

# Suivi d'objets dans une scène étendue par un système de vision distribué, application à la surveillance d'environnement autoroutier

---

## Object tracking over wide area using distributed vision, Application to freeway surveillance

par **Cina MOTAMED, Olivier WALLART**

Laboratoire LASL EA 2600, Université du Littoral Côte d'Opale, Bat 2, 50 rue F. Buisson 62228 Calais, motamed@lasl-gw.univ-littoral.fr

### *résumé et mots clés*

Cet article présente l'architecture logique d'un système de perception distribué. Le système est constitué de capteurs de vision autonomes analysant la scène selon des points de vue distincts et sans recouvrement. Le suivi global des objets est réalisé par l'intermédiaire d'une ré-identification des objets perçus par plusieurs caméras. Les principales difficultés du système d'interprétation sont la gestion des erreurs lors de la ré-identification des objets et des contraintes temps réel que le système doit respecter. La gestion des erreurs est prise en charge par un raisonnement approximatif basé sur la théorie des possibilités et les aspects temps réel du système sont gérés par un raisonnement distribué et temporel.

Perception multi-capteurs, Fusion temporelle, Suivi d'objets, Gestion de l'incertitude.

### *abstract and key words*

This work presents the architecture of a distributed perception system. Each vision unit analyzes the scene from its distinct point of view and without recovery. The global tracking is carried out by a re-identification of the objects perceived by several cameras. The difficulties of the system concern the management of the uncertainty at the identification stage and the real time constraints. The management of uncertainty is dealt with an approximate reasoning approach by using the possibility theory. A distributed and temporal reasoning approach takes into account real time constraints.

Multi-sensor perception, Temporal fusion, Tracking, Uncertainty management.

## 1. introduction

L'interprétation de séquences d'images pour la surveillance des activités humaines est un processus complexe qui impose de créer des modules spécialisés. Leurs rôles respectifs sont dans l'ordre, la détection, le suivi des objets mobiles, la classification des objets et l'analyse des comportements afin de reconnaître des scénarios modélisés. La complexité de ce type de système impose d'apporter des informations *a priori* permettant de configurer et d'orienter les différents étages de perception. Ces informations représentent le contexte d'une application de vidéosurveillance [2]. Une nouvelle génération des systèmes de vidéosurveillance exploitant la notion de réseaux de caméras est en cours d'investigation [24]. La principale évolution de ces systèmes est la distribution des fonctions locales d'interprétation bas niveau sur des caméras intelligentes dotées de capacités de traitements temps réel et de communication. Seules les informations de haut niveau transitent entre les nœuds de ce réseau. Les traitements bas niveau englobent la détection, le suivi local et la classification des objets. Ce type d'architecture est particulièrement utile pour faire face aux contraintes temps réel d'un système d'interprétation en ligne et plus particulièrement en environnement multi-caméras. Ce travail présente l'architecture logique d'un système de poursuite d'objets à partir de modules de visions distribués. Ces modules constitués chacun d'une caméra et d'une unité locale de traitement analysent la scène selon des points de vue distincts et sans recouvrement (voir figure 1). L'application envisagée consiste à développer un système de Détection Automatique d'Incidents (DAI) en environnement autoroutier. Les systèmes actuels de DAI basés sur la vidéo utilisent plusieurs systèmes de vision et couvrent l'ensemble du secteur à surveiller. L'analyse des déplacements et des stationnarités des objets dans le champ de chaque caméra permet de détecter un ensemble d'incidents modélisés [1][11]. Les scénarios recherchés sont les arrêts sur la voie d'urgence, les accidents, les bouchons, les véhicules à contresens et les véhicules lents. Le problème essentiel de ces dispositifs est la nécessité d'une couverture totale de la scène demandant ainsi le déploiement d'un nombre très important de systèmes de vision autonomes (tous les 300 mètres en moyenne). L'apport d'un système d'interprétation multi-capteurs dans ce domaine peut être une réduction du nombre de caméras tout en essayant d'assurer des fonctions utiles pour la sécurité.

L'architecture proposée traite uniquement la fonction de suivi global des objets par une ré-identification des d'objets perçus de manière séquentielle par plusieurs modules de vision. La ré-identification consiste à mettre en correspondance chaque observation (objet perçu) avec un objet préalablement observé (objet attendu) par le système. La surveillance visuelle de la scène sera effectuée de jour, tout en tolérant des variations d'illumination. La notion de surveillance est considérée au sens strict, c'est-à-dire que le système doit éviter de perdre ou de confondre les

objets dont les pistes ont été initialisées. Le système doit intégrer le concept de « sécurité positive » en détectant les situations où le risque d'erreur, associée à une décision, est jugé important. Une particularité importante du problème envisagé est l'éloignement des caméras entre elles pouvant atteindre plusieurs kilomètres. Pour un réseau de surveillance, l'intérêt de cette dispersion est principalement de nature économique car elle permet de couvrir des zones de surveillance vastes en minimisant le nombre de stations de mesures.

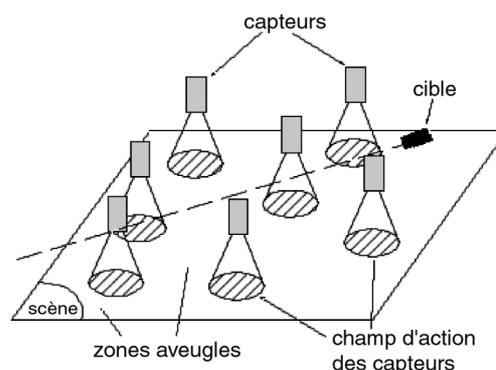


Figure 1. – Organisation multi-capteurs.

Le problème de ré-identification d'objets dans le domaine du transport a été principalement abordé dans les nouveaux systèmes de mesure de temps de parcours entre deux nœuds distants d'un réseau autoroutier [3][22]. L'indicateur des temps de parcours est utile pour les informations aux usagers, pour les systèmes de régulation du trafic et également pour la détection d'incidents. Les capteurs utilisés sont majoritairement des boucles magnétiques. Dans le but d'obtenir des mesures de temps plus précises, des systèmes à double boucles ont été développés. Ces doubles boucles appelées aussi trappes permettent de calculer la vitesse locale et de prédire l'ordre d'arrivée des objets à la prochaine trappe, sous l'hypothèse d'une accélération nulle. Coifman [4] a proposé un algorithme qui permet d'estimer la longueur des objets grâce à ces trappes placées sur chaque voie d'autoroute. Ensuite, le système identifie les séquences d'objets d'une trappe à l'autre en exploitant l'ordre prévu d'arrivée et la longueur des véhicules. Cependant le taux de mauvaise association reste élevé. Les difficultés d'association proviennent d'une part de la pauvreté de l'information sur les objets et d'autre part des erreurs de prédiction. En effet, ces systèmes considèrent que les objets suivent des chemins uniques imposés, avec un modèle de comportement cinématique donné. Cui et Huang ont étudié la possibilité de discriminer visuellement les véhicules à partir de leurs plaques d'immatriculation [5]. Cette approche directe n'a cependant pas pu être retenue à cause des contraintes liées au respect de la vie privée. De plus l'obtention de cette mesure n'est pas toujours aisée à grande vitesse.

Les avantages des capteurs à base de caméras sont la facilité d'installation dans une infrastructure routière existante et aussi la richesse en terme de quantité d'informations produites [1][4]. Les boucles magnétiques ou les radars [16] disposent de données moins discriminantes que les caméras mais ont l'avantage de fonctionner indépendamment des conditions d'éclairage.

Globalement, les systèmes actuels de ré-identification restent relativement limités car ils ne fonctionnent qu'avec un seul couple de capteurs dans des environnements simples, sans entrée ni sortie.

La réalisation des tâches de ré-identification des objets à partir de configuration multi-caméras distantes dans les scènes de transport pose plusieurs niveaux de difficultés :

- Les systèmes de vision travaillant sous éclairage indépendant doivent extraire des descripteurs stables dans le temps et invariants entre deux capteurs.
- Le système global doit gérer un flux d'objets qui peut devenir important.
- Sur les voies de transport, de nombreux objets mobiles peuvent présenter des apparences proches. Ce phénomène induit alors des ambiguïtés importantes lors de la mise en correspondance visuelle.
- L'absence de couverture totale de la zone de surveillance entraîne la mise en place de mécanismes de prédiction associés au déplacement des objets. Cependant la modélisation des comportements dynamiques liés aux activités humaines, en vue d'une prédiction, demeure difficile car elle doit tolérer de grandes variabilités. Le choix des trajets et la variation de comportement dynamique dans un environnement autoroutier peut dépendre du type de véhicule, du conducteur et aussi de l'environnement (°limitation de vitesse, trafic, bouchon, déviation, etc..).

