points rouges (ces objets obtiennent des coefficients de fiabilité proches du 0).

## 5.2. Analyse des étapes hiérarchiques de la solution proposée

Cette section présente une analyse des étapes de la solution proposée, dans le contexte des entités hiérarchiques (zone – contours – région, voir le schéma 4.16), analysées sur ces phases. Ces entités hiérarchiques sont caractérisées par des paramètres numériques, utilisées par les méthodes du traitement d'images, et dépendent ainsi des conditions propres à une vidéo.

L'objectif de la solution est de proposer une approche générale de détection et comptage de petits objets dans les vidéos, mais en raison de limites numériques et de la qualité de données (ici, des vidéos) elle ne peut pas être universelle. La présente section est alors destinée pour une analyse des limites de la solution.

L'analyse est présentée dans le tableau comparatif 5.3, qui associe à chaque étape les caractéristiques d'une entité analysée sur une étape, les paramètres, appliqués pour l'estimation de coefficient de fiabilité associés ainsi que les difficultés rencontrées et les exemples de vidéos, où elles apparaissent.

La première phase de sélection de zones d'intérêt par la variance d'intensité lumineuse de couleur de leurs pixels. Le paramètre associé est le seuil de la variance à partir duquel une zone d'image est considérée comme une potentielle zone d'intérêt. Il dépend

- du type du fond (une ou plusieurs couleurs avec des transmissions entre elles, comme dans le cas d'une couleur bleu et blanche d'une voile nuageuse sur le ciel bleu ou du soleil brillant)
- de la qualité d'une vidéo - homogénéité de la couleur uniforme, où des objets d'intérêt peuvent être détectés
- du contraste d'objets d'intérêt sur le fond de la zone.

Les difficultés sur cette phase sont issues principalement de la distance entre les objets et la caméra, qui a un effet sur la taille d'objets et sur leur contraste sur le fond. Si les objets sont petits, les

couleurs de leurs pixels ne contribuent pas suffisamment à la variance d'intensité lumineuse de couleurs dans la zone qui les contient, en raison de leur faible nombre par rapport au nombre de pixels dans une zone. S'ils ne sont pas assez contrastés sur le fond d'image, la valeur de la variance d'intensités lumineuses sera également inférieure au seuil estimé. D'autre part, cette variance peut aussi être générée par les variations des couleurs sur le fond lui-même (en pratique dans une vidéo réelle le fond n'est pas idéalement homogène, mais composé de couleurs avec des teintes similaires, qui peuvent générer des faux contours d'objets).

Les difficultés sur la deuxième phase de sélection des contours d'objets et de leur conversion en régions continues consistaient en une définition d'une valeur de niveau de gris définissant une intensité de contrastes de contours pour les détections fiables d'objets d'intérêt. L'autre difficulté a été liée à leurs formes non-régulières (e.g. des contours non-fermés), qui compliquait leur conversion en régions.

L'efficacité de cette étape est liée avec l'étape précédente d'évaluation de détection dans les zones d'intérêt, car son paramètre de seuil de variation d'intensité de couleurs sert à évaluer la fiabilité de détections dans les zones. Pour une vidéo avec des objets difficilement visibles la valeur de ce paramètre doit être basse pour associer à ces zones un coefficient de fiabilité différent du 0. En même temps une valeur basse de ce seuil entraîne l'estimation de valeurs de coefficients élevées pour des zones, où cette variation est générée par d'autres phénomènes que la présence d'objets d'intérêt dans une zone. Dans ces zones il y a un risque de considérer comme des contours d'objets d'intérêt les faibles variations entre les couleurs de pixels du fond. Ces faibles variations seront à l'origine de sélection de faux contours d'objets et, comme conséquence, des fausses détections peuvent apparaître dans l'ensemble d'objets détectés.

Sur la troisième phase de la solution, où les objets détectés sont comparés avec le modèle théorique d'objet d'intérêt, la difficulté principale consistait en une construction du modèle, le plus proche des objets d'intérêt. L'exécution de la troisième phase est alors liée à l'exécution des phases précédentes et dépend des contours détectés et des régions reconstruites. Il y a plus de chances de construire un modèle qui est proche de l'apparence d'objets d'intérêt, si les objets dans l'ensemble, qui a servi pour l'estimation des paramètres du modèle sont similaires et ont des bonnes fiabilités de détections (i.e. leurs contours sont nets et les formes sont relativement régulières). En pratique, et surtout dans les vidéos du corpus 2, c'est n'est pas toujours le cas, et le modèle peut être légèrement différente de l'apparence réelle d'objets d'intérêt.

L'analyse d'exécution de l'algorithme sur des vidéos de deux corpus a mené aux conclusions suivantes :

- Les facteurs définitifs pour l'exécution des phases de la solution proposée et pour le résultat final du comptage sont la taille et le contraste d'objets dans une vidéo, liés à la distance entre la caméra et les objets observés.
- Le critère de la taille d'objets en pixels, a besoin d'une limite d'une taille minimale (dont l'estimation est présentée dans la section 4.3.5), car avec l'augmentation de la distance entre les objets et la caméra, les objets d'intérêt réduisent en taille et deviennent similaires au bruit numérique. Ce phénomène peut introduire des détections irrégulières du bruit numérique, similaire aux objets d'intérêt par l'aspect spatial. Dans le cas d'un grand groupe de très petits objets d'intérêt, des fausses détections (le bruit) peuvent être proches aux objets d'intérêt et entrer ainsi dans les groupes d'objets d'intérêt, suivis par l'aspect temporel. Cette possible confusion entre les objets d'intérêt et les fausses détections, dû aux caractéristiques d'une vidéo peut introduire des erreurs dans l'estimation de paramètres du modèle d'objet d'intérêt.
- Le phénomène de chevauchement de très petits objets dans une scène réelle et leur position proche aux autres objets d'intérêt, résulte en leur fusion sur une image. Dans ce cas il devient compliqué de définir, combien d'objets correspondent à une région sur un masque binaire de détections et d'estimer les paramètres du modèle.



Ces points expliquent la différence observée entre la vérité terrain et les résultats d'exécution de l'algorithme.

Tableau 5.3 Analyse des étapes de la solution et de difficultés du traitement

Phase de la solution	Caractéristique appliquée	Paramètre associé sur l'étape	Objectif de l'étape	Phénomènes avec le plus d'influence sur l'étape	Difficulté impliquée	Exemples
1ère phase d'évaluation à priori de détection dans les zones	Variance d'intensité lumineuse de pixels	$Y_{\text{texture}}$ - seuil minimal de variation d'intensité lumineuse	Sélection d'une zone avec des potentiels objets d'intérêt	Uniformité du fond	Transition entre couleurs fait augmenter la variation d'intensité, même si la zone est sans objets d'intérêt	Ciel au coucher du soleil, soleil brillant sur une partie d'image
				Contraste d'objets sur le fond	Un contraste très faible n'est pas suffisant pour une variance d'intensité supérieure au seuil	Des objets très éloignés et petits, qui fusionnent avec le fond
2 <sup>ème</sup> phase de détection de contours	Intensité lumineuse de contours sur le masque de contours standard	$\alpha$ - intensité lumineuse minimale	Sélection des contours les plus contrastés	Fiabilité de sélection d'une zone, variation d'intensité lumineuse localement (pixels voisins)	Confusion de contours de très petits objets à un faible contraste avec des faibles variations des couleurs	Images à une basse qualité, granuleuse : Des zones avec de limites de couleurs ou luminosité
2 <sup>ème</sup> phase de conversion de contours dans des régions contenues	Intensités moyennes locales des patches, centrés sur les contours sélectionnés	$M_{\alpha}$ - moyennes d'intensités lumineuses associées aux voisinages locaux d'un contour	Fermeture de contours de très petits objets sans les agrandir	Intensité de contours, générés à l'étape précédente, fermeture de contours	Intensité de contours variable pour le même objet, localisation proche de contours sur une image	Fusion de contours de très petits objets, proches sur l'image
3 <sup>ème</sup> phase - estimation de conformité au modèle	Degré de similarité entre la forme apparente d'un objet et son modèle théorique	Paramètres de modèle en fonction du type de modèle (petit objets/très petit objet)	Analyse de conformité de détections aux caractéristiques de très petit objet	Fiabilité d'exécution des étapes précédentes et de la modélisation de forme apparente de très petits objets	Résidus des fausses détections (bruit similaire au modèle d'objets, faux contours) dans le résultat final dû au choix des paramètres aux étapes précédentes	Un faible contraste d'objets d'intérêt sur le fond d'image, qu'y génère des mêmes variations d'intensité pour des régions d'objets d'intérêt que pour le fond d'image

### **5.3 Résultats expérimentaux sur les vidéos du corpus 2 et la conception du modèle de « très petits objets »**

Dans la seconde partie d'expérimentations la solution proposée a été appliquée sur les vidéos du corpus 2, d'une complexité supérieure (voir la section 3.3.1 et le tableau 3.2). L'enregistrement de ces vidéos a été fait pendant la campagne de migrations des années 2016 et 2017, en respectant les critères identifiés lors de l'analyse du premier ensemble de vidéos fourni (section 3.3 contient les détails sur cette analyse et le tableau 3.2 – les critères de répartition de vidéos dans les corpus).

A la différence de la première partie d'expérimentations, où les formes apparentes permettaient les approximations à des ellipses, uniquement les tailles d'objets ont été utilisées pour générer le modèle d'objets d'intérêt et évaluer la conformité d'objets détectés à ce modèle. Le contraste d'objets sur le fond dans les vidéos du corpus 2, et ainsi leurs contours, ont été moins nets. En conséquence, leur conversion en régions continues a été moins fiable. Dans ces conditions les coefficients de fiabilité sont particulièrement importants, car ils permettent de conserver les détections de différents niveaux de fiabilité jusqu'à la fin d'exécution, où elles seront caractérisées par leurs coefficients de fiabilité finaux.

