Reconnaissance d'événements vidéos par l'analyse de trajectoires à l'aide de modèles de Markov

Video Trajectory-based Event Recognition using Hidden Markov Models

Alexandre Hervieu [1], Patrick Bouthemy [1] et Jean-Pierre Le Cadre [2]

[1] INRIA, Centre Rennes - Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France
alexandre.hervieu@inria.fr, Patrick, Bouthemy@inria.fr
[2] IRISA/CNRS, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France
lecadre@irisa.fr

Manuscrit reçu le

Résumé et mots clés

Nous présentons une méthode originale de classification de trajectoires dans des séquences vidéos pour la reconnaissance d'événements dynamiques. Les Modèles de Markov Cachés (MMC) sont utilisés afin de représenter chaque trajectoire et d'évaluer leurs similarités. Nous avons pu valider notre méthode en la comparant à plusieurs autres méthodes telles que la comparaison d'histogrammes, une méthode utilisant les Séparateurs à Vaste Marge (SVM) ainsi qu'une méthode de MMC utilisant des modélisations par mélanges de gaussiennes. Des descripteurs appropriés, invariants à la translation, à la rotation ainsi qu'au facteur d'échelle sont calculés sur les trajectoires, puis exploités dans une représentation par MMC. Une méthode statistique est également proposée pour le choix du nombre d'états pour la modélisation par MMC choisie. Nous avons testé notre méthode sur deux ensembles de trajectoires issues de vidéos (respectivement de Formule 1 et de ski) obtenues par une méthode de suivi dans des vidéos de sport.

Vision par ordinateur, Analyse de séquences d'images, Reconnaissance d'évènements, Modèles de Markov cachés.

Abstract and key words

We address the problem of dynamic event recognition in videos. This is motivated by increasing needs for content-based exploitation of video footage, as encouraged in numerous applications, e.g., retrieving video sequences in large TV archives, creating automatic video summarization of sport TV programs, or detecting specific actions or activities in video-surveillance. It implies to tackle the well-known semantic gap between computed low-level features and high-level concepts. Considering 2D trajectories is attractive since they form computable image features which capture elaborated spatio-temporal information on the viewed actions. Methods for tracking moving objects in an image sequence are now available to get reliable enough 2D trajectories in various situations. These trajectories are given as a set of consecutive positions (x, y) in the image plane over time. If they are embedded in an appropriate modeling framework, high-level information on the dynamic scene can then be reachable.

We aim at designing a general trajectory classification method that does not exploit strong *a priori* information on the scene structure, the camera set-up, the 3D object motions, while taking into account both the trajectory shape (geometrical information related to the type of motion and to variations in the motion direction) and the speed changes

of the moving object on its trajectory (dynamics-related information). Appropriate local differential features combining curvature and motion magnitude are defined and robustly computed on the motion trajectories.
Moreover, these features are not affected by the location of the trajectory in the image plane (invariance to translation), by its direction in the image plane (invariance to rotation) and by the distance of the viewed action to the camera (invariance to scale), and may allow comparison of trajectories from different cameras. A robust enough non-parametric feature extraction framework is also proposed since local differential features computed on the extracted trajectories are prone to be noise corrupted.
To efficiently process the invariant trajectory characterization, probabilistic networks, and more specifically hidden Markov models (HMM) are used since the inherent properties of this modeling help taking into account the temporal evolution of the spatio-temporal information contained in the trajectories. Classical HMM, relying on Gaussian mixture modes (GMM), are designed to model data of sufficient sizes. Hence they may fail treating small trajectories with only few dozens of observations. An original HMM modeling, based on a uniform quantization of the observation space and dealing efficiently with small trajectories, is proposed, and an efficient HMM state number selection is also developed.
To compare trajectories, a similarity measure is defined based on the Rabiner distance between HMM, and used to process the video event retrieval task.
By considering a feature having those invariances and characteristics (considering both the trajectory shapes and speed evolutions), we consider the trajectory as a dynamical pattern while other methods consider the trajectories as attached to the camera point of view. We have compared our approach with other methods to put forward the properties (spatio-temporal modeling and efficient processing of small trajectories) of the developed method. Methods relying on feature histogram comparisons, on HMM/GMM modeling and on support vector machine (SVM) were considered. A set of comparative experiments on real videos (especially Formula One and ski TV program) with classification ground truth has been conducted and showed that the proposed method supplies accurate results and offers better performances than other methods.

Computer vision, Image sequence analysis, Event recognition, Hidden Markov models.

1. Introduction

Le suivi d'objets dans des vidéos est désormais développé au point qu'il est possible d'obtenir des trajectoires fiables d'objets en mouvement dans des situations variées. Les travaux concernant l'analyse de ces trajectoires sont de plus en plus nombreux [1, 5, 6, 9, 12]. De telles données peuvent permettre de reconnaître des événements, des actions ou même des interactions entre objets. Le but est de fournir une information pertinente pour l'exploitation automatique de vidéos [7, 8].
Différentes approches ont été étudiées pour la caractérisation de trajectoires. F. Porikli a proposé une modélisation par les orientations locales, invariante aux translations 2D et au facteur d'échelle, et une méthode de classification basée sur les MMC permettant de modéliser la causalité temporelle des trajectoires étudiées [9]. Néanmoins, il s'est appuyé sur des modélisations à l'aide de mélanges de gaussiennes posant des problèmes quant à la représentation de trajectoires (choix du nombre d'états et de gaussiennes dans les mélanges), et s'avèrent peu adaptées pour des trajectoires de petites tailles.