En résumé, la tâche du système d'interprétation devient naturellement difficile car elle doit supporter le flux d'entrée des capteurs, les incomplétudes et les incertitudes des informations disponibles.

La section 2 présente la stratégie globale et l'architecture distribuée du système d'interprétation. La section 3 traite du choix des descripteurs visuels qui permettent de modéliser l'apparence des observations. La section 4 expose la phase de prédiction, qui a pour objectif de définir l'espace de recherche des candidats à la mise en correspondance. L'étape finale de décision de mise en correspondance, exploitant l'ensemble des informations visuelles et temporelles, est détaillée dans la section 5. La validation des algorithmes d'analyse d'images a été envisagée sur des données réelles à partir de séquences enregistrées, issues d'un couple de caméras distantes, placées le long d'une autoroute. L'étape de mise en correspondance a été validée à partir de données multi-capteurs générées par un simulateur de trafic autoroutier intégrant un réseau paramétrable de capteurs. Quelques résultats de simulation sont présentés dans la section 6.

## 2. architecture logique du système

Ce travail présente la mise en place d'une architecture logique d'un système de suivi multi-capteurs. Elle définit la façon dont le système doit combiner les informations et organiser ses ressources selon le but recherché. Cette architecture intègre des mécanismes de raisonnements évolués de type distribué et temporel et utilise des outils de gestion de l'incertitude afin de minimiser les décisions erronées. L'utilisation du raisonnement distribué est relativement récente dans le domaine de l'intelligence artificielle. Il est particulièrement adapté au caractère distribué de certains problèmes. La distribution peut être géographique, fonctionnelle ou méthodologique. Les intérêts recherchés par la distribution des tâches sont : la réduction de complexité, la sûreté de fonctionnement et la modularité. Dans le cas de perception multi-caméras, l'objectif principal est de limiter considérablement les échanges de données dans le système en évitant les transports de données vidéo brutes qui demanderaient une bande passante importante. Ainsi, seules les informations de haut niveau peuvent transiter dans le réseau de communication numérique à débit modéré.

Le raisonnement temporel a pour objectif de contrôler le processus global de perception. Les connaissances du système d'interprétation peuvent évoluer dans le temps, il est donc important que le système puisse utiliser un raisonnement temporel afin de mettre à jour (ou réviser) ses connaissances. De plus, il est souvent nécessaire de pouvoir mener un raisonnement sans disposer de toutes les informations utiles et de réviser ultérieurement une décision prise à partir de nouvelles informations.

Le but de la gestion de l'incertitude est de générer, à partir de données imparfaites, des conclusions satisfaisantes ou nuancées. Dans un contexte de surveillance, elle apporte une stratégie prudente d'interprétation. Le cadre de gestion de l'incertitude choisi est la théorie des possibilités. Ce cadre permet de modéliser et de gérer de manière flexible les informations de natures très diverses. En particulier, par rapport à la théorie des probabilités, l'approche possibiliste permet, comme la théorie de l'évidence, de bien modéliser la notion d'ignorance [8]. La modélisation des informations sous forme possibiliste est également bien adaptée pour manipuler les incertitudes de mesures issues de capteurs [17].

Afin de faciliter la mise place des différents composants du système d'interprétation, les principales informations contextuelles exploitées sont :

- Les informations de configuration de la scène qui englobent le positionnement global des capteurs (placement, orientation de la caméra), le calibrage des caméras par rapport à la scène, et une modélisation des zones reliant les caméras ( zones aveugles).
- Les informations liées aux classes des objets mobiles qui représentent des contraintes qui peuvent être visuelles (forme,

dimensions) et aussi dynamiques (différents comportements cinématiques possibles).

Dans la littérature, deux types d'approche d'organisation multi-capteurs liée aux missions de suivi des objets mobiles sont proposés.

La première approche est orientée vers la fusion d'informations. Par exemple, le filtre de Kalman distribué DKF combine plusieurs filtres de Kalman associés aux différents capteurs afin d'augmenter la robustesse du suivi [23][26]. Les filtres travaillent en parallèle, en réalisant leurs propres prédictions. Une deuxième étape permet de combiner dans une centrale de fusion, l'ensemble de ces prédictions afin de réaliser un suivi global de toutes les cibles, en exploitant la redondance des informations.

Une deuxième approche exploite une stratégie de type perception active [19]. Le système utilise de manière dynamique un ensemble de capteurs orientables vers les différents objets d'intérêt de la scène. Des agents logiques sont associés à chaque capteur. L'objectif recherché est l'organisation de ces agents en fonction des difficultés qui peuvent être générées soit par un nombre important d'objets, soit par la présence d'occultations statique (obstacle fixe) ou dynamiques (entre les objets mobiles). Les architectures logiques employées s'approchent du concept de société d'experts développé dans le domaine des systèmes distribués. Les sociétés sont formées dynamiquement et utilisent un ensemble d'agents qui sont en mesure d'apporter un avis sur des objets communs afin d'améliorer le suivi.

Dans les systèmes cités précédemment, les champs d'action des capteurs permettent de couvrir la totalité de la scène à analyser. Les capteurs sont soit co-localisés ou présentent un chevauchement partiel de leurs champs de perception et permettent ainsi au système d'interprétation d'utiliser la redondance et la complémentarité des informations. Dans notre configuration, la présence de grandes zones aveugles induit des informations souvent partielles et complémentaires mais jamais temporellement redondantes sur un même objet.

Nous rappelons que le système de perception doit être capable de réaliser une mise en correspondance entre les observations et les objets attendus. Après chaque mise en correspondance, les objets perçus deviennent à leur tour, et ce de manière récursive, des objets attendus. Afin d'optimiser cette gestion d'objets, il serait intéressant que le système puisse « focaliser » la mise en correspondance d'une observation avec un ensemble restreint de candidats. Cette focalisation, qui correspond à la notion de fenêtre de validation dans notre système, peut être aussi bien géographique que temporelle. La focalisation géographique est induite par le fait qu'un objet, entre deux observations, ne peut suivre qu'un nombre limité de chemins et donc ne peut que se retrouver devant un ensemble réduit de capteurs du réseau. Dans ces conditions, le système attend une prochaine réapparition d'un objet à un nombre limité d'endroits. La focalisation temporelle permet quant à elle, de définir la fenêtre temporelle d'apparition d'un événement attendu devant un capteur donné. Ces contraintes temporelles gèrent la durée de vie de chaque objet attendu et per-

mettent ainsi de maîtriser dynamiquement le nombre d'objets prévus à chaque instant sur l'ensemble du système.

L'architecture logique choisie est basée sur une approche de perception active selon une organisation de type sociétés d'experts qui permet d'exploiter pleinement les possibilités des focalisations. Les experts sont créés au sein de modules de vision locaux et ont pour mission de participer à la tâche de poursuite globale d'un objet donné pendant le temps d'une mise en correspondance. Ces sociétés d'experts sont générées dans la fenêtre géographique obtenue à l'aide des connaissances contextuelles de nature topologique associées aux chemins possibles entre les différents modules de visions. Ces sociétés ont des durées de vie maximales imposées par les fenêtres temporelles des objets dont elles dépendent.

Une « piste » dans le vocabulaire de poursuite de cibles représente une collection de mesures ayant une origine (source) commune. La gestion de ces pistes a pour but de s'intéresser à leurs initialisations, leurs maintenances et leurs terminaisons. La maintenance et la terminaison des pistes au sein de notre architecture sont réalisées à tour de rôle par les derniers experts ayant ré-identifié l'objet. Après chaque ré-identification par un expert noté  $E^1$ , une société d'experts est créée à son initiative dans l'objectif de pouvoir poursuivre la piste de l'objet de manière coopérative. Si l'un des experts, noté  $E^2$ , de la nouvelle société reconnaît l'objet, l'expert  $E^1$  est prévenu par  $E^2$ . La piste de l'objet est alors maintenue et la société d'experts courante est détruite. Le suivi est ensuite pris en charge par l'expert  $E^2$  qui, à son tour, crée une nouvelle société.

Au sein d'une société, la communication est réalisée par envoi de messages à tous les experts concernés, c'est-à-dire à l'ensemble des experts susceptibles de voir apparaître l'objet attendu. La tâche de suivi d'objets par leur ré-identification est alors réalisée sous forme coopérative et distribuée. Cette organisation contribue à diminuer le nombre de messages sur l'ensemble du réseau et permet d'augmenter l'autonomie du système global. Les messages au sein de ces sociétés contiennent 2 types d'information :

- les contraintes temporelles,
- une modélisation visuelle de l'objet.

La tâche de surveillance de haut niveau pour la détection d'incidents, sera quant à elle de type centralisé. Elle sera réalisée par le biais d'une unité de surveillance globale en se basant principalement sur les résultats du suivi des objets.