L'apport de l'aspect temporel est également plus important pour le traitement de vidéos du corpus 2. Un meilleur contraste d'objets d'intérêt sur le fond dans les vidéos du corpus 1 contribue à des variations d'intensité lumineuse dans les zones d'image considérablement supérieures à celles dans des zones vides, même si elles ont des faibles variations de couleurs au sein de la même teinte. Ce phénomène permet d'associer des coefficients de fiabilité égaux à 0 aux zones avec des faibles variances d'intensité lumineuse de couleur et aux contours peu intenses, sans perdre d'objets d'intérêt. En termes d'aspect temporel, cela signifie que d'un frame à l'autre la plupart des objets détectés sont des objets en mouvement (d'après les caractéristiques du corpus 1, la texture du fond est uniforme, sans autres objets). Dans les vidéos du corpus 2, quant à elles, le fond est partiellement ou entièrement texturée, qui signifie, que des objets de natures différentes peuvent aussi apparaître sur le fond d'image. Le critère de déplacement, implémenté dans l'aspect temporel, devient alors important pour la détection d'objets afin de mettre à 0 le coefficient de fiabilité d'objets, qui n'ont pas de déplacements ou ont des mouvements erratiques.

Le comptage d'objets détectés sans coefficients de fiabilité dans les vidéos de corpus 2 (i.e. un simple dénombrement de toutes les régions détectées par la méthode de composantes connexes) donne l'écart de la vérité terrain d'environ 200% (la comparaison complète de résultats est donnée dans le tableau 5.4). L'application de coefficients de fiabilité permet de considérablement réduire cet écart et d'approcher le résultat de dénombrement à la vérité terrain. En fonction d'une vidéo l'écart de la vérité terrain varie de 35% à 10% pour des vidéos avec quelques centaines à plus d'une mille d'oiseaux. Il est important de rappeler que cette « vérité terrain » est obtenue en comptant manuellement les oiseaux ce qui est en pratique très difficile à réaliser d'une manière exacte. La marge d'erreur de la « vérité terrain » par rapport à la réalité peut facilement atteindre plus de 20%. Les résultats du dénombrement automatique rentrent donc généralement dans cette marge d'erreur. Les résultats d'application de la solution sont donnés dans le tableau 5.4 et commentés dans la section 5.4.

### **5.4 Analyse des résultats d'application de la solution sur les vidéos typiques d'observation de migrations pour leur suivi et dénombrement**

Le tableau 5.4 compare les résultats d'application de la solution sur les vidéos réelles des campagnes de migrations avec les résultats d'application de deux algorithmes basés sur le seuillage pour détecter dans les vidéos (image par image) des régions, associées aux potentiels objets d'intérêt :

1. Algorithme Fiji qui applique un seuillage adaptatif, testé par l'équipe de Sogeti.
2. Un algorithme pour des outils mobiles qui applique un seuil fixe pour sélectionner des pixels de régions représentantes les objets d'intérêt. L'application mobile pour les tests de cette approche a été également développée par Sogeti.

3. Comptage de toutes les régions détectées par la méthode de composantes connexes (résultats sans application des coefficients de fiabilité). C'est-à-dire, que tous les objets détectés ont leurs coefficients de fiabilité mis à 1 (i.e. ils sont considérés comme des objets d'intérêt).

Les résultats de dénombrement par ces trois méthodes représentent la moyenne de détections pour chaque vidéo dans le tableau 5.4. Les figures 5.2 – 5.5 donne une aperçue graphique des résultats dans le tableau 5.4.

Tableau 5.4 : Résultats et expérimentations sur 9 vidéos d'une complexité élevée et leurs explications. Les nombres avec ~ sont estimés par une de méthodes rapides mais approximatives de dénombrement manuel, souvent utilisée par les ornithologues : la segmentation visuelle d'un vol d'oiseaux en blocs avec des nombres d'oiseaux approximativement égaux. Les cases sans nombres signifient un résultat incertain du dénombrement manuel à cause de la qualité de vidéo (pas de résultats de dénombrement manuel pour ces vidéos). La vidéo 129 représente un cas difficile, où les objets ont une forme relativement régulière, mais un contraste moyennement net avec une particularité de disparaître avec des reflets du soleil (l'image extraite est donnée sur la figure 9.9).

Vidéo	Vérité terrain	Résultats de dénombrement par des méthodes différents				
		1. Seuillage adaptatif de l'algorithme Fiji	2. Seuillage par l'application mobile avec un seuil fixe	3. sans coefficients de fiabilité	avec coefficient de fiabilité de l'aspect spatial	avec coefficients de fiabilité de l'aspect spatial et aspect temporel de paramétrage du modèle
135.mp4	~1000	-	-	2130	1563	1175
132.mp4	~1000	-	-	1698	1319	1054
114.mp4	410	198	62	785	577	412
115.mp4	1245	1098	395	2261	1643	1393
113.mp4	650	323	285	1758	1510	872
129.mp4	380	82	106	976	853	622
109.mp4	~1000	-	~500	1892	1537	1030
133.mp4	~1000	-	-	1857	1629	975
137.mp4	~1200	-	-	3375	1259	1054

La comparaison de ces trois approches permet de dégager les éléments suivants :

- L'algorithme Fiji (la méthode 1) et la méthode pour des appareils mobiles avec un seuil fixe (la méthode 2) donnent des résultats variés en fonction des caractéristiques d'une vidéo. Aucune d'elles n'a pas montré des nets avantages l'une par rapport à l'autre. Les possibilités d'application ces méthodes pour la détection et le comptage de petits objets en mouvement dans les vidéos sont limitées, car un seuillage, même adaptatif ne donne pas de moyens d'analyse approfondie.
- La méthode proposée permet une analyse des scènes de vidéos et des potentiels objets d'intérêt sur des niveaux hiérarchiques de traitements (zones d'intérêt, contours des objets d'intérêt et les régions englobées, figure 4.16). Cela explique les résultats de dénombrement sans coefficients de fiabilité pour une partie de vidéos plus proches à la vérité terrain, que ceux, obtenus par les méthodes 1 et 2, basées sur le seuillage. Mais pour une autre partie de vidéos les résultats d'application de la solution sans coefficients de fiabilité sont toujours loin de la vérité terrain (souvent plus élevés, car les fausses détections, générées par exemple par les faux contours ou le bruit numérique obtient les coefficients de fiabilité égaux à 1 et sont considérés comme des objets d'intérêt).
- Les résultats obtenus avec l'application des coefficients de fiabilité montrent, que les traitements adaptatifs prennent en considération les caractéristiques et la complexité d'une vidéo, grâce à l'évaluation des fiabilités des détections sur les étapes de la solution proposée et

leurs interprétations tout au long de son exécution en un coefficient de fiabilité final. Cette méthode donne les résultats les plus proches à la vérité terrain

- La solution proposée permet de détecter et de compter les oiseaux dans des grands vols, où le comptage à l'œil ne peut être qu'approximatif (la figure 9.3 dans l'annexe est un exemple). Ce phénomène s'explique par un faible contraste des très petits objets sur le fond, qui ne repérable à l'œil qu'avec un agrandissement. Numériquement ses différences s'expriment en légères variations des composantes-couleur, qui sont détectées par les paramètres ajustables de phases de la solution. Ces faibles variations sont également caractérisées par les coefficients de fiabilité, qui permettent de les inclure dans le résultat final.
- Les résultats de la solution inférieures à la vérité terrain s'expliquent par un phénomène fréquent dans des vidéos avec des vols nombreux et éloignés de la caméra – le chevauchement des oiseaux. Il se représente comme une superposition des oiseaux d'une ou de l'autre manière :
  - plusieurs oiseaux très proches l'un à l'autre se fusionnent en une région d'une taille plus grande que celle estimée pour le modèle. Cette région contient plusieurs objets d'intérêt, mais dans le résultat final ils sont caractérisés par un seul coefficient de fiabilité inférieur à 1 (ou considérés comme non-conformes au modèle d'objet d'intérêt avec le coefficient égal à 0).
  - certains oiseaux sont cachés les uns derrière les autres, car une vidéo est une projection en 2D d'une scène en 3D. Dans ce cas la région dans une vidéo, associée à cette superposition sera compté comme un objet d'intérêt avec son coefficient de fiabilité associé, i.e. au moins un objet ne sera pas compris dans le résultat final de dénombrement.

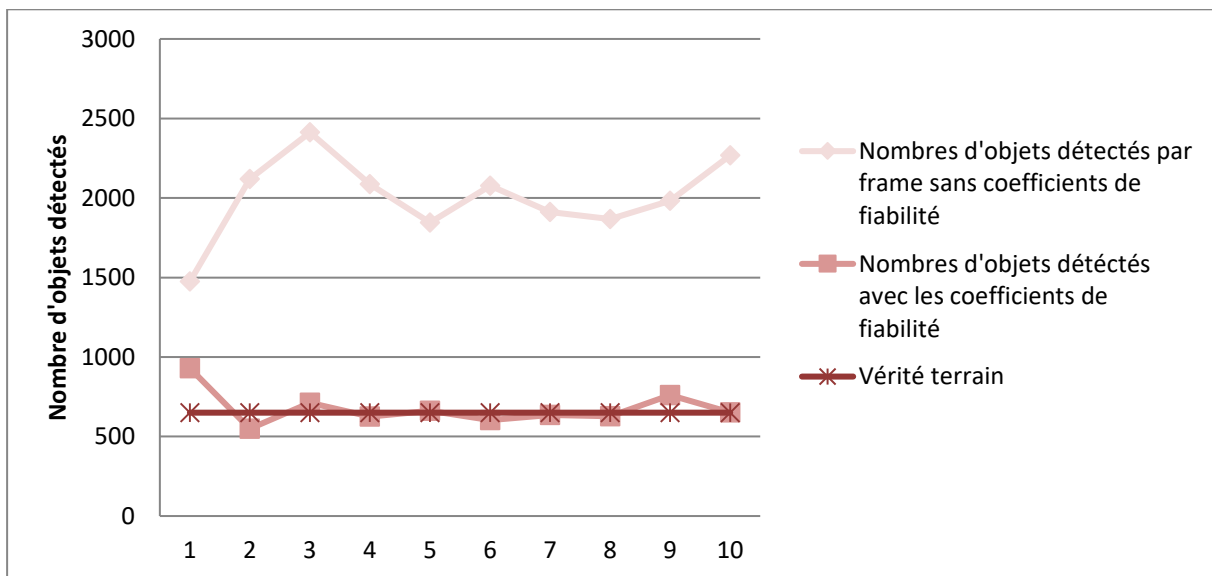


Figure 5.2 : Illustration graphique de résultats de détections par la solution proposée dans la vidéo 113, sur la base de 10 frames. Les nombres de détections par frame sans application de coefficients de fiabilité (qui consiste en sommation de toutes les objets détectés par les méthodes l'aspect spatial) ne sont pas réguliers et différents de la vérité terrain. Les coefficients de fiabilité, quant à eux, permettent d'approcher les nombres de détection à la vérité terrain.