Contrairement aux méthodes existantes qui analysent des mouvements issus d'une même caméra (par exemple, pour la modélisation de parcours dans la surveillance d'un parking), notre but est de pouvoir traiter toutes les trajectoires (quelles que soient leurs longueurs et leurs provenances) et d'en extraire des classes correspondant à des mouvements similaires, en terme de chemin suivi et de vitesse de parcours, sans connaissance sur la calibration des caméras et sur la structure de la scène ou sur le mouvement 3D de l'objet mobile considéré. La méthode proposée doit également être capable de traiter des trajectoires issues de caméras différentes. Ainsi, nous avons conçu une représentation des trajectoires invariante à la translation 2D, à la rotation 2D et au facteur d'échelle, et une méthode originale de reconnaissance à partir de MMC prenant en compte l'évolution temporelle des événements dynamiques observés, et pouvant traiter des trajectoires de toutes tailles.
Le plan de l'article est le suivant. La section 2 présente la représentation invariante des trajectoires. En section 3 est décrite la méthode de comparaison de trajectoires à partir des MMC ainsi que le choix du nombre d'états des MMC. Dans la section 4, nous introduisons différentes méthodes de classification afin de

les comparer à la méthode développée. Enfin, en section 5, nous présentons les ensembles de trajectoires test utilisés, et les résultats expérimentaux sur données réelles.

2. Représentation des trajectoires

Dans le domaine de l'analyse et de l'interprétation de vidéos, les invariances à un certain nombre de transformations sont souhaitables : invariance à la translation 2D (reconnaissance indépendante de la position dans l'image), à la rotation 2D (indépendance à la direction globale de déplacement) ainsi qu'au facteur d'échelle (indépendance à la distance entre caméra et lieu de l'événement, au moins dans une certaine mesure). Elles seront d'une grande importance dans nombre d'applications vidéo pour traiter des contenus issus de caméras différentes.

2.1. Approximation par noyaux

Supposons qu'une trajectoire T_k soit définie par un ensemble de n_k points correspondant aux positions successives de l'objet suivi dans la séquence d'images, on note $T_k = \{(x_1, y_1), ..., (x_{n_k}, y_{n_k})\}$. Avant de calculer les descripteurs de cette trajectoire, qui seront des valeurs différentielles, il est préférable d'avoir une représentation continue des courbes formées par les trajectoires. Nous avons donc effectué une approximation par noyaux gaussiens de T_k, nécessitant le choix de h (paramètre de lissage de la courbe), définie par :

$$u_t = \frac{\sum_{j=1}^{n_k} e^{-(\frac{t-j}{h})^2} x_j}{\sum_{j=1}^{n_k} e^{-(\frac{t-j}{h})^2}},\ v_t = \frac{\sum_{j=1}^{n_k} e^{-(\frac{t-j}{h})^2} y_j}{\sum_{j=1}^{n_k} e^{-(\frac{t-j}{h})^2}}. \quad (1)$$

Les expressions explicites $\dot{u}_t$, $\dot{v}_t$, $\ddot{u}_t$ et $\ddot{v}_t$ correspondant aux dérivées temporelles à l'ordre un et à l'ordre deux de u et v peuvent alors être obtenues.

2.2. Sélection du paramètre de lissage

Afin d'avoir une méthode automatique d'extraction de la représentation, un moyen de fixer le paramètre h est nécessaire. Ainsi, pour une trajectoire T_k de taille n_k, un critère d'erreur quadratique moyenne (EQM) est considéré (Härdle *et al.* [4]), défini par

$$EQM(h) = EQM(\hat{u}_{t,h}) = \frac{1}{n_k} \sum_{i=1}^{n_k} (\hat{u}_{i,h} - u_i)^2,$$

où les u_i correspondent aux « vraies » valeurs à estimer. Une approximation naïve $p(h)$ de $EQM(h)$, ou « estimation par substitution », est faite en remplaçant u_i par x_i :

$$p(h) = \frac{1}{n_k} \sum_{i=1}^{n_k} (x_i - \hat{u}_{i,h})^2. \quad (2)$$

En additionnant et soustrayant u_i à (2), on obtient

$$\begin{aligned} p(h) &= \frac{1}{n_k} \sum_{i=1}^{n_k} ((x_i - u_i) + (u_i - \hat{u}_{i,h}))^2 \\ &= \frac{1}{n_k} \sum_{i=1}^{n_k} \varepsilon_i^2 + EQM(h) - \frac{2}{n_k} \sum_{i=1}^{n_k} \varepsilon_i (\hat{u}_{i,h} - u_i) \quad (3) \end{aligned}$$

où $\varepsilon_i = x_i - u_i$.