### 3. la mesure de compatibilité visuelle

L'objectif de ce paragraphe est de présenter l'estimation d'une mesure de compatibilité visuelle entre objets perçus par différentes caméras, à des instants différents. Le système multi-

caméras présent est composé d'une seule caméra fixe par zone surveillée. Une première étape consiste en l'extraction des objets par un module de détection du mouvement. Dans le cas de caméras fixes, l'algorithme le plus direct est basé sur une distance entre l'image courante  $I$  et une image de référence  $R$ . Cette image de référence est réactualisée en continu afin d'intégrer les variations lentes d'éclairage [20][30]. La distance utilisée, exploite la différence des trois composantes couleurs RVB des images. Après seuillage, le résultat de la détection pour chaque pixel  $P$  correspond à des régions de l'image contenant des masques de la projection des objets mobiles. Le seuil de détection  $\omega$  est réglé à deux fois la variance du bruit dans l'image estimée à partir d'une séquence préliminaire sans objet mobile. La décision de détection est définie par l'algorithme suivant :

$$\begin{aligned} &\text{si } \left( \max_{c=R,V,B} |I_c^k(P) - R_c^{k-1}(P)| > \omega \right) \\ &\text{alors } D^k(P) = 1 \\ &\text{sinon } D^k(P) = 0 \end{aligned} \quad (1)$$

Malheureusement, dans les scènes autoroutières, les masques de détection présentent un grand nombre d'imperfections, pouvant être :

- les erreurs de non-détection lorsque l'objet a des couleurs proches du fond,
- les problèmes d'ombres qui se traduisent par un élargissement des régions de détection,
- les frontières des masques d'objets qui sont bruitées lorsque les objets possèdent des vitesses élevées.

Nous avons choisi d'utiliser des descripteurs robustes et complémentaires permettant de tolérer en partie les imperfections citées précédemment. La mesure de compatibilité entre objets pour chaque descripteur intègre une évaluation de l'incertitude liée à la notion d'intervalle de confiance [12]. Cette évaluation de l'incertitude est réalisée de manière statistique, pour chaque descripteur, à partir d'une séquence préliminaire. Cette gestion de l'incertitude, qui a été réalisée dans un cadre possibiliste, a pour rôle de faciliter la combinaison des informations lors de l'estimation de la mesure de compatibilité globale.

### 3.1. mesures dimensionnelles

Dans le cas d'objets 3D tel que les véhicules, il est souvent difficile de trouver directement des informations tridimensionnelles à partir de la seule projection de l'objet dans l'image. Dans le domaine du transport, quelques travaux ont été consacrés à ce problème [7][14]. Ces techniques permettent d'ajuster aux observations des modèles de type polyédrique plus ou moins détaillés. Nous avons décidé également de nous appuyer sur une approche 3D dans laquelle les véhicules sont modélisés

par un parallélépipède. Cette modélisation 3D simplifiée permet de tolérer des variations angulaires de prise de vue et d'accéder directement aux paramètres dimensionnels de l'objet que nous appelons « gabarit ». Afin de faciliter la phase d'ajustement du modèle aux observations, nous considérons que le parallélépipède est orienté selon la route. Le critère d'ajustement global est une minimisation de l'écart entre la surface de projection d'un modèle parallélépipédique avec celle du masque de l'objet. Cet ajustement flexible permet de tolérer une partie des imperfections issues de la phase de détection d'objets. Les descripteurs de gabarit retenus sont la longueur, la largeur, et la hauteur réelle du parallélépipède.

Pour des raisons de facilité de mise en œuvre, le calibrage du système de mesure est obtenu par une technique géométrique dédiée, développée dans les systèmes d'analyse de trafic routier [4] ou piéton [31]. L'utilisateur doit simplement pointer les coins du rectangle et fournir les dimensions réelles de ce dernier. Ensuite l'algorithme génère une grille de calibration à l'intérieur du rectangle en enlevant les effets de perspective (voir figure 2). Dans notre contexte, le rectangle est dirigé vers la direction de la route afin de positionner deux points de fuite sur l'image. Ces deux points sont ensuite utilisés dans l'ajustement du modèle parallélépipédique. La grille est construite de manière récursive en utilisant les deux propriétés géométriques suivantes :

- L'intersection de la projection des diagonales du rectangle correspond à la projection  $I$  de son centre.
- Les droites passant par le point  $I$  et les points de fuite sur l'image passent par les milieux  $m_i$  des côtés du rectangle.

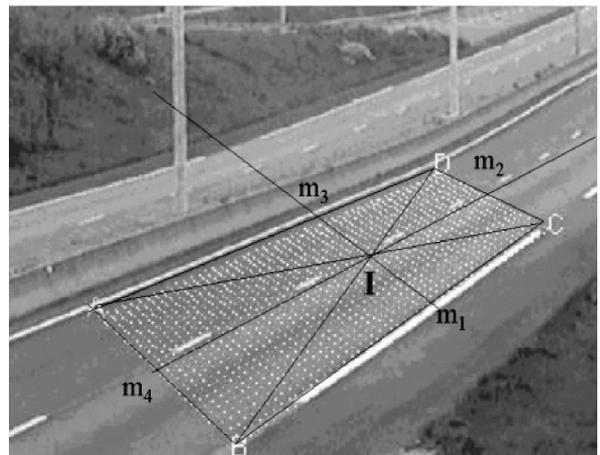


Figure 2. – Grille de calibration.

À chaque itération, l'algorithme positionne cinq nouveaux points rendant disponibles, de manière récursive, 4 nouveaux rectangles intermédiaires. Le processus continue et s'arrête en fonction de la résolution définie par l'utilisateur. Il apparaît clairement que la qualité de cette calibration dépend de la précision du positionnement des sommets du rectangle initial par l'opérateur.

Les dimensions réelles des segments au sol du parallélépipède sont déduites directement à partir de la grille de calibration. La hauteur dans l'image de l'objet est obtenue à partir d'un segment vertical du parallélépipède. Malheureusement, ce segment n'appartient pas au sol et la grille de calibration ne peut être exploitée directement. Un facteur d'échelle approché, calculé à partir d'un segment au sol situé au pied de segment vertical, est alors utilisé. Cette approximation se justifie lorsque la hauteur des objets par rapport à la distance caméra-objet est faible. Dans un cadre plus général, l'ajustement de modèles 3D quelconques ne peut être envisagé qu'avec des outils classiques de projection perspective [13]. La figure 3 illustre quelques résultats d'ajustement du modèle parallélépipédique sur des observations réelles.

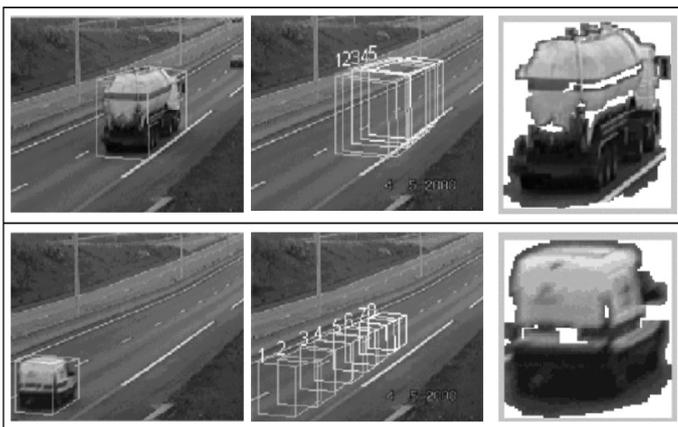


Figure 3. – Exemples de détection et de modélisation parallélépipédique.

### 3.2. mesure de compatibilité en couleur

La comparaison d'histogrammes de couleur est un moyen très utilisé pour reconnaître des objets par leur apparence [28]. Cette technique a été développée principalement dans le domaine de l'indexation d'images. Elle est particulièrement adaptée pour tolérer des changements d'échelle, des changements d'orientation, des occultations partielles mais également de faibles déformations des objets. En effet, dans toutes ces situations de déformations géométriques, les distributions de couleurs sont proches et les histogrammes associés restent stables.

Cependant, les techniques basées sur les histogrammes présentent une grande sensibilité aux variations d'éclairage. Une solution pour atténuer l'influence des conditions d'éclairage consiste à rechercher des informations invariantes par rapport aux modèles d'illumination. Les modèles les plus couramment utilisés sont basés sur des transformations linéaires. On trouve en particulier la modélisation scalaire utilisée dans l'espace RVB normalisé et la modélisation diagonale qui permet de rendre indépendant les canaux de couleurs. Ces techniques fonctionnent en qualité d'invariants par rapport à la composante lumi-

nance et opèrent dans un espace de chrominance. Nous avons utilisé la normalisation par l'énergie connue sous le nom « Grey-world » qui permet de préserver les trois dimensions des données (RVB) initiales [10]. Cette technique utilise une modélisation diagonale et donne de bons résultats en configuration multicaméras [32]. Par contre, elle requiert que les variations soient globales sur l'image. La normalisation par l'énergie se présente sous la forme :

$$r'_i = n \cdot \frac{r_i}{\sum_j r_j}, g'_i = n \cdot \frac{v_i}{\sum_j v_j}, b'_i = n \cdot \frac{b_i}{\sum_j b_j} \quad (2)$$

Notation :

$P_i : (r_i, v_i, b_i)^t$  est la couleur d'un pixel,

$P'_i : (r'_i, v'_i, b'_i)^t$  représente la couleur d'un pixel  $i$  normalisée par l'énergie,

$n$  : est le nombre de pixels dans l'image.