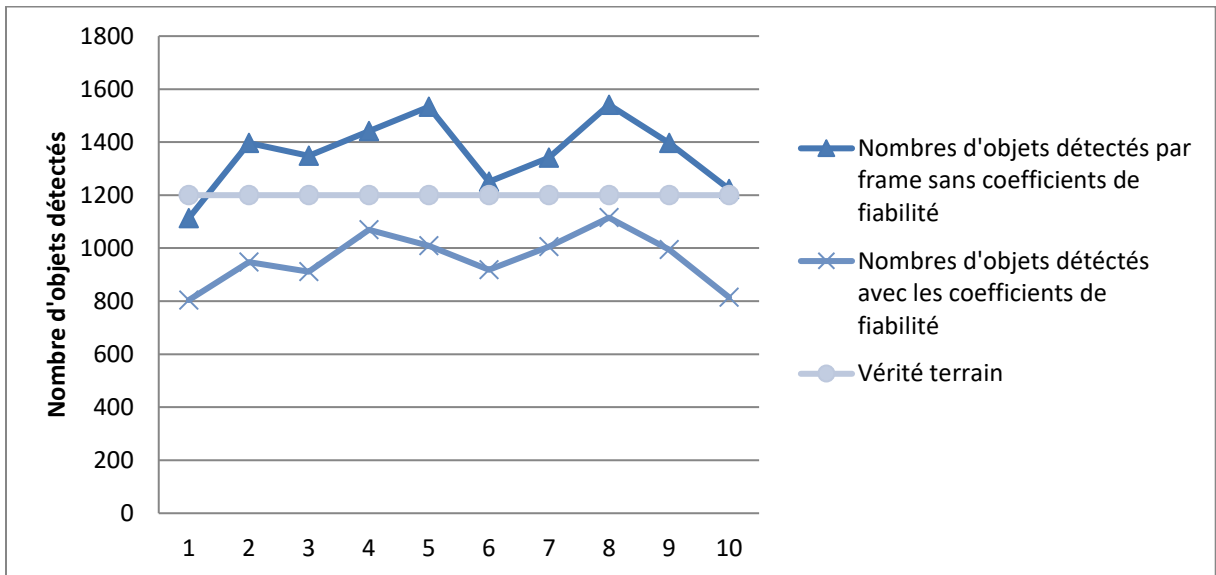


Figure 5.3 : Illustration graphique de résultats de détections par la solution proposée dans la vidéo 137, sur la base de 10 frames. Ici, on observe le même phénomène (les nombres de détections sans coefficients de fiabilité sont irréguliers et très différents de la vérité terrain). Une différence plus importante de la vérité terrain et des nombres de détections avec l'application de coefficients de fiabilité s'explique par une complexité supérieure de la vidéo traitée (les objets d'intérêt plus petits et la présence du phénomène de chevauchement).

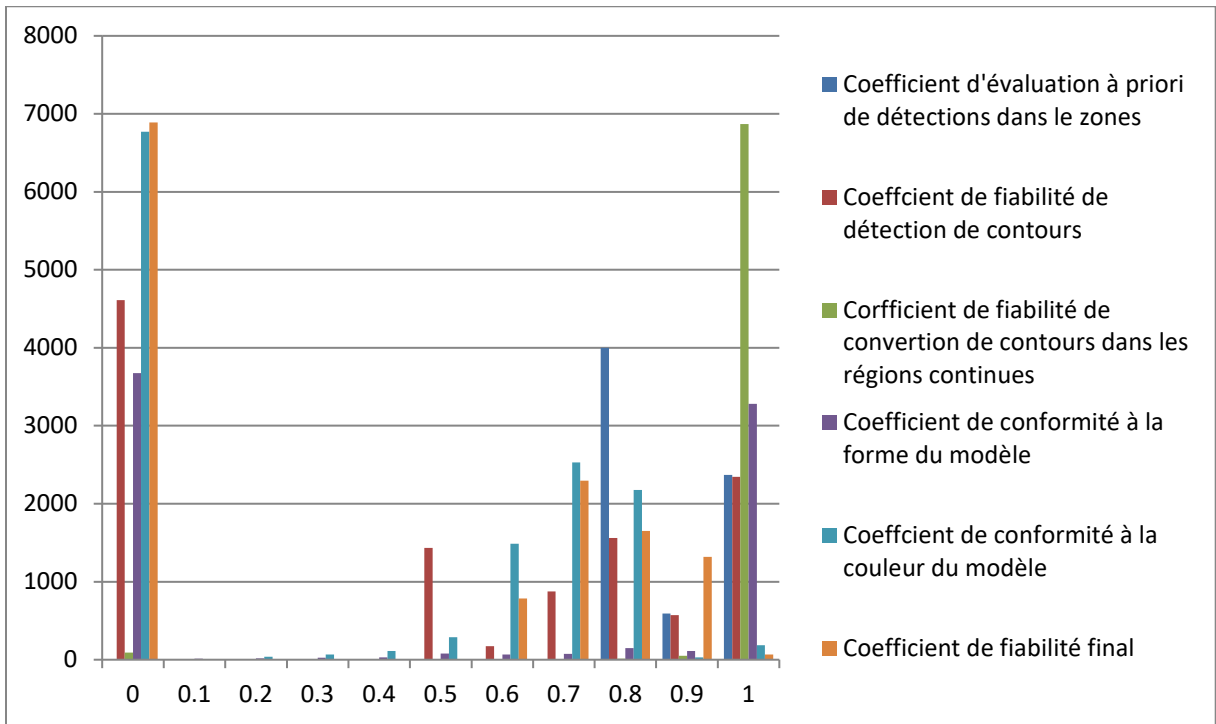


Figure 5.4: Histogramme de coefficients, attribués aux objets détectés dans la vidéo 137 sur la base de 5 frames

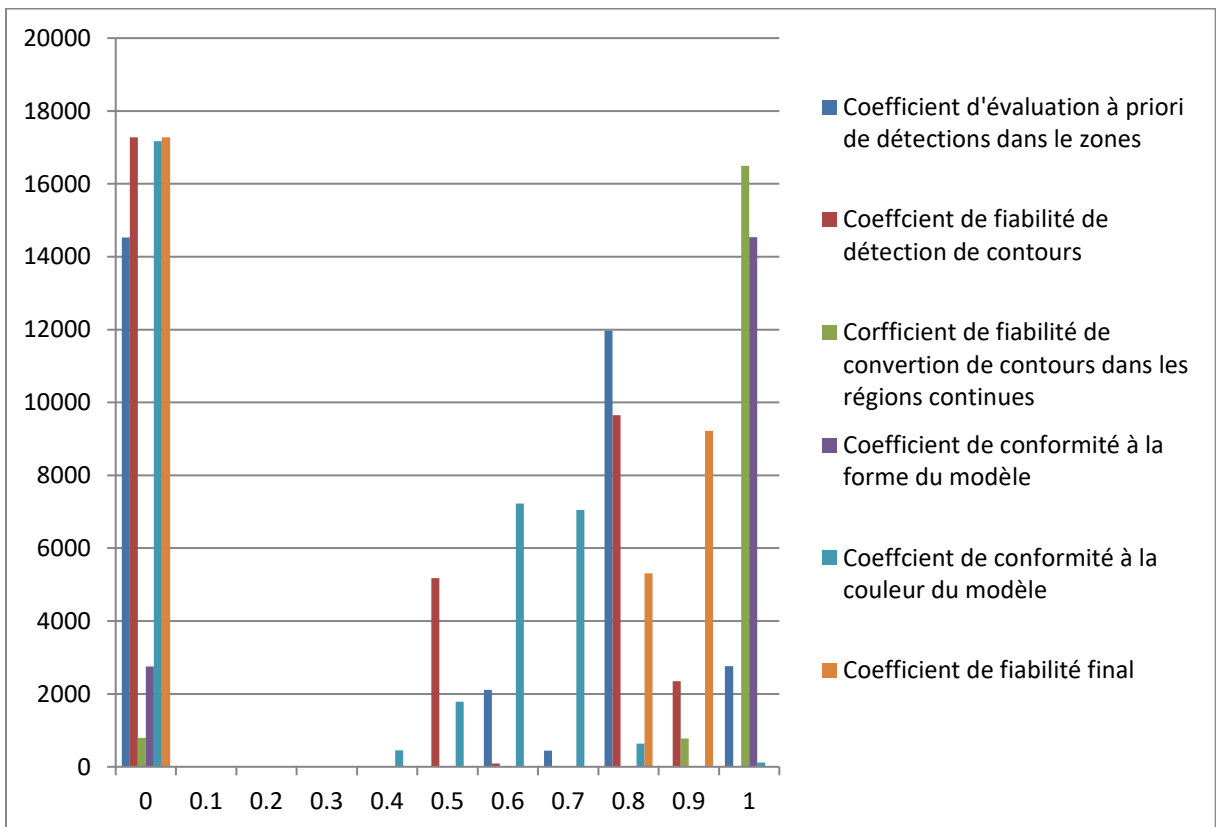


Figure 5.5 : Histogramme de coefficients, attribués aux objets détectés dans la vidéo 113 sur la base de 10 frames