Le premier terme de (3), $\frac{1}{n_k} \sum_{i=1}^{n_k} \varepsilon_i^2$, est indépendant de h. De plus, en considérant une méthode de validation croisée [4], on montre que l'espérance du dernier terme de (3) est zéro si, à la place de $\hat{u}_{i,h}$, on considère $\hat{u}_{-i,h}$, *i.e.*,

$$E[-\frac{2}{n_k} \sum_{i=1}^{n_k} \varepsilon_i (\hat{u}_{-i,h} - u_i)] = 0,$$

où $\hat{u}_{-i,h}$ est un estimateur de validation croisée « leave-one-out » donné par :

$$\hat{u}_{-i,h} = \frac{\sum_{j \neq i} e^{-(\frac{i-j}{h})^2} x_j}{\sum_{j \neq i} e^{-(\frac{i-j}{h})^2}}.$$

Considérons alors le critère de validation croisée $VC(h)$ défini par :

$$VC(h) = \frac{1}{n_k} \sum_{i=1}^{n_k} (x_i - \hat{u}_{-i,h})^2.$$

Choisir h_{opt} tel que $VC(h)$ est minimisé revient donc à minimiser, en moyenne, $EQM(h)$. Le choix du paramètre est ainsi effectué en sélectionnant, parmi un ensemble de h, la valeur h_{opt} minimisant le critère de validation croisée $VC(h)$.

2.3. Invariance du descripteur local de trajectoires

La plupart des méthodes pour la classification de trajectoires développées jusqu'à maintenant utilisent les coordonnées spatiales dans l'image comme représentation (excepté [9]). Ces coordonnées sont en effet utiles afin d'étudier les ressemblances exactes entre trajectoires, mais notre approche est de considérer l'aspect général des trajectoires (en terme de forme et de vitesse). Prendre en compte les orientations locales successives des trajectoires est plus intéressant et cela permet de comparer leurs formes générales. Considérons les valeurs $\gamma_t = \arctan(\dot{v}_t / \dot{u}_t)$, nous avons alors une représentation invariante à la translation ainsi qu'au facteur d'échelle. Afin d'avoir une représentation également invariante aux rotations, considérons maintenant sa dérivée temporelle $\dot{\gamma}_t$. À l'aide de quelques calculs simples de dérivation et de trigonométrie, on prouve que :

$$\dot{\gamma}_t = \frac{\ddot{v}_t \dot{u}_t - \ddot{u}_t \dot{v}_t}{\dot{u}_t^2 + \dot{v}_t^2} = \kappa_t . \| V_t \|, \quad (4)$$

avec $\kappa_t = \dfrac{\ddot{v}_t \dot{u}_t - \ddot{u}_t \dot{v}_t}{(\dot{u}_t^2 + \dot{v}_t^2)^{\frac{3}{2}}}$ la courbure locale de la trajectoire et $\| V_t \| = (\dot{u}_t^2 + \dot{v}_t^2)^{\frac{1}{2}}$ la norme de la vitesse locale au point (u_t, v_t). Le numérateur de $\dot{\gamma}_t$, $\ddot{v}_t \dot{u}_t - \ddot{u}_t \dot{v}_t = \text{dét} \begin{pmatrix} \ddot{v}_t & \dot{v}_t \\ \ddot{u}_t & \dot{u}_t \end{pmatrix}$ est un déterminant, donc invariant à la rotation 2D. Le dénominateur de $\dot{\gamma}_t$ est $\dot{u}_t^2 + \dot{v}_t^2 = \| V_t \|^2$, invariant aux rotations 2D (norme de la vitesse). Par conséquent $\dot{\gamma}_t$ est invariant aux rotations 2D. Le vecteur de descripteurs utilisé pour représenter une trajectoire T_k (de taille n_k) donnée est alors le vecteur contenant les valeurs successives de $\dot{\gamma}_t$: $\phi = [\dot{\gamma}_1, \dot{\gamma}_2, ..., \dot{\gamma}_{n_k - 1}, \dot{\gamma}_{n_k}]$.

3. Distance entre trajectoires et classification

Afin de traiter efficacement la représentation invariante de trajectoires décrite dans la section précédente, nous avons recours aux Modèles de Markov Cachés (MMC). Cette modélisation doit permettre d'intégrer la causalité temporelle inhérentes à des trajectoires spatio-temporelles issues de vidéos. Une formalisation appropriée est introduite pour prendre en compte des ensembles de données de faibles tailles (*i.e.*, des trajectoires courtes). La sélection du nombre d'états de ces MMC est également traitée. Une distance entre trajectoires reposant sur la distance de Rabiner entre MMC [11] est ensuite présentée.

3.1. Approche par Modèles de Markov Cachés

La prise en compte de la causalité temporelle inhérente aux trajectoires issues de vidéos est réalisée à l'aide de MMC. Les états associés à la modélisation par MMC sont définis par les bins d'une quantification effectuée sur les descripteurs $\dot{\gamma}$. Pour modéliser la distribution des $\dot{\gamma}$, nous avons choisi de fixer un intervalle $[B_1, B_2]$ contenant un certain pourcentage P des valeurs de $\dot{\gamma}$ mesurées (pour l'ensemble des trajectoires considérées) autour de la valeur moyenne m des $\dot{\gamma}$. L'objectif est d'éliminer les mesures aberrantes et de maintenir les chaînes de Markov à un nombre d'états limité et représentatif. Ensuite, l'intervalle obtenu $[B_1, B_2]$ est divisé en un nombre N' de bins S_i qui vont former les états « intérieurs ». Deux bins extremaux (non bornés) S_1 et S_N sont ajoutés, respectivement définis par $]-\infty, B_1]$ et $[B_2, +\infty[$. Les états des MMC associés aux trajectoires seront donc les $N = N' + 2$ bins de la quantification (la Fig. 1 présente une quantification des $\dot{\gamma}$ observés).