Afin d'évaluer la compatibilité des histogrammes des objets, nous utilisons la distance basée sur le test statistique de  $\chi^2$  qui est largement répandu, et en particulier dans les systèmes d'indexation d'images.

$$\chi^2(H^1, H^2) = \sum_i \frac{(H^1(i) - H^2(i))^2}{H^1(i) + H^2(i)} \quad (3)$$

### 3.3. combinaison de mesures de compatibilité

La distance entre les descripteurs d'objets ne peut représenter directement la notion de compatibilité entre objets. Par exemple, deux objets peuvent présenter une distance d'histogramme très grande tout en étant compatibles si les incertitudes de leurs mesures sont importantes. Afin de rendre cohérente la notion de compatibilité des objets à partir de la distance entre descripteurs, il est essentiel de recalculer les informations dans un espace commun de décision.

Les mesures de gabarit, intègrent des imprécisions issues :

- de l'étape d'extraction du masque des objets mobiles,
- de l'ajustement du modèle parallélépipédique,
- du processus de calibration.

Chaque mesure de gabarit est modélisée par une quantité floue de forme triangulaire considérée comme une distribution de possibilité. Elle constitue une approximation d'une loi gaussienne centrée sur la mesure [12]. Ce type de modélisation permet d'estimer une compatibilité entre deux mesures en prenant en compte leurs imprécisions. La distribution de possibilité triangulaire possède un noyau centré sur la valeur de la mesure brute et un

support fixe représentant l'imprécision de la mesure. L'imprécision est obtenue à partir d'un ensemble d'échantillons (supérieur à 40) de véhicules en estimant l'écart type entre les mesures brutes d'objets et leurs mesures réelles (données constructeur). Dans notre configuration expérimentale, l'imprécision sur la longueur est la plus importante (voir tableau 1).

Tableau 1. – Support d'imprécision des gabarits.

Compatibilité en gabarit	Support d'imprécision
Compatibilité en largeur	1,2 m
Compatibilité en longueur	1,8 m
Compatibilité en hauteur	1,2 m

Pour chaque descripteur de gabarit, le degré de compatibilité, entre deux candidats à la mise en correspondance, est basé sur un indice de concordance de mesures floues. Différents indices sont disponibles dans la littérature. L'indice possibiliste proposé par Dubois et Prade [8] est le plus couramment utilisé. Il est bien adapté pour des distributions de possibilités simples et symétriques de type triangulaire ou trapézoïdal. Pour deux mesures floues  $A$  et  $B$ , il correspond à la hauteur  $h$  de l'intersection :

$$h(A, B) = \sup_{\omega \in \Omega} \min(A(\omega), B(\omega)) \quad (4)$$

La distance d'histogrammes intègre principalement des imprécisions issues de l'extraction des masques des objets. Le recalage de la distance d'histogrammes est réalisé par une transformation permettant d'approximer ces incertitudes de mesures (voir figure 4). La présence d'une distance à l'intérieur de la zone d'imprécision, matérialisée par un support  $\Delta$  permet de garantir la notion de compatibilité des objets par leurs histogrammes. La largeur du support d'imprécision est obtenue, sous forme statistique, lors d'une phase préliminaire à partir de l'écart type des distances d'histogrammes des objets par rapport à eux-mêmes

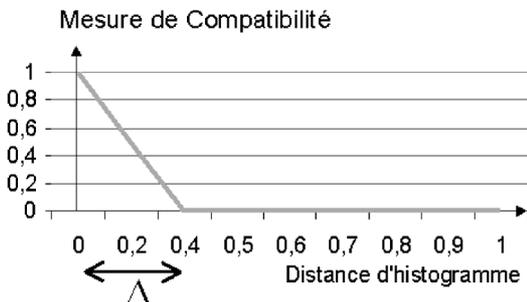


Figure 4. – Mesure de compatibilité des histogrammes.

mais prises à des instants et lieux différents. D'un point de vue expérimental, nous avons utilisé une banque de 40 véhicules détectés sur une autoroute par deux caméras distantes de 3 km. Afin de résumer la notion de compatibilité entre objet, une mesure de compatibilité globale est calculée en combinant les mesures de compatibilité associées aux descripteurs de gabarit et de couleur. Dans notre situation, les informations à combiner sont complémentaires et intègrent les imprécisions de mesures. La modélisation de l'incertitude associée a permis d'englober, de manière large, l'ensemble des candidats compatibles pour chaque descripteur. Il nous apparaît alors légitime de mettre en place à ce stade, un opérateur d'agrégation sévère de type conjonctif, permettant ainsi de limiter les candidats à la mise en correspondance. Nous avons choisi l'opérateur *min*.

## 4. les contraintes temporelles

Ces contraintes permettent d'une part de définir la fenêtre temporelle de recherche d'un événement attendu mais surtout de gérer la durée de vie de chaque objet manipulé par le système. Pour prédire le déplacement des objets mobiles dans les zones aveugles, nous utilisons des intervalles temporels flous : Dop (*Domain Occurrence Possibility*) [9][31].

### 4.1. modélisation cinématique

Soit la Dop ( $O_i, C^j, C^k$ ), elle représente un intervalle temporel flou généré par le module vision (l'expert) associé au capteur  $C^j$ . Elle explique la possibilité d'apparition au cours du temps de l'objet  $O_i$  dans le champ d'action d'un capteur spécifique  $C^k$ . Une Dop est représentée au moyen d'une distribution de possibilité. La largeur du support de la Dop exprime l'imprécision de la prédiction associée à l'objet mobile. Cette imprécision dépend fortement de la qualité des connaissances issues du contexte et des caractéristiques dynamiques de l'objet suivi.

La Dop ( $O_i, C^j, C^k$ ) associée à l'objet mobile  $i$  entre le capteur  $j$  et  $k$ , est calculée en se basant sur :

- la vitesse mesurée  $v_0$  au départ du capteur  $j$
- un modèle d'accélération appelé « **A** » entre le capteur  $j$  et  $k$
- et la distance  $d$  séparant les capteurs  $j$  et  $k$ .

La prédiction utilise un modèle de déplacement défini par :

$$d = v_0 \cdot t + \frac{A \cdot t^2}{2} \quad (5)$$

$$\Rightarrow t = \frac{-v_0 + \sqrt{v_0^2 + 2 \cdot A \cdot d}}{A} \quad (6)$$

La mesure de la vitesse est estimée en tenant compte de la moyenne des vitesses inter-image lors du suivi du centre de gravité de l'objet. La calibration de la mesure exploite directement la grille de calibration utilisée pour l'estimation des dimensions des gabarits. Cette mesure est modélisée par une distribution de possibilité triangulaire centrée sur la moyenne et une base reflétant une dispersion  $\sigma_v$  fixe. L'estimation de la moyenne de la vitesse de chaque objet s'effectue sur une dizaine d'images si le nombre d'observations le permet. D'un point de vue expérimental, la dispersion  $\sigma_v$  retenue est de l'ordre de 10 km/h.

Afin d'avoir une approche générique, nous prenons comme hypothèse qu'entre deux capteurs, les objets suivent uniquement des modèles d'accélération associés à leurs classes. Les modèles d'accélération des classes sont directement liés à l'infrastructure (type de route, dénivellement, limitation de vitesse ...), mais dépendent également de l'évolution dynamique du trafic (ralentissement, bouchon ...) entre deux capteurs. Le modèle d'accélération est représenté par un intervalle flou trapézoïdal au sens d'une distribution de possibilité normalisée.

Les informations utiles à la phase de prédiction (vitesse, accélération) sont supposées imparfaites. Il est donc nécessaire de disposer d'opérateurs mathématiques pouvant les manipuler en propageant efficacement l'ensemble des incertitudes sur les décisions. Dans le cas d'utilisation de sous-ensembles flous, il existe une solution générale basée sur le principe d'extension [33]. Ce principe constitue une des bases de la théorie des sous-ensembles flous. L'idée est d'étendre naturellement l'ensemble des concepts des mathématiques classiques aux cas des quantités floues [15]. Nous avons développé, à partir de la formule 6, un estimateur flou de la date d'arrivée en exploitant directement l'arithmétique binaire floue et le principe d'extension pour la fonction « racine carrée ».