## 6. Conclusion

Les travaux des recherches dans le cadre de cette thèse avaient pour le premier objectif d'étudier le problème du dénombrement de petits et très petits objets dans des vidéos et la possibilité de son automatisation. Le deuxième objectif consistait en conception sous forme d'une séquence de traitements de vidéos, qui résulte en un nombre d'oiseaux présents dans un vol enregistré d'une solution répondant aux points suivants :

- faciliter la récolte de données statistiques sur les migrations saisonnières d'oiseaux
- réduire la main-d'œuvre mobilisée lors des campagnes de migrations
- améliorer la précision de dénombrement

L'analyse de vidéos fournies pour le cas d'étude – des vidéos avec des vols d'oiseaux migratoires – et des résultats intermédiaires des étapes de la solution proposée ont mis en évidence les limites d'analyse des scènes dans ce type de vidéos :

- Difficulté de distinguer les très petits objets d'intérêt (les oiseaux dans un vol éloigné) du fond d'image
- Une forte influence de la qualité d'une vidéo sur les résultats de détection d'objets d'intérêt
- Similarité des représentations numériques d'objets d'intérêt et de fausses détections lors de l'exécution des étapes de la solution (e.g. les couleurs ou les formes similaires d'objets d'intérêt et de fausses détections)

L'étude de l'état de l'art des travaux existants a révélé les limites d'applications de méthodes existantes sur les vidéos avec des très petits objets, telles qu'une impossibilité d'associer à un objet les méthodes existantes (e.g. correspondance de formes ou des points-clés). Une nécessité d'introduire une approche différente de méthodes appliquées dans les travaux existants a été constaté.

Une solution pour la détection et le comptage, basée sur la fiabilité des détections d'objets dans les vidéos a été alors proposée. En comparaison avec les méthodes existantes étudiées, le principe hiérarchique de la solution – d'une entité globale (une image) à une entité locale d'un objet – possède des avantages suivants :

- permet une extension sur les applications variées
- permet d'ajuster le traitement d'une vidéo en fonction de sa qualité
- permet une abstraction du type d'objets recherchés au début de traitement, en commençant par l'analyse des zones d'images
- la modélisation d'objet d'intérêt et un intervalle de couleurs associées permet d'adapter le modèle d'objet à des conditions propres à une vidéo ou d'introduire plusieurs modèles différentes pour une vidéo.

Ces caractéristiques de la solution conçue pour la détection est le dénombrement de petits objets en mouvement permettent de répondre au besoin global d'une solution modulaire et extensible pour l'application dans un vaste domaine de détection d'objets en mouvements dans les vidéos. Le cas d'étude pour le suivi et le dénombrement des oiseaux-migrateurs est un exemple de base qui a servi pour la conception et développement du concept de la solution, qui applique sur ses étapes des méthodes de traitement d'images propres à ce cas de référence.



Le développement des techniques associées aux étapes de l'algorithme, propres aux autres applications, telles que la vidéosurveillance du trafic routier, des lieux publics, ou encore la détection des insectes ou de la poussière dans les vidéos est une des axes principaux des travaux ultérieurs sur le sujet abordé dans la thèse.

## 7. Discussion

### 7.1 Contributions des travaux de recherches effectués

La détection d'objets dans les vidéos est un besoin, qui survient souvent dans les applications de vidéosurveillance. Pour cette raison une partie importante du travail a été consacrée à l'étude de l'état de l'art dans le domaine du traitement d'images et à l'analyse la possibilité d'application de méthodes existantes pour le cas particulier de détection de très petits objets en mouvement dans les vidéos. Les types d'objets d'intérêt, pour lesquels les méthodes étudiées dans l'état de l'art sont conçues, sous-entendent des représentations d'objets dans les vidéos, qui permettent de définir de descripteurs d'objets (tels que l'envergure des ailes des oiseaux). Mais en cas d'une grande distance entre la caméra et les objets observés, ils ne peuvent pas être décrits que d'une manière approximative, i.e. les distinguer d'autres objets présents dans une vidéo et du bruit devient difficile, voire impossible.

Les travaux de recherches dans le cadre de la thèse contribuent dans le domaine du traitement d'images dans les points suivants :

- proposition d'une approche innovante pour le dénombrement d'objets dans les vidéos – application de coefficients de fiabilité de détections d'objets en tant que des unités de dénombrement
- conception d'une méthode d'estimation de fiabilité de détections d'objets, basée sur l'analyse de leurs aspect spatial (forme, couleur) et temporel (régularité de déplacement)
- développement d'une séquence hiérarchique de méthodes du traitement d'images, adaptées aux conditions locales d'exécution et de fiabilité de l'exécution de l'étape précédente

Les coefficients de fiabilité dans l'intervalle de 0 à 1 caractérisent une détection implicitement en tant qu'un objet d'intérêt ou une fausse détection pour en garder le premier et en supprimer la deuxième (si son coefficient de fiabilité vaut 0). Ainsi un objet sera ajouté proportionnellement à sa fiabilité d'être un objet d'intérêt dans le résultat de comptage final par une simple addition du coefficient de fiabilité associé sans aucun autre tri supplémentaire (l'addition du coefficient valant zéro, qui correspond à une fausse détection, n'aura pas d'effet sur le résultat final, et l'addition d'une valeur proche à 1 permettra correctement sélectionner un objet d'intérêt). Le résultat de ce dénombrement est interprété ainsi comme une moyenne pondérée d'objets, détectés dans une vidéo, où toutes les étapes de la solution contribuent dans l'estimation du poids, associé à chaque objet. L'application de coefficients de fiabilité permet ainsi d'estimer un nombre approximatif d'objets d'intérêt, si les conditions dans une vidéo ne permettent pas de définir leur nombre exact. Sans application de coefficient de fiabilité les résultats de comptage diffèrent significativement la vérité terrain (tableau 5.4).

Le tableau 5.4 montre aussi, qu'à la différence de la méthode de composantes connexes, la moyenne pondérée de coefficients de fiabilité permet d'obtenir des résultats proches à la vérité terrain. Pour les vidéos d'une complexité supérieure, où le dénombrement par la méthode de composantes connexes résulte en un écart de 200% de la vérité terrain contre 10% dans les meilleurs cas d'application de coefficients de fiabilité. Cette marge d'erreur de 10% est acceptable pour l'analyse statistique de migrations, où il s'agit des grands nombres d'oiseaux migratoires sur plusieurs postes d'observation.

La deuxième contribution du travail dans le domaine du traitement d'images et vidéos et la conception d'une méthode novatrice de l'aspect spatial et temporel pour la détection et dénombrement de petits objets dans les vidéos avec des contraintes sur la forme apparente d'objets, leurs tailles et leurs contrastes sur le fond d'image.

La solution innovante consistant à combiner ces deux aspects permet d'analyser la représentation spatiale de petits objets d'intérêt et son évolution dans le temps dans des conditions, où les détections ne sont pas fiables (e.g. quand des objets de natures différentes obtiennent dans une vidéo des caractéristiques similaires dans le contexte du traitement numérique, comme des formes apparentes similaires). Les phases de la solution développée permettent d'avoir une évaluation graduelle des fiabilités de détection dans le contexte de deux aspects et de méthodes appliquées au moyen de coefficients de fiabilité. Ainsi il est possible d'éviter la suppression d'objets détectés, mais de caractériser par leurs coefficients de fiabilité, à quel point ils correspondent à des objets d'intérêt. Ce concept permet de considérer au final comme des objets d'intérêt, ceux qui ont les caractéristiques attendues pour un objet d'intérêt.

La troisième contribution des travaux des recherches consiste en application de notions de la statistique et de la logique floue comme les bases essentielles d'une solution pour le traitement d'images. Si les méthodes de la statistique ont des applications dans ce domaine, l'utilisation de méthodes de la logique floue n'y est pas répandue. Dans la présente thèse les opérations ensemblistes et les fonctions d'appartenance sont essentielles pour l'application de notions de fiabilité dans la séquence de traitements implémentés.

Le tableau 5.4 prouve l'efficacité d'application de méthodes de la logique floue dans le contexte de détection et dénombrement de petits et très petits objets d'intérêt. Les techniques d'agrégation et de défuzzification du mécanisme d'inférence de Mamdani servent comme des liens entre les domaines et les méthodes variées, utilisés dans la solution. Elles permettent de transmettre la notion de fiabilité entre les phases de la séquence de traitements hiérarchiques, liées entre elles. Ces méthodes permettent d'implémenter et d'appliquer le concept du dénombrement de détections, obtenues avec une certaine degré de certitude et dans des conditions complexes. Son efficacité est présentée par des résultats d'exécution détaillés (avec et sans utilisation de coefficients de fiabilité) permet de le valider pour le cas de suivi de migrations au moyen de la traitement des vidéos. Les points essentiels des contributions des travaux de recherches dans le cadre de cette thèse sont résumés dans la liste suivante :

- *Proposition d'une méthode-novatrice* pour le dénombrement de petits objets en mouvement dans les vidéos : à ce jour il existe peu de travaux sur ce sujet
- *Analyse des limites des méthodes existantes* et des méthodes classiques pour la détection et dénombrement de très petits objets en mouvement dans le contexte de suivi de migrations :
  - o approches basées sur le suivi d'objets ou la reconnaissance de formes demandent une taille suffisante pour l'analyse des objets détectés
  - o approches basées sur le flux optique appliquent des traitements pixel par pixel, ce qui augmentent le temps d'exécution pour l'application de l'algorithme sur des vidéos HD. Mais pour la sélection de petits objets dans des vidéos la plus complète, cette qualité est nécessaire.
- La méthode numérique développée est capable de résoudre le problème de difficulté à repérer des très petits objets visuellement dans une vidéo.
- Une chaîne complète de détection et dénombrement de très petits objets, basée sur un aspect spatio-temporel du modèle de référence pour le dénombrement et la pondération des objets détectés, a été proposée.
- Une *notion innovante de fiabilité de détections* en fonction des conditions dans une vidéo a été introduite et appliquée tout au long d'exécution de l'algorithme pour un dénombrement

implicite des objets détectés (de l'estimation de la fiabilité d'une zone d'intérêt au niveau globale de l'algorithme jusqu'au niveau local de l'estimation de la conformité d'un objet à son modèle).