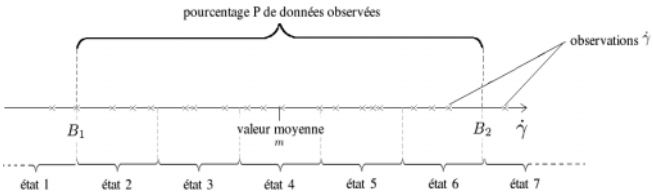


Figure 1. Exemple de quantification effectuée sur les $\dot{\gamma}$ calculés sur les trajectoires avec ici 5 bins, correspondant aux 5 états « intérieurs » ($N' = 5$ et donc $N = 7$).

Chaque trajectoire est modélisée à l'aide d'un MMC avec un nombre d'états donné par le nombre N de bins. Un MMC à N états est caractérisé par :

- la matrice de transition $A = \{a_{ij}\}$ avec

$$a_{ij} = P[\, q_{t+1} = S_j \mid q_t = S_i \,], \;\; 1 \leq i, j \leq N,$$

où q_t est l'état à l'instant t et S_i correspond à la valeur de l'état i.

- la distribution initiale des états $\pi = \{\pi_i\}$, où

$$\pi_i = P[\, q_1 = S_i \,], \;\; 1 \leq i \leq N.$$

- les probabilités d'observation conditionnelle b, où

$$b_i(X_t) = P[X_t \mid q_t = S_i], \;\; 1 \leq i \leq N$$

avec X_t l'observation au temps t.

F. Porikli a proposé une méthode de classification basée sur les MMC permettant de modéliser la causalité temporelle des trajectoires étudiées [9]. Néanmoins, il s'est appuyé sur une modélisation par MMC ayant une topologie de type « left-to-right » (*i.e.*, les transitions temporelles entre états du MMC s'effectuent entre états de numéros croissants) et des probabilités d'observation conditionnelle représentées par des mélanges de gaussiennes (MMG). Cette modélisation nécessite, pour chaque trajectoire, le choix non immédiat du nombre d'états (*i.e.*, du nombre de composantes des mélanges de gaussiennes) qui s'avère complexe lorsque l'on considère des trajectoires de petites tailles (avec peu d'observations). De plus, cette méthode est sensible aux tailles des trajectoires, ce qui signifie que des trajectoires de tailles très différentes seront considérées comme différentes. Notre but est d'avoir une méthode pouvant traiter des trajectoires de toutes tailles et également indépendante des tailles des trajectoires.

Ainsi, nous avons pris pour les états des chaînes de Markov les bins des histogrammes (le même nombre N pour toutes les trajectoires), et les observations X_t sont les $\dot{\gamma}_t$ calculés. Pour modéliser les probabilités conditionnelles d'observation $b_i(X_t)$, nous avons utilisé des gaussiennes centrées en μ_i (*i.e.*, le centre du bin S_i considéré). L'écart type σ de ces gaussiennes est choisi tel que l'intervalle $[\mu_i - \sigma, \mu_i + \sigma]$ correspond à la largeur des bins (Fig. 2). Cette modélisation à l'aide de MMC permet de traiter des trajectoires de toutes tailles.

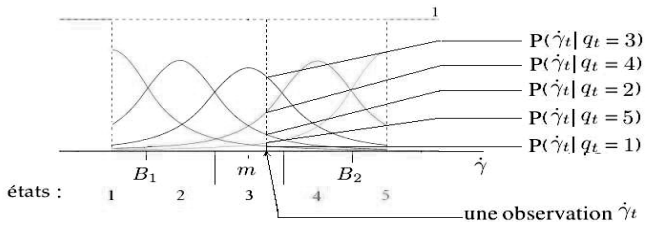
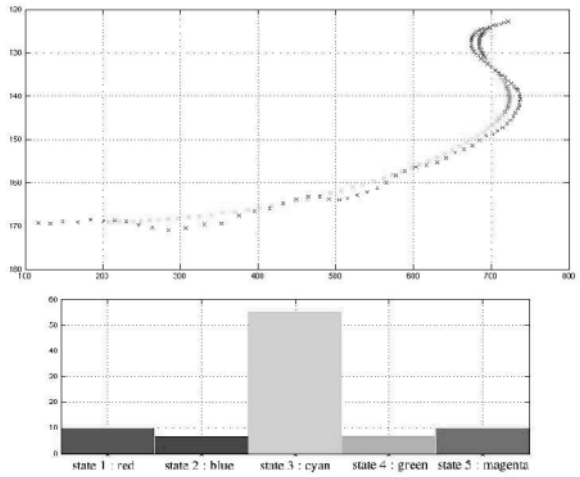


Figure 2. Modèle des probabilités d'observation conditionnelle, avec ici $N = 5$.