## 4.2. classification des objets

La classification pour les besoins de modélisation des accélérations, est effectuée premièrement sur le critère « gabarit des objets ». Les différentes classes modélisées sont : les poids lourds, les utilitaires, les véhicules légers et les deux roues. La classification est réalisée à partir de la valeur du volume englobant ces véhicules. Ce volume est calculé à partir des trois mesures de gabarit et est représenté par un nombre flou triangulaire.

Cette classification est rendue plus fine en se basant sur le comportement cinématique des objets pendant les instants d'observations. Ce deuxième critère de classification utilise la vitesse instantanée des véhicules. Il permet ainsi une subdivision des classes préliminaires (gabarit) en 3 autres sous classes liées à la vitesse (lente, normale et rapide). Cette augmentation du nombre de modèles d'accélération disponibles par le système permet d'obtenir une meilleure représentativité des comportements des objets de la scène.

Une approche classique de classification floue est mise en place afin d'intégrer les imprécisions des classes mais également de pouvoir gérer les imperfections des mesures issues des observations (vitesse, gabarit). Les classes sont construites par un expert à partir de données contextuelles (fiches constructeur, type de voie). Les degrés d'appartenance d'une observation aux classes sont évalués à partir de la mesure de concordance possibiliste.

Dans un contexte de classification floue d'objet, un objet peut appartenir de manière nuancée à plusieurs classes. Il est alors intéressant d'utiliser des modèles d'accélération intermédiaires en fusionnant plusieurs modèles. Afin de ne pas élargir les supports des Dops, les caractéristiques (centre  $C$ , support  $S$ , noyau  $N$ ) du modèle d'accélération résultant sont évaluées sous la forme de barycentres des caractéristiques des modèles concernés, pondérées par les degrés d'appartenance de l'observation obtenus lors de la classification.

## 4.3. la modélisation et la mise à jour des accélérations

Nous rappelons que ces modèles d'accélération associés à chaque classe d'objet sont estimés pour chaque couple de capteurs voisins. Ils sont réglés au départ par l'expert humain en analysant les informations contextuelles disponibles. Les informations utiles sont : la distance, le dénivellement et tous les autres paramètres contextuels pouvant intervenir sur le comportement des objets entre deux capteurs (limitation de vitesse, etc.). Ils représentent une initialisation des modèles d'accélération (moyenne  $\mathbf{a}$ , dispersion  $\sigma_a$ ). En conséquence, ces modèles devront être les plus larges possibles (au sens de l'imprécision). La modélisation trapézoïdale utilisée, est centrée sur la moyenne  $\mathbf{a}$  avec une base de  $3\sigma_a$  et un noyau à  $2\sigma_a$ .

Les modèles d'accélération évalués par l'expert de manière empirique donnent de bons résultats dans des environnements peu encombrés. Par contre, dès que le nombre d'objets mobiles augmente, les Dops associées à ces modèles peuvent devenir des informations considérées comme « faibles ».

Afin d'adapter et d'affiner les modèles d'accélération aux comportements réels des véhicules, un processus de mise à jour des modèles d'accélération pour chaque classe a été mis en place. Le processus prend en compte les  $n$  derniers représentants détectés de cette même classe et calcule la valeur moyenne et la variance de leurs accélérations. Ensuite, deux filtres de mise à jour de premier ordre sont utilisés pour réviser le modèle d'accélération. Dans le but de ne pas rendre trop sélectif le modèle d'accélération, l'algorithme de mise à jour maintient une dispersion minimale à  $\sigma_s$ .

$$a^{t+1} = (1 - \alpha) \cdot a^t + \alpha \cdot \sigma^t \quad (7)$$

$$\sigma_a^{t+1} = \max((1 - \beta) \cdot \sigma_a^t + \beta \sigma_0^t, \sigma_s) \quad (8)$$

*Notation :*

$a^{t+1}, \sigma_a^{t+1}$  : paramètre du modèle d'accélération courant,  
 $\sigma^t, \sigma_{ot}$  : la moyenne et la variance des accélérations des  $n$  dernières observations,  
 $\sigma_s$  : variance minimale,  
 $\alpha, \beta$  : coefficient d'actualisation.

## 5. la mise en correspondance

Nous avons choisi une organisation logique du système d'interprétation de type distribué et coopératif basée sur des sociétés d'experts. Pour chaque objet détecté, le système prévoit un ensemble possible de déplacements et reste en attente d'informations pour vérifier les hypothèses prévues. Chaque expert décide d'associer ou non l'observation à l'objet attendu dont il dépend. Cette décision utilise les compatibilités temporelles et visuelles de l'observation aux différents objets. La compatibilité temporelle correspond à la valeur numérique de la Dop à l'instant de l'observation. L'opérateur de combinaison des deux mesures de compatibilité (visuelle et temporelle) est une conjonction (opérateur min). Ce choix sévère reprend le même raisonnement présenté dans le calcul de la compatibilité visuelle globale. La conjonction permet alors de sélectionner les seuls objets vérifiant la double contrainte. Malheureusement même avec une réduction significative des candidats, la décision d'association demeure difficile. Nous présentons ici les stratégies associées aux trois phases de base du système de suivi : l'initialisation, la maintenance et la terminaison de pistes.

### 5.1. stratégies d'initialisation et de terminaison de pistes

En fonction de la couverture du réseau de capteurs et de l'organisation de la scène, il n'existe généralement pas de bijection entre l'ensemble des objets attendus et l'ensemble des observations au niveau d'un capteur.

Il est important que la création des nouveaux objets puisse être prévue dans le système d'interprétation afin d'initialiser leurs pistes. Ces objets doivent être identifiés dès qu'ils apparaissent devant un des capteurs. Dans notre système, cette décision est prise lorsque l'observation ne présente aucune compatibilité avec les objets attendus. Cette notion d'objet nouveau notée '\*' ou objet non prévu apparaît dans la théorie de l'évidence sous le nom du « monde ouvert » [27]. Il exploite la notion de transfert des croyances vers l'hypothèse de l'objet non attendu. Nous avons également utilisé une approche similaire sous la forme suivante :

$$\forall O_i \in \Phi, C_j(*) = 1 - \max_{O_i \in \Phi} C_j(O_i) \quad (9)$$

*Notation :*

$\Phi$  : l'ensemble des objets attendus,  
 $C_j(O_i)$  : la compatibilité de l'observation avec l'objet  $i$  au niveau du capteur  $j$ .

Cependant, cette méthode de création de nouveaux objets suppose que ces derniers ne doivent pas apparaître pendant les périodes où des objets similaires peuvent être attendus.

La terminaison de piste est déclarée en local pour les objets qui ne respectent pas leurs contraintes temporelles. Ces objets peuvent correspondre, soit à ceux qui n'ont pas été détectés lors de leur passage devant un capteur lorsqu'ils étaient attendus, soit à ceux qui terminent réellement leurs pistes. Le non-respect des contraintes temporelles ou les non-détections induisent soit des pertes simples, soit des pertes suivies de créations de nouveaux objets. L'organisation temps réel et distribuée du système d'interprétation n'est malheureusement pas adaptée pour gérer ces cas particuliers directement au niveau local. Les décisions de terminaisons sont prises par les derniers experts et sont remontées à l'unité de surveillance globale. Cette dernière aura la possibilité de corriger, plus facilement, une partie des erreurs de terminaison prématurée en comparant visuellement les objets nouveaux à ceux dont les pistes viennent d'être terminer.

### 5.2. stratégies de maintien de pistes en présence d'ambiguïtés

Les ambiguïtés, liées à la phase de maintien de pistes, sont en grande partie liées à la quantité d'objets circulant en parallèle mais aussi au placement géographique des caméras. En effet, plus la distance entre les capteurs est grande, plus les prédictions deviennent larges temporellement. Ainsi, plusieurs sociétés peuvent opérer en parallèle sur des objets ressemblants.

Dans notre contexte dynamique où les informations évoluent de manière incrémentale, la stratégie envisagée est de type prudent et consiste à ne pas prendre de décision de manière systématique. La fusion temporelle, dans un système de perception, consiste à combiner des informations au cours du temps par un ou plusieurs capteurs, puis à apporter ou à améliorer une décision. Cette approche permet de repousser les décisions peu crédibles en attendant d'autres informations complémentaires. Les nouvelles informations ne garantissent pas toujours des levées d'ambiguïtés, mais assurent une exploitation optimale des informations susceptibles de conforter les décisions. Ce type de stratégie d'attente va malheureusement à l'encontre de la rapidité de réaction d'un système temps réel devant faire face à un flot d'objets important. Une solution intermédiaire consiste alors à déclencher le mécanisme de maintien d'hypothèses uniquement en cas de besoin et de manière contrôlée dans le temps [32].