- Les méthodes de l'aspect spatial proposé ont été développées sur la base des principes de la logique floue et les fonctions d'appartenance aux ensembles flous pour les appliquer en tant que la fiabilité d'une détection
- La structure de la séquence des traitements proposée permet d'interpréter la fiabilité des traitements sur les étapes de la solution en tant que des entités hiérarchiques indépendamment les unes des autres, en permettant des extensions de l'algorithme aux applications variées
- L'application de l'approche temporelle permet d'affiner la sélection d'un ensemble d'objets pour la construction du modèle de référence d'objet d'intérêt. Dans le cas où les caractéristiques spatiales des objets détectés se rapprochent à ceux définis dans le modèle, l'approche temporelle permet d'appliquer l'hypothèse sur un mouvement régulier d'objets d'intérêt

## 7.2 Perspectives et travail ultérieur

L'approche basée sur la construction du modèle d'objet d'intérêt dans l'aspect spatial et temporel a été implémentée dans le cadre du projet Octobre Bleu. Il représente la base fondamentale de sa bibliothèque d'algorithmes pour le traitement d'images et vidéos, qui permettent son extension prévue (e.g. pour la vidéosurveillance de trafic routier).

L'algorithme a été testé sur un cas d'application pour le besoin de GIFS d'automatisation de suivi de migrations sur l'ensemble de vidéos avec des caractéristiques communes suivantes :

- le ciel bleu d'une couleur visuellement uniforme ou avec une voile de nuages blancs
- les oiseaux sur un fond bleu ou blanc
- les objets fixes sur l'arrière-plan généralement du même type : végétation, sol

Ces caractéristiques sont propres pour les vidéos dans le cas de suivi de migrations d'oiseaux. Mais la conception d'exécution hiérarchique de l'algorithme, qui débute par le traitement d'entités globales (zones d'intérêt) et converge vers l'entité locale (objets), permet son extension pour des autres types de vidéo et des cas d'application. Les recherches dans le domaine du traitement d'images pour le développement de ces extensions représentent une des axes du travail ultérieur.

Le deuxième axe concerne l'enrichissement de la notion du modèle d'objet. Premièrement le modèle peut être complété par l'analyse supplémentaire en fonction d'une application pour fournir des moyens de distinction d'objets plus affinée. Deuxièmement, le problème de chevauchement d'objets d'intérêt peut être également résolu par une analyse des proportions des paramètres du modèle et ceux de détections en fonction de coefficients de fiabilité associés.

Les travaux de recherches étaient concentrés sur les principes fondamentaux de l'algorithme : l'application de la notion de fiabilité de détection pour le dénombrement d'objets dans les vidéos et de l'aspect spatio-temporelle d'estimation de coefficients de fiabilité, qui représentent cette notion dans l'algorithme. Pour initialiser les références indispensables pour la définition de coefficients en fonction d'une entité hiérarchique analysée (e.g. les couleurs attendues pour les objets d'intérêt et le fond des zones d'intérêt) les moyennes statistiques de la moyenne et la déviation standard ou les valeurs, fournies par l'utilisateur sont actuellement utilisées.

Le troisième axe des futurs travaux consiste alors en conception et développement de méthodes alternatives à l'application d'intervalles de confiance de paramètres et de caractéristiques (e.g. la

couleur et la taille d'objets détecté), permettant leur évolution dynamique en fonction de changements des conditions dans une vidéo.

Parmi les autres perspectives sont la conception des approches d'analyse d'objets d'intérêt au moyen de coefficients de fiabilité pour des déplacements plus complexes qu'une trajectoire linéaire et pour des grands groupes d'objets :

- analyse d'une trajectoire de déplacement circulaire
- gestion de cas, quand les coordonnées d'objets d'intérêt de changent pas (éloignement et rapprochement d'objets à la caméra)
- conception et développement d'une méthode de dénombrement, si des nouveaux objets d'intérêt apparaissent tout au long d'une vidéo (e.g. des grands vols d'oiseaux migratoires, passant devant la caméra)

## **8 Architecture modulaire du projet Octobre Bleu**

Le suivi et le dénombrement des oiseaux migratoires a été le contexte applicatif prévue pour le projet Octobre Bleu. Le concept hiérarchique à la base de l'algorithme proposée est reflété dans l'architecture modulaire du projet, destinée aux implémentations de ces extensions. L'approche modulaire permet ainsi l'abstraction des méthodes concrètes sur les étapes de l'algorithme, et un enrichissement par d'autres méthodes du traitement d'images, destinées aux autres cas d'application.

Du traitement plus général au plus local, l'architecture compris plusieurs modules :

- Module de classification de zones d'intérêt
- Module de détection de contours d'objet et de leur conversion de contours d'objet dans des régions continues
- Module de construction du modèle de petits objets en mouvement par l'aspect spatial et temporel
- Module de sélection d'objets conformes au modèle
- Module d'estimation des coefficients finaux et de du dénombrement d'objets détectés

Chaque module peut contenir plusieurs fonctions, implémentant des méthodes variées du traitement d'images. Ainsi un ensemble de solutions peut être appliqué pour le traitement d'une vidéo concrète, en fonction de sa complexité.

### **8.1 Perspectives d'extensions sur d'autres types d'applications**

Dans le cas d'applications pour la vidéosurveillance des lieux publics et le control du trafic routier sur les autoroutes, prévues pour le projet Octobre Bleu l'utilisation d'autres méthodes du traitement sera nécessaire pour la raison des différences entre les objets d'intérêt (les oiseaux-migrateurs dans les vidéos ne ressemblent pas aux piétons ou des voitures et les zones d'intérêt sont également différentes). Des modifications et des extensions pourront être facilement apportés dans l'architecture modulaire du projet par l'ajout de fonctions supplémentaires sans modifier sa structure, ni le principe du traitement. Pour en donner quelques exemples :

- La sélection de zones d'intérêt peut être effectuée en fonction d'un arrière-plan, qui pourrait être comparé avec un modèle du fond d'image. Dans ce cas, de même que pour les objets

d'intérêt, il serait possible d'introduire un modèle avec des paramètres et des fonctions d'analyse (e.g. estimation de similarité d'images à la place d'analyse de variation d'intensité de couleur dans l'espace colorimétrique YUV)

- L'étape de sélection des contours d'objets d'intérêt peut être enrichie par d'autres filtres, appliqués en fonction de caractéristiques d'une zone d'intérêt
- En fonction d'objets d'intérêt dans le contexte d'application, la construction du modèle ou des modèles d'objet d'intérêt sera orientée vers des approximations variées de leurs apparences dans des vidéos (e.g. une primitive géométrique la plus similaire à un oiseau sera une ellipse plus proche à un cercle, mais pour une voiture cela peut être une ellipse plus étirée ou un rectangle). Ainsi des nouveaux modèles enrichiront les fonctionnalités du projet

## 9 Annexe

### 9.1. Exemples de détections dans certaines vidéos du corpus 2

L'annexe donne des illustrations de résultats dans le tableau 5.4. Des extraits de deux frames consécutifs de quatre vidéos du corpus 2 et les masques binaires d'objets détectés sont présentés dans cette section. Toutes les régions blanches dans les masques binaires participent dans la somme finale d'objets détectés, mais avec les coefficients de fiabilité différents. Les régions, entourées des cercles rouges obtiennent de coefficients proches de 0. Les autres – les coefficients de fiabilité variables dans l'intervalle (0.7 - 1]. Ces objets sont considérés comme des détections correctes à différentes degrés de fiabilité:

- 80-90% d'objets retenus ont un coefficient de fiabilité final  $> 0.8$
- 10-20% ont un coefficient de 0.7 à 0.8

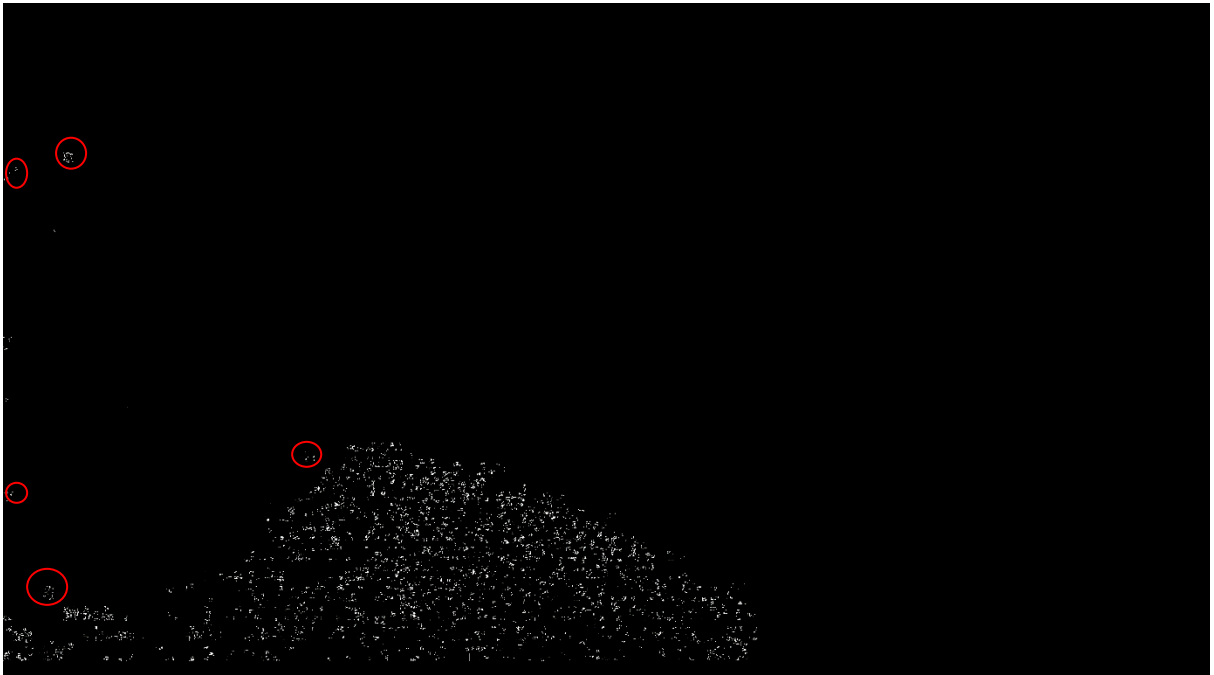


Figure 9.1 : Exemple de détection d'objets d'intérêt dans la vidéo 115 (premier frame). Vérité terrain (comptage à la main) : 1245, résultat de dénombrement par la solution : 1393



Figure 9.2 : Exemple de détection d'objets d'intérêt dans la vidéo 115 (second frame).