Afin d'estimer les paramètres du modèle, A et π, nous avons adopté la méthode par moindres carrés définie dans [10] (méthode s'appuyant sur un processus de comptage). Soit $H_t^{(i)} = P(\dot{\gamma}_t | q_t = i)$ (correspondant à un poids pour le processus de comptage effectué), les estimations de A et π sont, pour une trajectoire T_k comprenant n_k observations, données par (Fig. 3):

$$a_{ij} = \frac{\sum_{n=1}^{n_k-1} H_n^{(i)} H_{n+1}^{(j)}}{\sum_{n=1}^{n_k-1} H_n^{(i)}} \quad \text{et} \quad \pi_i = \frac{\sum_{n=1}^{n_k} H_n^i}{n_k}. \tag{5}$$



$$A = \begin{pmatrix} 0.80 & 0.12 & 0.05 & 0.02 & 0.00 \\ 0.14 & 0.61 & 0.16 & 0.08 & 0.01 \\ 0.02 & 0.11 & 0.71 & 0.13 & 0.03 \\ 0.01 & 0.09 & 0.17 & 0.56 & 0.17 \\ 0.01 & 0.02 & 0.07 & 0.13 & 0.77 \end{pmatrix}$$

$$\pi = [0.11\ 0.08\ 0.62\ 0.08\ 0.11]$$

Figure 3. Partie supérieure : tracé d'une trajectoire réelle (en noir) extraite d'une vidéo de Formule 1 et sa représentation lissée (en couleurs). Partie médiane : histogramme correspondant. Les couleurs correspondent aux différents états du MMC. Partie inférieure : matrice de transition A et distribution initiale des états π estimées.

3.2. Choix du nombre d'états des MMC

Dans la modélisation décrite dans la sous-section précédente, le nombre d'états « intérieurs » N' considérés dans la quantification (dans $[B_1, B_2]$) et décrivant les trajectoires reste à déterminer. Pour cela, un critère de décision statistique exprimant un équilibre entre le nombre d'états et la confiance dans les valeurs de l'histogramme correspondant a été développé. Le but est de définir un critère tel que le nombre d'états soit le plus faible possible (pour éviter une sur-représentation du modèle) tout en maximisant la confiance dans les valeurs associées à l'histogramme correspondant (afin d'avoir une représentation fiable).
Soit Θ_j la valeur vraie (inconnue) correspondant à la proportion de valeur $\dot{\gamma}$ dans l'état (ou bin) j de l'histogramme normalisé représentant la distribution des $\dot{\gamma}$ associés à une trajectoire T_k (de taille n_k). Soit également $\widehat{\Theta}_j$ son estimateur défini comme la proportion de $\dot{\gamma}$ observés dans l'état considéré:

$$\widehat{\Theta}_j = \frac{K_j}{n'_k},$$

où

$$K_j = \sum_{l=1}^{n'_k} X_{j,l} = \sum_{l=1}^{n'_k} \mathbb{1}_{\{\dot{\gamma}_l \in S_j\}}, j = 2 \dots N' + 1$$

est le nombre d'observations dans l'état S_j, et

$$n'_k = \sum_{i=2}^{N'_k+1} K_j$$

correspond au nombre total d'observations dans $[B_1, B_2]$. $X_{j,l} = \mathbb{1}_{\{\dot{\gamma}_l \in S_j\}}$ est la fonction d'appartenance de $\dot{\gamma}_l$ dans les états « intérieurs », et l'hypothèse est faite que $X_{j,l}$ suit une loi de Bernoulli sur $[B_1, B_2]$.
Alors, en utilisant le théorème central limite [3],

$$\frac{\widehat{\Theta}_j - \mathbb{E}[\widehat{\Theta}_j]}{\sqrt{\mathbb{V}[\widehat{\Theta}_j]}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),\ \forall j = 2 \dots N'_k + 1. \tag{6}$$

Donc, de façon asymptotique,

$$\widehat{\Theta}_j \to \mathcal{N}(\mathbb{E}[\widehat{\Theta}_j], \mathbb{V}[\widehat{\Theta}_j]), \quad \forall j = 2 \dots N' + 1.$$

$\widehat{\Theta}_j$ est trivialement un estimateur non biaisé de Θ_j, de sorte que l'intervalle de confiance IC_{95} (avec un pourcentage de confiance de 95 %) de Θ_j est défini par:

$$IC_{95}(\Theta_j) = [\widehat{\Theta}_j - \alpha_{95}\mathbb{V}[\widehat{\Theta}_j],\ \widehat{\Theta}_j + \alpha_{95}\mathbb{V}[\widehat{\Theta}_j]], \tag{7}$$

où le quantile α_{95} est la valeur assurant que, en considérant l'équation 7,

$$P(\Theta_j \in\ ]\widehat{\Theta}_j - \alpha_{95}\mathbb{V}[\widehat{\Theta}_j],\ \widehat{\Theta}_j + \alpha_{95}\mathbb{V}[\widehat{\Theta}_j][) \geq 0.95.$$

La variable aléatoire $X_{j,l}$ suit une loi de Bernoulli, ainsi,

$$\mathbb{V}[X_{j,l}] = \Theta_j(1 - \Theta_j).$$

En utilisant $\widehat{\Theta}_j$ comme estimateur non biaisé de Θ_j, $\mathbb{V}[X_{j,l}]$ peut être approchée par:

$$\mathbb{V}[X_{j,l}] \simeq \widehat{\Theta}_j(1 - \widehat{\Theta}_j).$$

Ainsi,

$$\mathbb{V}[\widehat{\Theta}_i] = \mathbb{V}[\frac{1}{n'_k}\sum_{l=1}^{n'_k} X_{i,l}] = \frac{1}{n'^2_k} V[\sum_{l=1}^{n'_k} X_{i,l}]$$
$$= \frac{1}{n'^2_k} n'_k \mathbb{V}[X_{i,l}] \simeq \frac{\widehat{\Theta}_i(1-\widehat{\Theta}_i)}{n'_k}.$$