La fusion temporelle a été utilisée avec succès en reconnaissance et en classification dans des environnements évolutifs. Elle permet de réviser les confiances des décisions intermédiaires par

l'arrivée de nouvelles informations qui peuvent être de natures hétérogènes et/ou être disponibles de manière asynchrone [6][21].

La figure 5 représente un ensemble de situations faisant intervenir deux objets similaires ayant des possibilités de dates d'arrivée communes, c'est-à-dire que leurs Dops présentent une intersection non nulle sur les capteurs aval.

Au sein du premier scénario, même s'il existe une intersection entre les Dops des objets, les observations sont arrivées à des dates non ambiguës. Dans le deuxième scénario, la première observation est associée de manière instantanée au premier objet (Dop inférieure), par conséquent la deuxième observation n'est plus en situation critique.

Nous proposons de gérer en partie les problèmes d'associations par la reconnaissance de scénarios de type 3 et 4 à l'aide d'une approche de fusion temporelle. Dans le cas du scénario de type 3, le système est capable de résoudre le problème d'association après la détection de la deuxième observation. Dans le cas de la reconnaissance du scénario de type 4, après la deuxième observation, le système est capable seulement de décider qu'il ne pourra pas résoudre le problème. Cependant, dans tous les scénarios, le système réagit rapidement, ses décisions sont prises dès que les observations utiles sont détectées sans attendre la fin de leurs Dops.

### 5.3. décision de mise en correspondance

L'ensemble des degrés de compatibilité de l'observation avec les objets attendus, y compris les objets nouveaux, constitue une distribution de préférence liée à la décision d'association. Afin de gérer de manière quantitative la notion d'ambiguïté, il est nécessaire d'associer un degré de confiance à chaque association. Nous avons choisi de continuer à exploiter le cadre de la théorie des possibilités pour gérer ce niveau d'incertitude. Nous noterons par la suite  $\Phi^*$  l'ensemble  $\Phi$  des objets attendus muni du singleton  $\{*\}$  (nouveau objet). Cette distribution est ensuite normalisée pour obtenir une distribution de possibilité consistante. À chaque observation, cette distribution rassemble les avis des différents experts, opérant de manière parallèle, sur le module de vision ayant détecté cette observation. À partir de cette distribution, chaque expert évalue la nécessité d'une hypothèse d'association en tenant compte de l'avis des autres experts.

$$N(H_i) = 1 - P(\overline{H}_i) \tag{10}$$

Notation :

$$\overline{H}_i = \Phi^* - \{H_i\}$$

$$\Phi^* = \Phi \cup \{*\}$$

$$P(\overline{H}_i) = \max_{H_j \in \overline{H}_i} P(H_j)$$

Scénarios $\square$	Décisions d'association $\square$
<p><b>Scénario 1</b></p>	<p>*L'observation 1 est associée à l'objet 1 sans ambiguïté. <math>\uparrow</math></p> <p>*L'observation 2 est associée à l'objet 2 sans ambiguïté. <math>\square</math></p>
<p><b>Scénario 2</b></p>	<p>*L'observation 1 est associée à l'objet 1 sans ambiguïté. <math>\uparrow</math></p> <p>*L'observation 2 est associée à l'objet 2 car la Dop 1 est déjà supprimée au moment de la décision. <math>\square</math></p>
<p><b>Scénario 3</b></p>	<p>*L'observation 1 est ambiguë. <math>\uparrow</math></p> <p>*L'observation 2 est associée à l'objet 2 sans ambiguïté; l'association de l'observation 2 permet de lever l'ambiguïté sur l'observation 1 par fusion temporelle. <math>\square</math></p>
<p><b>Scénario 4</b></p>	<p>*L'observation 1 est ambiguë. <math>\uparrow</math></p> <p>*L'observation 2 est également ambiguë; après cette deuxième observation le système décide du maintien des ambiguïtés. <math>\square</math></p>

Figure 5. – Scénarios en présence de deux objets proches.

Chaque expert observe alors la nécessité d'association de chaque observation avec les objets attendus. Si l'association présente un degré de nécessité important alors l'expert valide l'association. En revanche, dans le cas contraire, la notion d'ambiguïté est prononcée. Un processus de fusion temporelle est alors enclenché par une société d'experts de niveau supérieur (niveau II). L'expert ayant détecté l'ambiguïté crée une société de niveau II qui intègre l'ensemble des sociétés initiales concernées en contactant les experts responsables de leur création. Les décisions des experts initiaux sont bloquées et deviennent tributaires des décisions de niveau II.

En pratique, ce nombre d'objets en interaction peut être important. En effet, plusieurs scénarios d'ambiguïtés peuvent se combiner de manière récursive générant une augmentation importante de données à fusionner. Il est alors indispensable d'utiliser des outils permettant de structurer efficacement la fusion. Nous nous sommes alors orientés vers les arbres de type MHT (Multiple Hypothesis Testing) utilisés au départ par Singer et Reid [25] en poursuite radar. Ces arbres d'hypothèses sont développés dynamiquement en fonction des observations.

Dans notre cas, les hypothèses constituent une configuration d'associations entre les observations et les objets attendus. L'arbre MHT propage, de manière récursive, les branches représentant les hypothèses. Le but de la fusion temporelle est alors de mettre en valeur à partir de l'expansion de l'arbre, les hypothèses les plus favorables. La viabilité d'une structure de type MHT impose une gestion optimale du nombre d'hypothèses. Pour cela, nous intégrons donc uniquement des associations « fortes » en terme de degré de compatibilité globale. Nous rappelons que cette notion de compatibilité globale est contrôlée temporellement par l'intermédiaire des Dops, ce qui permet de limiter de manière dynamique le nombre d'hypothèses au sein de l'arbre. À chaque observation, la fusion temporelle doit estimer la qualité des différentes hypothèses. Nous utilisons pour cela la même technique employée par les experts initiaux en local. Cette mesure de qualité est obtenue en calculant les degrés de nécessité de chaque hypothèse de l'arbre.

Le degré de possibilité d'une hypothèse est basé sur une notion de degré de confiance calculé à partir d'une combinaison des degrés de compatibilité des associations élémentaires. L'opérateur de combinaison que nous avons choisi est la multiplication qui permet d'imposer un comportement de type conjonctif, tout en restant moins sévère que le min. Ensuite, les résultats de ces combinaisons sont rassemblés afin d'établir, après normalisation, une distribution de possibilité des hypothèses.

Nous rappelons que la mesure de nécessité représente l'écart relatif entre la possibilité d'une hypothèse vis-à-vis des autres hypothèses concurrentes générées par l'arbre. Si le degré de nécessité de l'une des branches est considéré suffisant, les associations de celle-ci sont validées et l'arbre est détruit. Afin de maîtriser l'expansion de l'arbre, le système décide d'arrêter ce dernier prématurément lorsque le nombre d'objets ambigus devient trop important (supérieur à 5 objets).

L'exemple didactique de la figure 6 illustre le comportement de la fusion temporelle en présence d'un scénario de type 3 concernant deux véhicules similaires (O1,O2). L'apparition d'une première observation (Obs1) a engendré une ambiguïté devant le capteur C1. Une deuxième observation (Obs2) sur le capteur C2 a permis de lever le doute (voir tableau 2).

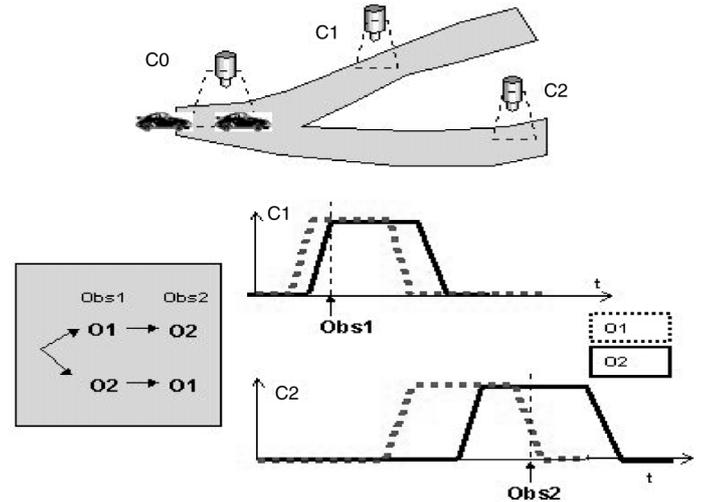


Figure 6. – Exemple de résolution de scénario de type 3.

Tableau 2. – Arbre MHT.