Figure 9.3 : Exemple de détection d'objets d'intérêt dans la vidéo 135 (premier frame). Vérité terrain (comptage à la main) : ~1000, résultat de dénombrement par la solution : 1175. Les cercles rouges contiennent les objets avec des coefficients de fiabilité bas (de 0 à 0.2)

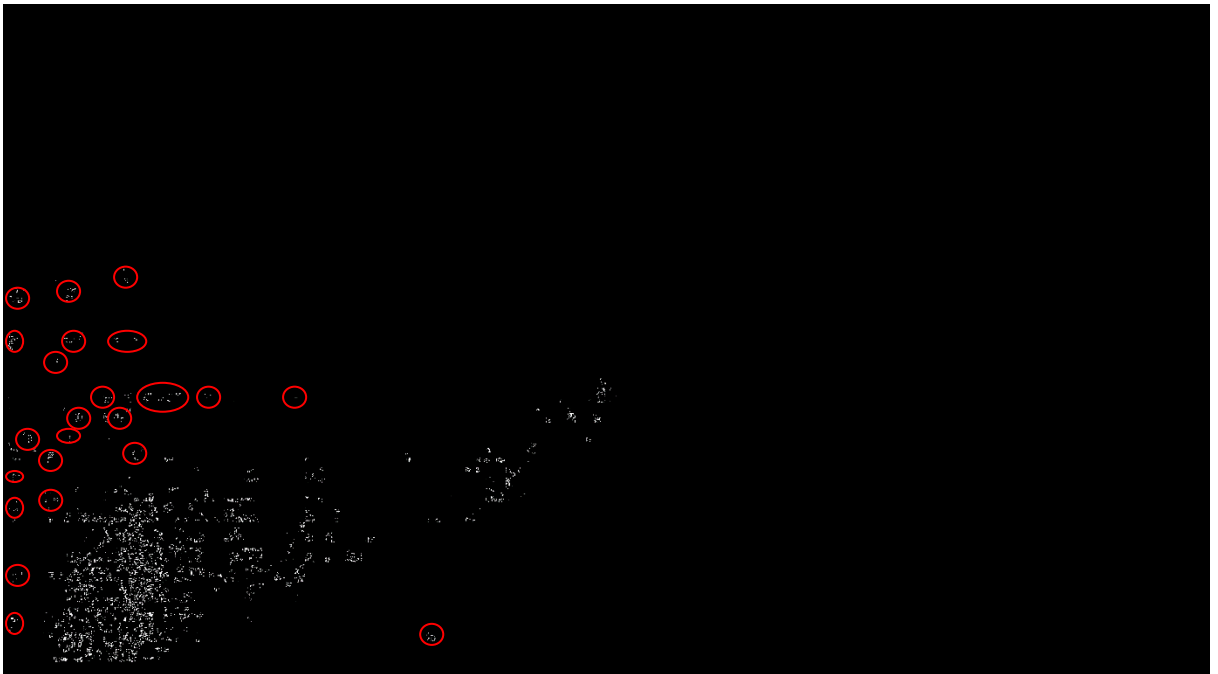


Figure 9.4 : Exemple de détection d'objets d'intérêt dans la vidéo 135 (second frame).

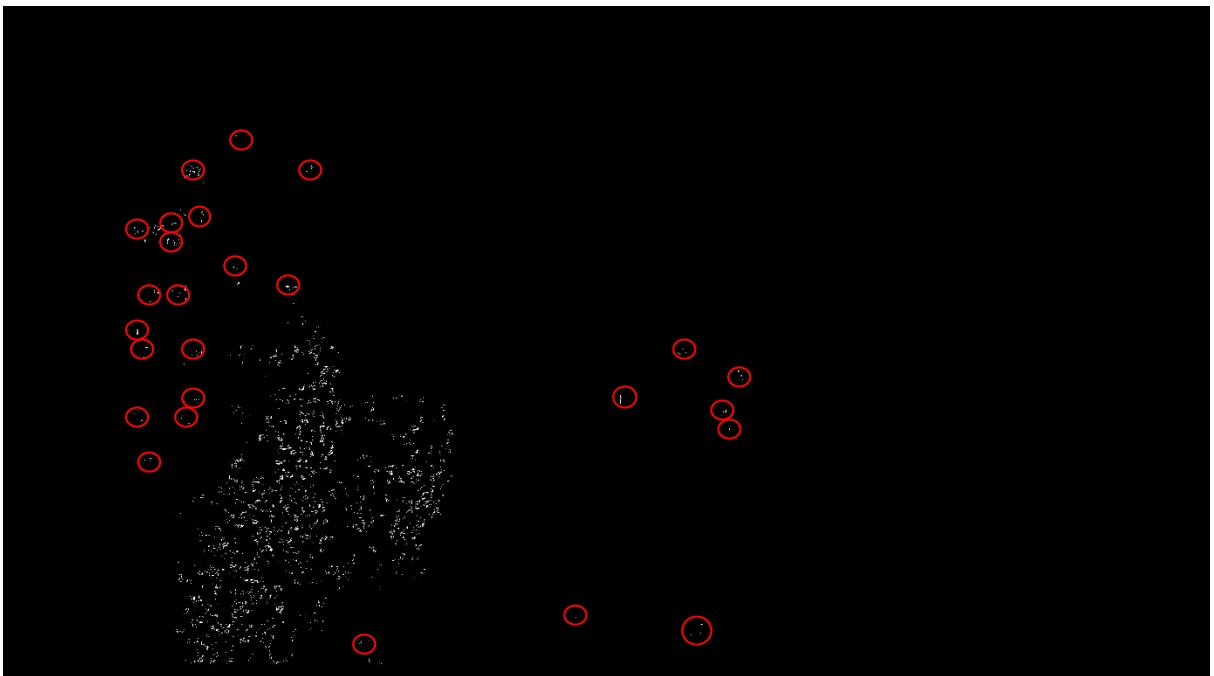


Figure 9.5 : Exemple de détection d'objets d'intérêt dans la vidéo 133 (premier frame). Vérité terrain (comptage à la main) : ~1000, résultat de dénombrement par la solution : 975. Dans ce vidéo les objets avec des coefficients proches du 0 ont des très petites tailles (pratiquement ne font que 2 pixels) et sont présentent également parmi les objets d'intérêt. Ils ne sont pas entourés de cercles rouges, afin de ne pas y enclure des bonnes détections.

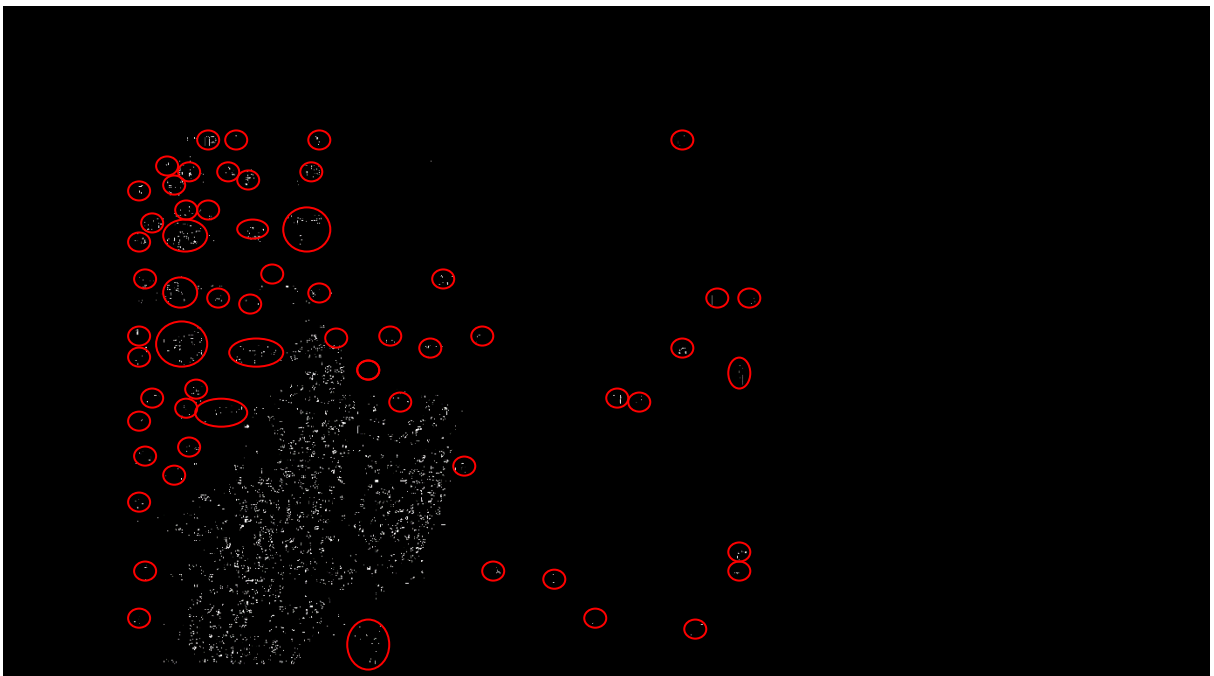


Figure 9.6 : Exemple de détection d'objets d'intérêt dans la vidéo 133 (second frame).