L'intervalle de confiance $IC_{95}(\Theta_j)$ a une taille $|IC_{95}(\Theta_j)|$ pouvant être estimée par :

$$|IC_{95}(\Theta_j)| = 2\alpha_{95}\mathbb{V}[\widehat{\Theta}_j] \simeq 2\alpha_{95}\frac{\widehat{\Theta}_j(1-\widehat{\Theta}_j)}{n'_k}$$
$$\simeq 2\alpha_{95}\frac{K_j n'_k - K_j^2}{n'^3_k} \simeq 2\alpha_{95}\frac{K_j(n'_k - K_j)}{n'^3_k}.$$

Soit $m_{IC,k}$ la valeur moyenne des tailles des intervalles de confiance pour la trajectoire T_k, définie par :

$$m_{IC,k} = \frac{\sum_{j=2\ldots N'_k+1} |IC_{95}(\Theta_j)|}{N'_k}.$$

$|IC_{95}(\Theta_j)|$ (et donc $m_{IC,k}$) est une fonction décroissante de N'_k puisque K_j est une fonction décroissante de N'_k. Le critère de décision défini par l'équilibre entre le nombre d'états « intérieur » N'_k et la valeur moyenne des intervalles de confiance $m_{IC,k}$ est alors obtenu en choisissant $\widetilde{N'}_k$ minimisant $m_{IC,k} + \delta N'_k$:

$$\widetilde{N'}_k = \min_{N'_k}(m_{IC,k} + \delta N'_k).$$

Il reste à choisir δ, qui est un paramètre d'échelle permettant une comparaison pertinente de $m_{IC,k}$ et de N'_k. En considérant une distribution donnée, les estimations asymptotiques des proportions $\widehat{\Theta}_j, j = 2\ldots N'_k+1$ sont constantes. Ainsi, si $m_{IC,k}$ est une fonction décroissante de n'_k, alors le facteur d'échelle permettant que $m_{IC,k}$ et N'_k soient comparés, doit également avoir une forme décroissante en fonction de n'_k.

Une fonction décroissante $\delta(n'_k) = \frac{\beta}{n'_k}$ (fonction décroissante ayant la même forme que $m_{IC,k}$) est empiriquement considérée. Tout d'abord, une valeur $\hat{\delta}$ donnant des valeurs de $\widetilde{N'}_k$ compactes et distinctes pour les différentes classes de trajectoires considérées (voir la Fig. 7 qui contient les classes de trajectoires considérées) est choisie, maximisant les distances inter-classes et minimisant les distances intra-classes (analyse discriminante linéaire du premier ordre sur δ). Considérant les valeurs $\widetilde{N'}_{C_i}$ de nombre d'états trouvées pour les différentes classes C_i (*i.e.*, la valeur moyenne arrondi à l'entier le plus proche des $\widetilde{N'}_k$ trouvés pour les instances de la classe considérée), les intervalles de δ menant à cette valeur représentative $\widetilde{N'}_{C_i}$ (pour chaque classe C_i) sont alors considérés.

Par suite, une régression est effectuée sur les bornes inférieures et supérieures de ces intervalles, en considérant que :

$$\delta(n'_k) = \frac{\beta}{n'_k} + e_k.$$

La valeur β minimisant $\sum_i e_k^2$ est alors trouvée en effectuant une régression par moindres carrés.

Cette première fonction $\delta(n'_k)$ obtenue est ensuite utilisée pour fixer le nombre d'états $\widetilde{N'}_k$ pour l'ensemble des trajectoires considérées, menant à une nouvelle régression, et ainsi de suite jusqu'à ce que la fonction $\delta(n'_k)$ soit stable (*i.e.*, jusqu'à ce qu'il n'y ait plus de changement dans la séquence de $\widetilde{N'}_k$ associée aux trajectoires). Cette méthode permet d'avoir, pour toutes les classes de trajectoires, une fonction $\delta(n'_k)$ pertinente. Empiriquement, une valeur $\beta = 0.0175$ a été trouvée, de sorte que $\delta(n'_k) = \frac{0.0175}{n'_k}$ est une fonction d'échelle appropriée permettant un choix automatique du nombre d'états, pour toutes les trajectoires de toutes les classes.

La Fig. 5 résume les opérations permettant le choix de δ, avec tout d'abord la sélection de $\hat{\delta}$, le choix des $\widetilde{N'}_{C_i}$ et le processus itératif de régression jusqu'à obtention de $\delta_{n'k}$.

Afin d'effectuer une comparaison pertinente (*i.e.*, de comparer des MMC ayant le même nombre d'états), le choix d'un unique nombre d'états $\widetilde{N'}$ pour l'ensemble des trajectoires considérées est nécessaire. Ainsi, en utilisant la fonction $\delta(n'_k)$, le nombre d'états $\widetilde{N'}$ est donné par

$$\widetilde{N'} = \arg\min_{N'} \sum_k (m_{IC,k} + \delta N'), \quad (8)$$

en considérant l'ensemble des trajectoires (Fig. 4).