Numéro des hypothèses	Observation 1 sur capteur C1 (compatibilité)	Observation 2 sur capteur C2 (compatibilité)	Possibilité de l'hypothèse	Nécessité de l'hypothèse
1.1	O1(1,0)		1	0,1
1.2	O2(0,9)		0.9	0
2.1	O1(1,0)	O2(1,0)	1	<b>0.73</b>
2.2	O2(0,9)	O1(0,3)	0.27	0

## 6. simulation

Afin de tester le comportement des algorithmes d'association sur des situations complexes, nous avons développé un simulateur de trafic permettant de générer des données multi-capteurs. Nous présentons ici une configuration sur autoroute constituée par quatre modules de vision (voir figure 7).

Une séquence particulière de véhicules est générée à partir du capteur amont (caméra C1). Le nombre de véhicules dans cette séquence augmente progressivement et redescend en fin de séquence (voir figure 8.a). Cette variation du trafic permet de mettre en évidence le comportement des algorithmes d'associations en fonction du volume de trafic. Dans cette expérimenta-

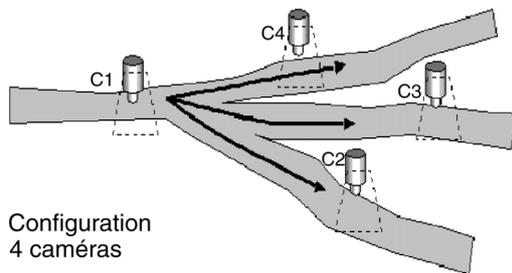


Figure 7. – Configuration des capteurs.

tion, la composition du trafic est basée sur la génération aléatoire de 5 formes visuelles distinctes. Un seul modèle d'accélération est utilisé. Les distances entre la caméra amont et les caméras aval sont respectivement de 2 kms, 2.5 kms et 3 kms. La vitesse moyenne de déplacements de chaque véhicule au départ du capteur amont est représentée par la figure 8.b. Le profil de ces vitesses tient compte des modèles de déplacement de véhicules sur des voies rapides [18]. La courbe de vitesse diminue avec l'accroissement du nombre de véhicules. Les dates d'apparition des observations sont générées de manière aléatoire à l'intérieur des Dops sur les capteurs aval, selon une loi gaussienne centrée sur la Dop. La largeur de support du modèle d'accélération ( $3\sigma_s$ ) a été réglée à  $0.1 \text{ m/s}^2$ . Les véhicules sont équi-répar-

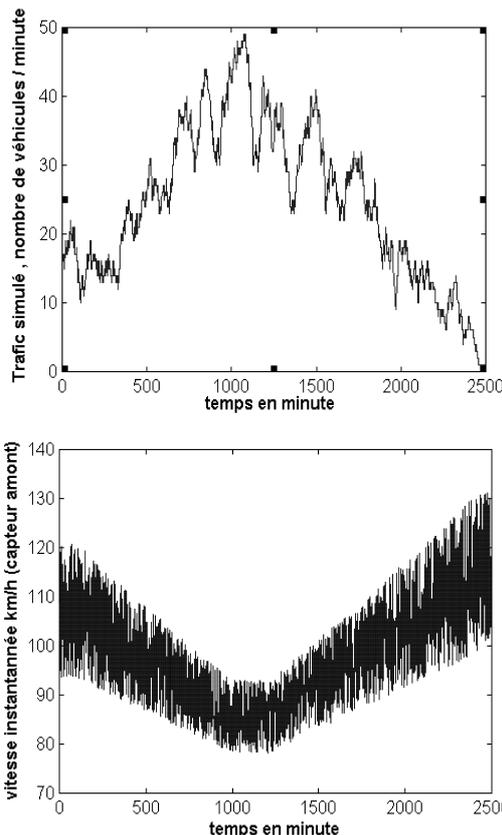


Figure 8a, 8b. – Trafic et vitesse des véhicules au départ du capteur amont C1.

tis sur les trois capteurs en aval. La détection de l'ambiguïté en local est obtenue par une mesure de nécessité inférieure à 0.2. En revanche, la terminaison de l'arbre MHT est déclenchée par une mesure de nécessité de branche supérieure à 0.3.

La figure 9 présente à partir d'un trafic simulé (courbe en pointillée), le nombre d'objets ambigus associés et le nombre de non-décisions d'association par la fusion temporelle (algorithme A). Ces indicateurs sont comparés aux performances d'un algorithme d'association simple (algorithme B). Cet algorithme associe chaque observation au meilleur candidat, de manière instantanée, en tenant compte des mesures de compatibilité globale.

La figure 10 résume les résultats d'associations en trois types d'indicateur (en ordonnée) par rapport au nombre initial de véhi-

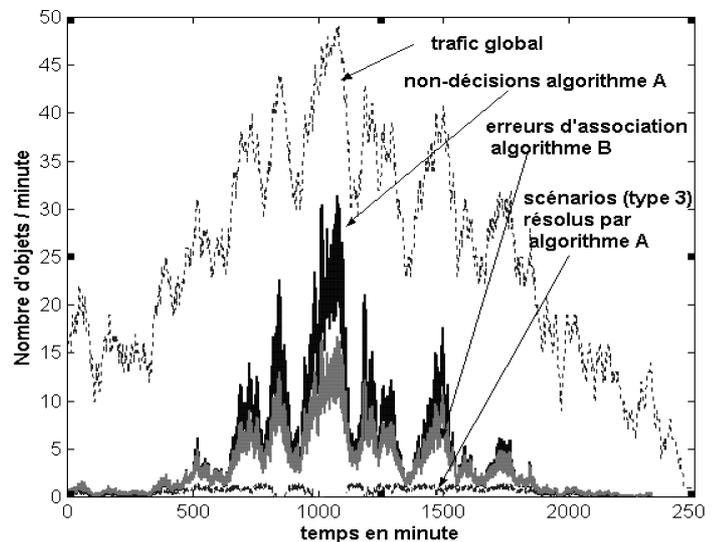


Figure 9. – Résultats d'association algorithme A (fusion temporelle) et B (meilleur candidat instantané).

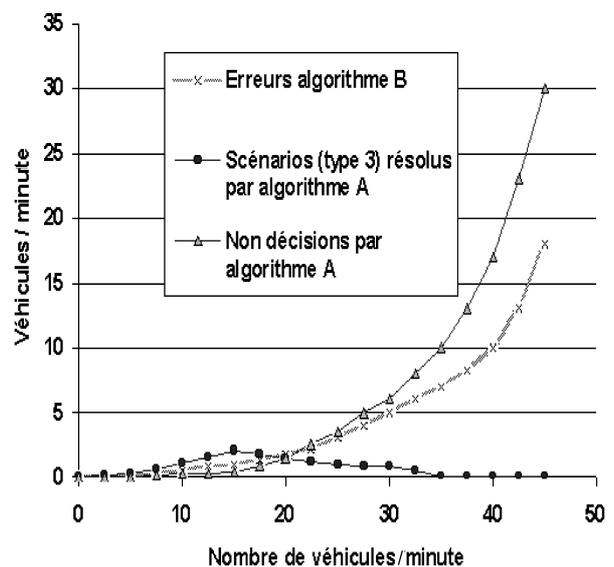


Figure 10. – Évaluation des performances des algorithmes A et B.

cules détectés par le capteur amont (en abscisse). Les indicateurs sont respectivement les erreurs de l'algorithme B, le nombre de situations de type 3 résolues par l'algorithme A et enfin le nombre d'objets non associés par l'algorithme A. L'algorithme A, basé sur la fusion temporelle, permet de résoudre la majorité des ambiguïtés lors de trafic modéré (inférieur à 20 véhicules / minute). Dans des situations de trafic élevé, les ambiguïtés ont tendance à s'accumuler en générant en continu des scénarios de type 4 qui sont détectés mais non-résolus par l'algorithme A. Ainsi, nous pouvons remarquer qu'en présence de trafic dense, la stratégie prudente de la gestion de l'incertitude du système n'est plus adaptée, car elle entraîne un taux de non-décisions plus important que le taux d'erreurs de l'algorithme B.

## 7. conclusion

Nous avons proposé dans ce travail une architecture logique d'un système multi-capteurs de suivi d'objet opérant sur des scènes vastes. Son principal rôle est la ré-identification des objets évoluant dans une scène étendue à partir de modules de vision autonomes. La solution proposée est basée sur l'intégration conjointe de mécanismes de raisonnement distribué et temporel et d'outils de gestion de l'incertitude.

Le raisonnement distribué est basé sur une distribution des traitements bas niveau sur les modules de vision autonomes. Une structure logique de coopération basée sur la notion de société d'experts permet ensuite de réaliser les tâches de suivi global.

Les contraintes temporelles sur les déplacements d'objets constituent les informations de base du raisonnement temporel car elles permettent de donner une durée de vie aux différents événements manipulés par le système. Ces contraintes basées sur des modèles d'accélération des classes d'objets sont mises à jour en fonction du comportement observé de la scène. Cet aspect apporte un niveau d'adaptabilité du système en fonction de son environnement.