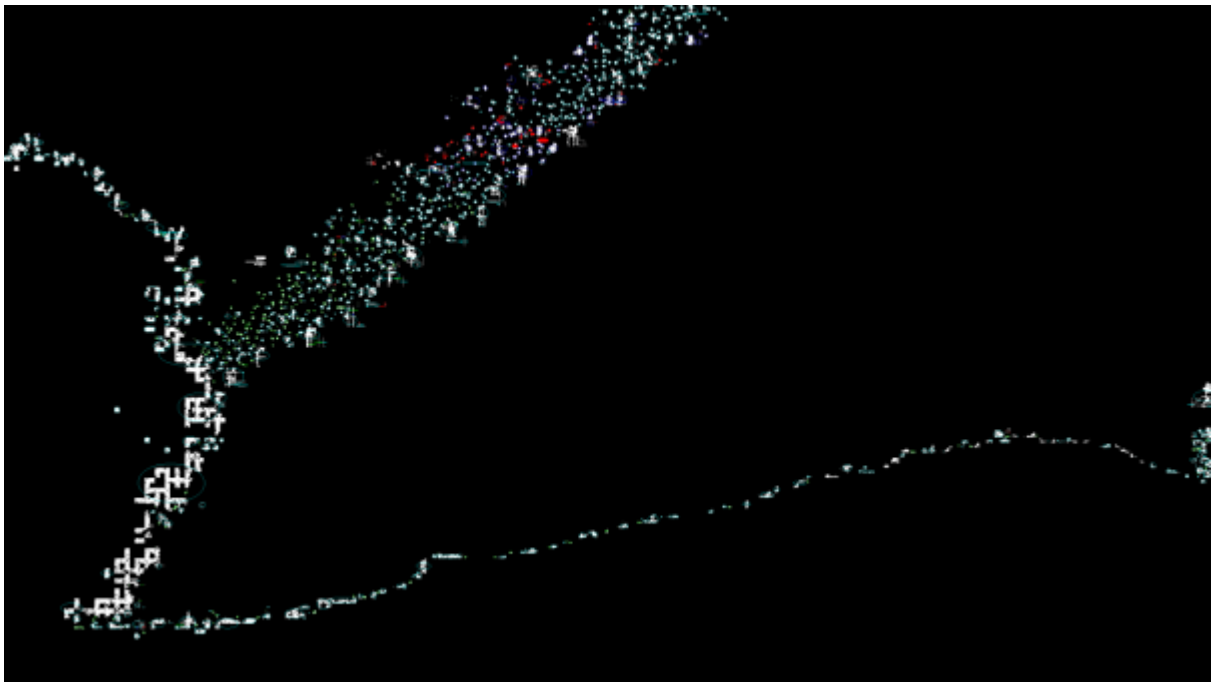


Figure 9.7 : Exemple de détection d'objets d'intérêt dans la vidéo 137. Image intermédiaire d'objets après la conversion de contours en régions continues (détection spatiale) et avant l'application de l'aspect temporel. Cet exemple montre la difficulté de détection de petits objets dans les vidéos réelles d'observation d'oiseaux migratoires, quand les objets dans une scène, similaires aux objets d'intérêt par les paramètres numériques et leurs formes apparentes lors du traitement. Les coefficients de fiabilité permettent de ne pas compter ces objets dans le résultat final et l'approcher à la vérité terrain. Vérité terrain  $\sim 1200$ , résultat de détection avec les coefficients de fiabilité : 1054

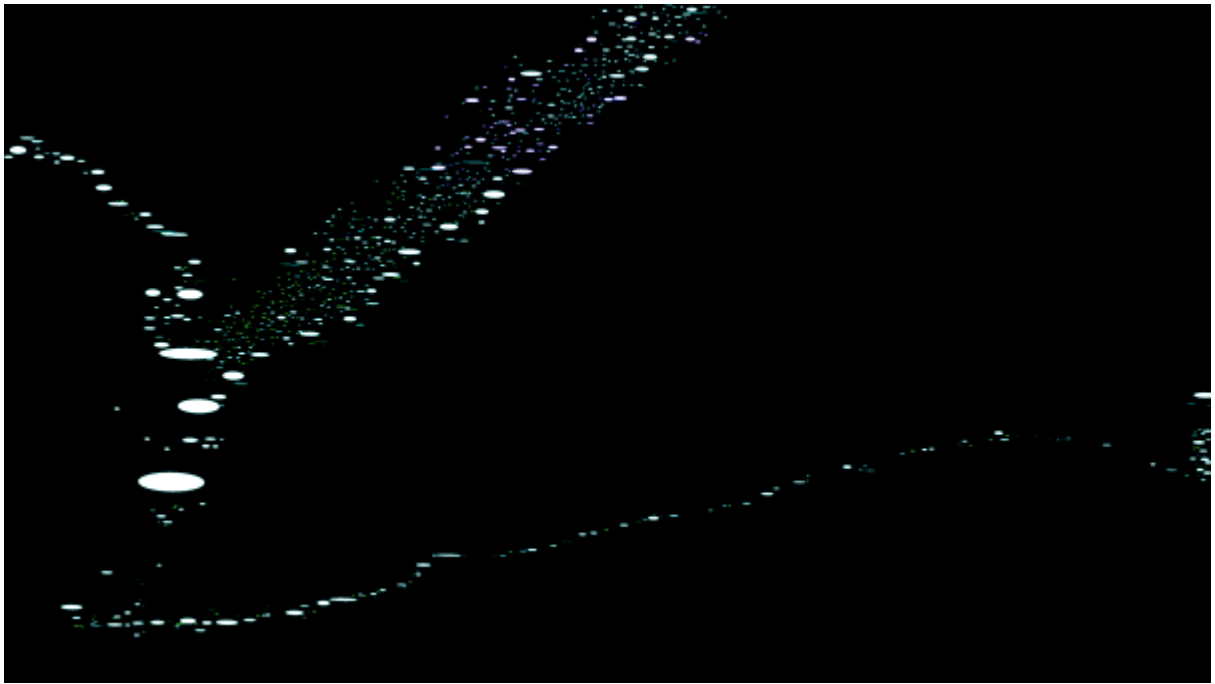


Figure 9.8: Résultats après la sélection spatiale et l'application de coefficients de fiabilité sur la vidéo 137 : les grandes ellipses blanches représentent schématiquement des grandes objets – les branches d'arbres en mouvement avec le vent, la végétation en légère mouvement avec le vent. Les petites ellipses d'une couleur verte sont des objets avec les coefficients de fiabilité proches de 0.9, bleu clair – de 0.7 à 0.9, bleu foncé – de 0.5 à 0.7, rouge – de 0.3 à 0.5. Les coefficients de fiabilité d'un groupe d'objets d'intérêt, inférieurs à la majorité d'objets d'intérêt (les oiseaux dans le vol entourés des ellipses bleues foncées) s'expliquent par les coefficients de l'aspect temporel plus bas. Cette situation arrive, quand le déplacement d'un groupe n'est pas suffisant pour le prendre en considération, où quand un déplacement n'est pas pris en compte en raison de dissimilarité d'objets dans un group suivi. Les objets de la figure, qui ne sont pas présents à la différence de la figure 9.7 n'ont pas été sélectionnés par l'aspect temporel.



Figure 9.9 : Un cas difficile, où les oiseaux disparaissent derrière les reflets du soleil, qui implique des fluctuations de leurs nombres au cours de la vidéo.

## 10 Bibliographie

- [1] Techniques de suivi des animaux. Portail éducol : <http://eduscol.education.fr/localisation/pedago/argos1/techniq.htm>
- [2] CNRS. Centre de Recherches sur la Biologie des Population d'Oiseaux. Bagueage d'oiseaux : <http://crbpo.mnhn.fr/spip.php?article5&lang=fr> , <http://www.isnea.eu/un-nouveau-reseau-de-suivi-des-oiseaux-migrateurs/>
- [3] Suivi par Radar de la Migration Postnuptiale dans l'Estuaire de la Loire. Association ACROLA-BIOTOPE. Janvier 2008. ([http://www.acrola.fr/BIOTOPE\\_DONGES\\_radar\\_01.pdf](http://www.acrola.fr/BIOTOPE_DONGES_radar_01.pdf))
- [4] Systèmes de localisation par satellite-récepteur. Portail éducol : [http://eduscol.education.fr/localisation/phys/systemes\\_loc\\_recepteurs/default.htm](http://eduscol.education.fr/localisation/phys/systemes_loc_recepteurs/default.htm)
- [5] LPO Auvergne-Rhône-Alpes. <https://auvergne-rhone-alpes.lpo.fr/actions/suivi-de-la-faune-et-sciences-participatives/programmes-de-suivis/oiseaux-migrateurs/article/le-suivi-des-oiseaux-migrateurs>
- [6] Les 10 Ans de « Max » Suivi à Long Terme d'une Cigogne Blanche par Satellites, Adrian AEBISCHER & André FASEL, Etat de Frigourg
- [7] Association Mission Migration : [https://www.migraction.net/index.php?m\\_id=22003&item=3](https://www.migraction.net/index.php?m_id=22003&item=3)
- [8] Jing Yi Tou, and Chen ChuanToh, "Optical Flow-Based Bird Tracking and Counting for Congregating Flocks", *Intelligent Information and Database Systems*, Springer Berlin Heilderberg, pp. 514-523, 2012
- [9] Jyh-Yeong Chang , Wen-Feng Hu , Mu-Huo Cheng , Bo-Sen Chang "Digital image translational and rotational motion stabilization using optical flow technique" *IEEE Transactions on Consumer Electronics*, Volume: 48 , Issue: 1 , Feb 2002
- [10] Allen, Paul E., and Thorpe, Charles E. *Some Approaches to finding Birds in Video Imagery*, Research, the Robotics Institute, Carnegie Mellon University, Pittsburg, Pennsylvania, US, December 1991.
- [11] Reyes, Elsa, *A Comparison of Image Processing Techniques for Bird Detection*, Thesis, the Faculty of California Polytechnic State University, 2014.
- [12] Cansın Yıldız « An Implementation on Histogram of Oriented Gradients for Human Detection »
- [13] Navneet Dalal and Bill Triggs Histograms of Oriented Gradients for Human Detection INRIA Rh.one-Alps Dept. of Computer Engineering Bilkent University Ankara,Turkey
- [14] Roy-Erlend, Berg, *Real-time people counting system using video camera*, Thesis, Department of the Computer Science and Media Technology, Gjøvik University College, Norway, 2008.
- [15] Zhang Xiao-yan, Wu Xiao-juan, Zhou Xin, Wang Xiao-gang, and Zhang Yuan-yuan, "Automatic Detection and Tracking of Maneuverable Birds" *International Conference on Computational Intelligence and Security*, Springer Berlin Heilderberg, pp. 514-523, 2008
- [16] Zhang X., Wu X., Zhou X., Wang X., Zhang Y. (2009) Birds' Flap Frequency Measure Based on Automatic Detection and Tracking in Captured Videos. In: Lim C.T., Goh J.C.H. 13th International Conference on Biomedical Engineering. IFMBE Proceedings, vol 23. Springer, Berlin, Heidelberg