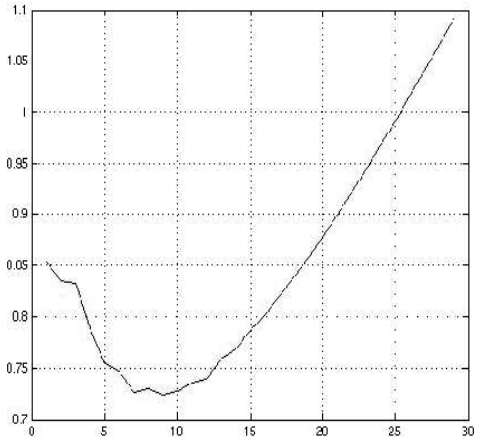


Figure 4. Fonction $\sum_k (m_{IC,k} + \delta N')$ utilisée pour fixer le nombre d'états « intérieurs » $\widetilde{N'}$ avec les données des 8 classes de trajectoires de Formule 1 de la Fig. 7.

3.3. Mesure de similarité et reconnaissance

La distance utilisée pour comparer deux MMC associés à deux trajectoires est celle proposée par Rabiner [11]. Étant donné deux MMC ayant pour ensembles de paramètres λ_1 et λ_2 ($\lambda_i = (A_i, b_i, \pi_i), i = 1,2$) représentant respectivement les trajectoires T_1 et T_2. On a

$$D(\lambda_1, \lambda_2) = \frac{1}{n_2}[\log\ P(O^{(2)}|\lambda_2) - \log\ P(O^{(2)}|\lambda_1)],$$

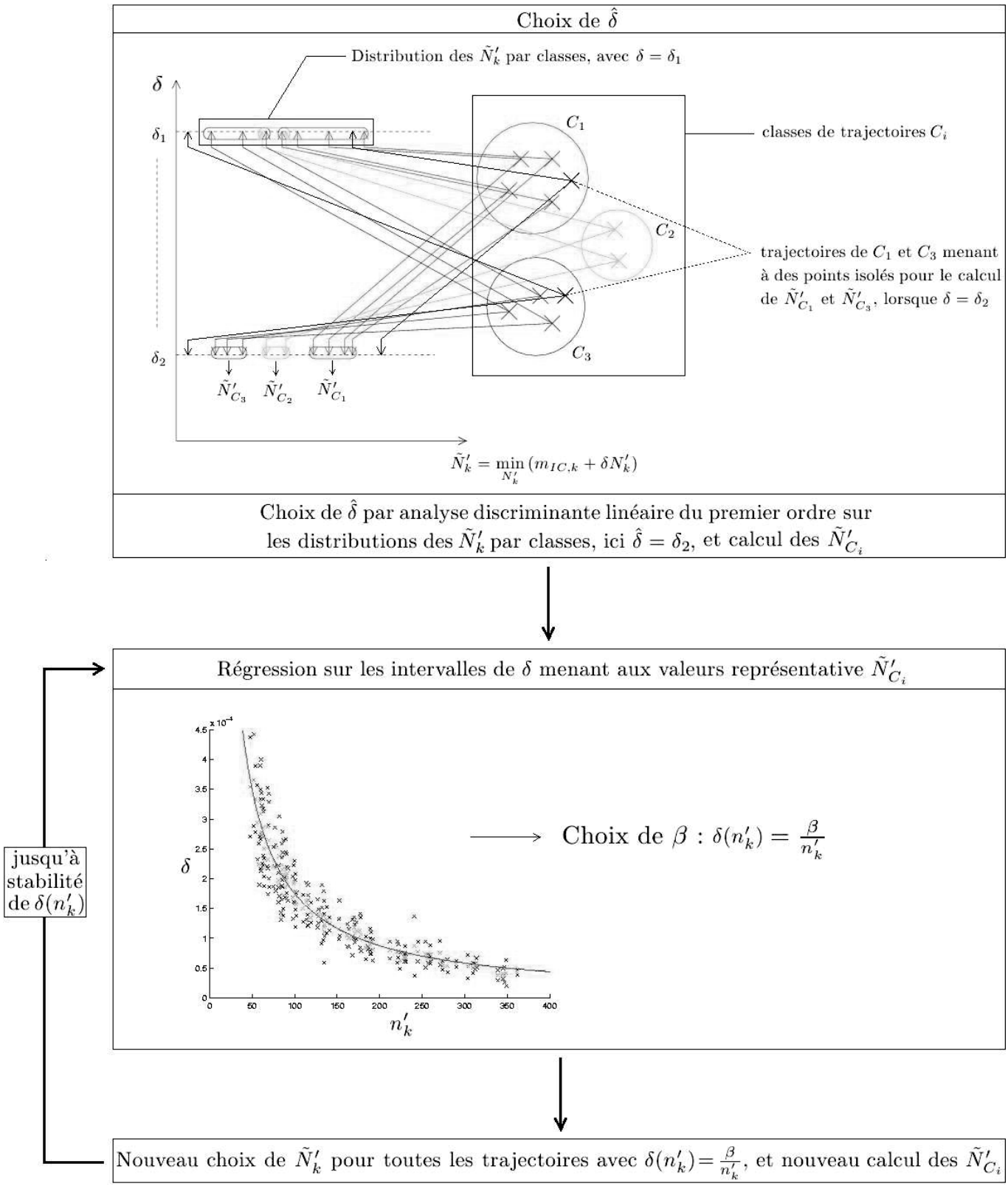


Figure 5. Schéma présentant l'algorithme de détermination de $\delta(n'_k)$.