La gestion de l'incertitude est une fonction importante du système. Elle est présente dans le calcul des mesures de comptabilité visuelles, dans la classification des objets, dans la prédiction des objets, et enfin dans la phase décisionnelle d'association. Cette gestion de l'incertitude permet au système de surveillance de valider uniquement les associations dont les risques d'erreurs sont faibles.

Dans une application donnée, la configuration spatiale des capteurs (le nombre et le placement) est un paramètre de réglage sensible car elle influence directement la précision des contraintes temporelles. Il serait alors intéressant de développer des outils qui permettent de proposer, pour une application de surveillance donnée, une configuration optimale en tenant compte des contraintes économiques et de la qualité minimale des décisions recherchées. Une autre extension de ce travail peut

être l'étude de l'influence de la fiabilité (ou perte de capteurs) dans les systèmes de surveillance. Une stratégie de reconfiguration automatique de l'architecture logique du système peut être mise en œuvre afin de maintenir une notion de qualité de service minimale. Elle permettrait alors une adaptation du raisonnement en fonction des ressources évolutives du système d'interprétation.

Un autre niveau d'amélioration consiste en l'apport de capteurs hétérogènes comme les boucles magnétiques ou les radars, qui permettent de proposer des informations redondantes et complémentaires durant l'interprétation de jour et permettent surtout d'étendre les fonctionnalités du système pendant la nuit ou en présence de situations de visibilité défavorables (brouillard, pluie ...).

## BIBLIOGRAPHIE

- [1] J.M. Blosseville, S. Bouzar, F. Lenoir, R. Glachet, « Traitement d'image pour la mesure et la surveillance du trafic : une avancée significative », Revue TEC, N°126, pp. 29-34, 1994.
- [2] F. Brémond, « Environnement de résolution de problèmes pour l'interprétation de séquences d'images », Thèse. Université de Nice-Sophia Antipolis. Octobre 1997.
- [3] B. Coifman, « Vehicule Reidentification and travel time measurement in real time on freeways using the existing loop detectors infrastructure », Transportation Research, Record 1643, Transportation research board, 1998, pp181,191.
- [4] B. Coifman, D. Beymer, P. McLauchlan, and J. Malik, « A real-time computer vision system for vehicle tracking and traffic surveillance », Transportation Research: Part C, vol 6, no 4, 1998, pp. 271-288.
- [5] Y. Cui, Q. Huang, « Character extraction of licence-plates from video », Proc IEEE Computer society Conference on Computer vision and Pattern recognition IEEE , pp. 502-507, 1997.
- [6] M. Delplanque, « L'application des techniques modernes de traitement de l'information pour le guidage de véhicules sous-marins. » Thèse de Doctorat Universiyé de Lille 1, 1998.
- [7] L. Dresschler, H.H. Nagel, « Volumetric model and 3D trajectory of a moving car derived from monocular TV frame sequences of a street scene », CVGIP vol 20, pp. 199-228, 1982.
- [8] D. Dubois et H. Prade, « Théorie des Possibilités, Applications à la représentation des connaissances en informatique », Ed. Masson, Paris, 1988.
- [9] D. Dubois, H. Prade , « Processing fuzzy temporal knowledge », IEEE transactions on systems man and cybernetics, vol 19, n°4, pp. 729-744, 1989.
- [10] G.D. Finlayson, B. Schiele, J.L. Crowley, « Comprehensive colour image normalization », Proc. In Fifth European Conference on Computer Vision, Freiburg, Germany, June 1998.
- [11] M. Forthofer, S. Bouzar, F. Lenoir, J.M. Blosseville, D. Aubert, « Automatic incident detection : wrong way vehicle detection using image processing ». Proc. of Third Annual World Congress on Intelligent Transportation System, Octobre 1996.
- [12] « Guide to the Expression of Uncertainty in Measurement », Report ISO/TAC4/WG3, 1993.
- [13] R. Horaud, O. Monga, « Vision par ordinateur, outils fondamentaux », Editions Hermès, 1995.

- [14] D. Koller, K. Danilidis, H. Nagel, « Model based object tracking in monocular image sequence of road traffic scenes », *International Journal of Computer Vision*, 3(10), pp. 257-281, 1993.
- [15] A. Kaufmann, M.M. Gupta, « Introduction to fuzzy arithmetic », Van Nostrand Reinhold Ed., New York, 1991.
- [16] J. Larson, K. Katwyk, K. Liu *et al.*, « A real time laser-based prototype Detection system for measurement of delineations of moving vehicles ». UCB-ITS-PWP-98-20, Path, University of California, Berkeley, CA, 1998.
- [17] V. Lasserre, « Modelisation floue des incertitudes de mesures de capteurs », Thèse de Doctorat de l'Université de Savoie, 1999.
- [18] J.P. Lebacques, J.B. Lesort, « Les réseaux de transport et les services », Proc. 14th international symposium on transportation and traffic theory, Amsterdam, pp. 3-5, 1999.
- [19] T. Matsuyama, « Cooperative distributed vision : dynamic integration of visual perception, camera action, and network communication », 4th international Workshop on Cooperative Distributed Vision, March 22-24, Kyoto Japan, 2001.
- [20] C. Motamed, « Motion detection and tracking using belief indicators for video-surveillance applications », Proc. of First IEEE Workshop on Performance Evaluation in Tracking and Surveillance, PETS'2000, Grenoble, mars 2000.
- [21] A. Nifle, R. Reynaud, « Classification de comportement fondée sur l'occurrence d'événement en théorie des possibilités », *Traitement du Signal*, 14(5), pp. 523-533, 1997.
- [22] J. Palen, « The need for surveillance in intelligent transportation Systems », *Intelli-motion*, vol. n°1, University of California PATH, Berkeley CA, pp. 289-303, 1997.
- [23] B.S.Y. Rao, H.F. Durrant-Whyte, J.A. Sheen, « A fully decentralized multi-sensor system for tracking and surveillance », *The International Journal of Robotics Research*, vol. 12 n°1, february, 1993.
- [24] C.S. Regazzoni, C. Sacchi, G. Gera, « Intelligence distribution of a third generation people counting system transmitting information over an urban digital radio link », Proc. of 2nd European workshop on Advanced Video-based Surveillance Systems, Kingston, UK, pp. 53-69, Aug. 2001.
- [25] B. Reid, « An algorithm for tracking multiple targets », *IEEE Transactions on Automatic Control*, AC-24(6), pp. 843-854, 1979.
- [26] C. Schleiffer, R. Reynaud and G. Sella, « Fusion de données multi-capteurs géographiquement répartis, problème des données manquantes », 15<sup>ème</sup> colloque Grets, Juan-les-pins, septembre, 1995.
- [27] P. Smets, R. Kennes, « The transferable belief model », *Artificial Intelligence*, Vol. 66 n°2, pp. 191-234, 1994.
- [28] M.J. Swain and D.H. Ballard, « Color indexing », *Intern. Journal of Computer Vision*, 7(1), pp. 11-32, 1991.
- [29] P. Vannoorenberghe, « Détection du mouvement par analyse de séquences d'images monoculaires. Application à l'estimation du flux des piétons en milieu urbain », Thèse de Doctorat, Université du Littoral Côte d'Opale, 1997.
- [30] P. Vannoorenberghe, C. Motamed, J.-G. Postaire, « Réactualisation d'une image de référence pour la détection du mouvement dans les scènes urbaines », *Revue Traitement du signal*, Vol. 15 n°2, pp.139-148, 1998.
- [31] O. Wallart, C. Motamed, M. Benjelloun, « Temporal knowledge for cooperative distributed vision », 7th IEE Conference on Image Processing and its applications, Manchester, July 1999.
- [32] O. Wallart, « Contribution à l'interprétation de scène par une approche multi-capteurs », Thèse de Doctorat, Université du Littoral Côte d'Opale, 2002.
- [33] L.A. Zadeh, « Fuzzy sets as a basis for a Theory of Possibility », *Fuzzy Sets and Systems*, N°1, pp. 3-28, 1978.

*Manuscrit reçu le 28 mai 2001*

## LES AUTEURS

Cina MOTAMED



Cina MOTAMED est né en 1965. Il a obtenu un Doctorat en Génie Informatique de l'Université de technologie de Compiègne en 1992. Il mène des recherches au sein du laboratoire LASL de Calais sur les systèmes d'interprétation de séquences d'images, principalement dans le domaine de la vidéosurveillance.

Olivier WALLART



Olivier WALLART est né en 1973. Il a obtenu un D.E.A. en Informatique Industrielle de l'Université de Lille 1 en 1996 et un Doctorat en Génie Informatique, Automatique et Traitement du Signal de l'Université du Littoral en 2002. Son activité de recherche se consacre essentiellement aux problèmes de suivi d'objets mobiles en environnement extérieur.