- [17] Li Wen, and Song, Dezhen, « Automatic Bird Species Detection Using Periodicity of Salient Extremities », *International Conference on Computational Intelligence and Security*, Springer Berlin Heilderberg, pp. 514-523, 2008.
- [18] Wassim Bouachir, “Thèse : Suivi d’objets par caractéristiques locales, encodant la structure”, Département de Génie Informatique et Génie Logiciel, Ecole Polytechnique de Montréal, Université de Montréal.
- [19] Aurélie Bugeau, “Thèse : Détection et suivi d’objets en Mouvement dans des Scènes Complexes. Application à la Surveillance de Conducteurs”. Université de Rennes 1, 4 Janvier 2011
- [20] Kwang Moo Yi, Kimin Yun, Soo Wan Kim, Hyung Jin Chang, Hawook Jeong and Jin Young Choi « Detection of Moving Objects with Non-Stationary Cameras in 5.8ms: Bringing Motion Detection to your Mobile Device » Seoul National University, Seoul, Korea *Computer Vision Foundation, CVFR Workshop 2013*
- [21] Motion Detection Based on Optical Flow and Self-adaptive Threshold Segmentation Shui-gen Wei a, Lei Yang a, Zhen Chen, Zhen-feng Liu
- [22] Paris, S. Hasinoff, S. W. and Kautz, J. “Local Laplacian Filters: Edge-aware image processing with a laplacian pyramid”, 2011 *ACM Transactions on Graphics*, 30, 4.
- [23] D. Sapra, Angel, Geronimo, David, Dornaika, Fadi, Rouhanni, Mohammad, and Lopez, Antonio M. “Moving Object Detection from Mobile Platforms Using Stereo Data Registration” *Computational Intelligence Paradigms*, pp. 25-37, 2012.
- [24] Yu Huang, Joan Llach, Chao Zhang, « A Method of Small Object Detection and Tracking Based on Particle Filters », Thomson Corporate Research, Princeton, US, State Key Lab of Machine Perception, Peking University, Beijing, China
- [25] Lucas J. van Vliet, Ian T. Young, and Guus L. Beekers, “A Nonlinear Laplace Operator as Edge Detector in Noisy Images” *Computer Vision, Graphics, and Image Processing 45*, Academic Press, Inc, pp. 167-195, 1989.
- [26] L. Guilbaud and P. Lacour, Déplacement de contours par passage par zéro du laplacien à l’aide de divers paramètres, Signal and Image Processing Department, Telecom ParisTech.
- [27] Horecki, D. Pautus and K. Wojciechowski, K.-H. Franke “Object localization using color histograms”, *International Conference on Image Processing*, IEEE, Volume 3, pp. 492-495, Vancouver BC, 2000
- [28] B. Vishnyakov, V. Gorbatshevich, S. Sidiyakin, Y. Vizilter, I. Malin, A. Egorov « Fast Moving Objects Detection Using iLBP Background Model », State Research Institute of Aviation Systems (FGUP GosNIIAS), Moscow, Russia
- [29] P. Dunne, “Histogram-based Detection of Moving Objects for Tracker Initialization In Surveillance Video” *International Journal of Grid and Distributed Computing*, Vol.4, No.3, September 2011.
- [30] Benjamin Lefaudeux « Thèse : Détection, localisation et suivi des obstacles et objets mobiles à partir d’une plateforme de stéréovision », p. 61, Informatique temps réel, robotique et automatique, ParisTech, 30 Septembre 2013
- [31] Shun-Wen Cho, Chiou-Shann Fuh, Kai Essig « Content-Based Image Retrieval System by Hierarchical Color Image Region Segmentation », Department of Computer Science and Information Engineering, National Taiwan University Taipei, Taiwan



- [32] Vincent Garcia, Eric Debreuve, Michel Barlaud « Méthode de suivi d'objets basée sur des trajectoires temporelles de points d'intérêt, Colloque GRETSI, 11-14 septembre 2007, Troyes
- [33] Sung Hyuk-Cha, "Taxonomy of Nominal Type Histogram Distance Measure" *American Conference on Applied Mathematics*, Harvard, Massachusetts, USA, March 24-26 2008.
- [34] Yang Chao, A comparison of medical image analysis algorithms for edge detection, Bachelor Thesis, Sweden, 2010
- [35] Mamta Juneja, Parvinder Singh Sandhu; "Performance Evaluation of Edge Detection Techniques for Image in Spatial Domain", *International Journal of Computer Theory and Engineering* Vol. 1, No. 5, December 2009
- [36] David P. Doane, Lori W. Seward, "Measuring skewness: A Forgotten Statistics?," *Journal of Statistics Education*, Volume 19, Number 2, 2011
- [37] Raman Maini and Dr. Himanshu Aggarwal, Study and Comparison of Various Image Edge Detection Techniques, Punjabi University Patiala, Punjab, India, *International Journal of Image Processing (IJIP)*, Volume 3 : Issue 1
- [38] Satbir Kaur and Ishpreet Singh « Comparison between Edge Detection Techniques », CSE & IT Department Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, *International Journal of Computer Applications*, Volume 145 – No.15, July 2016
- [39] Sung-Hyuk Cha "Comprehensive Survey on Distance/Similarity Measures Between Probability Density Functions", *International Journal of Mathematical Models and Methods in Applied Sciences* Issue 4, Vol. 1, 2007, p. 303
- [40] Edith Anderson Feisner, Ron Reed, "Color Studies", Fairchild Books, New York, Third Edition, p. 13, 2014
- [41] Matthews, B. W. (1975). "Comparison of the predicted and observed secondary structure of T4 phage lysozyme". *Biochimica et Biophysica Acta (BBA) - Protein Structure*. 405 (2): 442–451.
- [42] Jin Zhao, Bimal K. Bose, "Evaluation of Membership Functions for Fuzzy Logic Controlled Induction Motor Drive", Huazhong University of Science and Technology, The University of Tennessee, 2002
- [43] Wang, Chonghua, "A Study of Membership Functions on Mamdani-Type Fuzzy Inference System for Industrial Decision-Making" Theses and Dissertation. Paper 1665, 2015
- [44] Yehu Shen and Qicong Wang "Sky Region Detection in a Single Image for Autonomous Ground Robot Navigation", *International Journal of Advanced Robotic Systems*; July 2013
- [45] Konica Minolta. Color comparison : <http://sensing.konicaminolta.us/2014/04/identifying-color-differences-using-l-a-b-or-l-c-h-coordinates/>
- [46] Cambridge in Color. A Learning Community for Photographs <http://www.cambridgeincolour.com/tutorials/digital-camera-sensor-size.htm>
- [47] Analyse de la symétrie d'une distribution, Université Libre de Bruxelles, [http://www.itse.be/statistique2010/co/241\\_Cours\\_asymetrie.html](http://www.itse.be/statistique2010/co/241_Cours_asymetrie.html)
- [48] Cours de Master Statistique. Régression linéaire. Arnaud Guyader. Université de Rennes
- [49] Horecki, D. Pautus and K. Wojciechowski, K.-H. Franke "Object localization using color histograms", *International Conference on Image Processing*, IEEE, Volume 3, pp. 492-495, Vancouver BC, 2000

- [50] P. Dunne, “Histogram-based Detection of Moving Objects for Tracker Initialization In Surveillance Video “*International Journal of Grid and Distributed Computing*”, Vol.4, No.3, September 2011.
- [51] Sung Hyuk-Cha, “Taxonomy of Nominal Type Histogram Distance Measure” *American Conference on Applied Mathematics*, Harvard, Massachusetts, USA, March 24-26 2008.
- [52] Alexandre Girard, Bruno Charbonnier, Guy d’Urso « Une approche par traitement statistique d’images du comptage de poils de brosses à mascara » , Chatou Cedex, France
- [53] Jacques BREMONT, Vincent BOMBARDIER, Oliver PEREZ-ORAMAS « Détection de contours par règles linguistiques floues », Centre de Recherche en Automatique de Nancy, CNRS ESA 7039 Equipe PRAISSIH Université Henri Poincaré, Vandoeuvre-lès-Nancy, France
- [54] Gaurav Sharma, Wencheng Wu, Edul N. Dalal « The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations » Department and Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14627-0126 2 Xerox Corporation, 800 Phillips Road, Webster, NY 14580