où $O^{(j)} = \dot{\gamma}_1 \dot{\gamma}_2 ... \dot{\gamma}_T$ est la séquence des observations associée à la trajectoire T_j et $P(O^{(j)}|\lambda_i)$ exprime la probabilité d'observer $O^{(j)}$ avec le modèle λ_i. n_j dénote la taille de la trajectoire T_j. La version symétrisée de cette distance est

$$D_s(\lambda_1,\lambda_2) = \frac{1}{2}[D(\lambda_1,\lambda_2) + D(\lambda_2,\lambda_1)].$$

Une classe est représentée par un ensemble de MMC correspondant chacun à une des trajectoires utilisées pour apprendre cette classe. La reconnaissance est réalisée à l'aide d'une méthode d'agrégation par lien moyen, *i.e.*, en calculant la moyenne des distances entre la trajectoire testée T_i et toutes les trajectoires T_{l_j} de la classe G_j par :

$$D(T_i,G_j) = \frac{\sum_{T_{l_j} \in G_j} D_s(T_i,T_{l_j})}{\#G_j}. \quad (9)$$

La trajectoire testée est affectée à la classe qui correspond à la distance $D(T_i,G_j)$ minimale.

4. Autres méthodes de classification

L'approche que nous avons développée considère les trajectoires comme des motifs dynamiques là où la plupart des autres méthodes lient les trajectoires vidéos aux points de vue à partir desquels elles sont filmées (par exemple, modélisation automatique, dans un parking, des chemins de passage usuels à partir des vidéos prises par une caméra fixe de surveillance).
Aussi, afin d'évaluer notre approche de façon pertinente et de mettre en valeur des propriétés intéressantes, nous avons développé ou adapté d'autres méthodes pour une évaluation comparative.

4.1. Distance de Bhattacharyya entre histogrammes

Afin de démontrer l'importance de la prise en compte de la causalité temporelle, *i.e.*, les transitions entre états, nous avons implanté une méthode de classification basée sur la distance de Bhattacharyya entre histogrammes. La distance de Bhattacharyya D_b entre deux histogrammes normalisés h_i et h_j est définie par :

$$D_b(h_i,h_j) = 1 - \sum_{k=1}^{N} \sqrt{h_i^k h_j^k}$$

où h_i^k est la valeur dans le bin k de l'histogramme associé à la trajectoire i. La reconnaissance est ensuite obtenue de façon analogue à l'approche par MMC, *i.e.*, à l'aide d'une méthode d'agrégation par lien moyen. La procédure de choix du nombre d'états décrite en sous-section 3.2 est appliquée également au choix automatique du nombre de bins des histogrammes.

4.2. Méthode de classification par SVM

Un outil efficace de classification est le SVM (Séparateur à Vaste Marge) [2]. En entrée des SVM, nous avons choisi d'utiliser les paramètres des MMC associées aux trajectoires. Les paramètres en entrée des SVM doivent être représentés sous formes de vecteurs. Par conséquent, pour chaque trajectoire, nous avons créé un vecteur X_i contenant les paramètres du MMC correspondant. Par exemple, considérons le MMC de paramètres λ_i correspondant à la trajectoire T_i (pour des facilités de présentation, nous développons un exemple avec $N = 3$) avec

$$A_i = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \pi_i = [a_1\ a_2\ a_3].$$

Alors $X_i = [a_{11}\ a_{12}\ a_{13}\ a_{21}\ a_{22}\ a_{23}\ a_{31}\ a_{32}\ a_{33}\ a_1\ a_2\ a_3]$ sera le vecteur caractérisant la trajectoire T_i. Nous utilisons une technique de classification par SVM à l'aide d'un noyau gaussien. Les résultats obtenus sont issus d'un schéma de classification « un contre tous », les paramètres du SVM ayant été fixés à l'aide d'une validation croisée.

4.3. Modélisation par MMC/MMG

Une méthode de comparaison entre trajectoires issues de vidéo, inspirée par les travaux de Porikli [9], a été aussi développée. La distance entre MMC s'appuyant sur des modèles de mélanges de gaussiennes (MMG) a été étendue à l'analyse des $\dot{\gamma}_t$. Dans ce cas, nous visons la mise en valeur des performances de la méthode proposée pour le traitement efficace des petites trajectoires. Un MMC ergodique, insensible aux tailles des trajectoires (contrairement à la modélisation « left-to-right » proposée par Porikli) a été utilisé, la modélisation des états et des probabilités d'observations conditionnelles étant effectuée à l'aide de MMG. L'initialisation des MMG est réalisée par un algorithme de type K-means. Pour déterminer le nombre d'états à utiliser dans le MMC, un critère de type CIB (critère d'information bayésien) a été utilisé. Dans cette méthode, un MMC/MMG est créé pour chaque trajectoire. Les différentes tâches de reconnaissance vidéo considérées ont été réalisées avec les MMC/MMG comme pour notre méthode, en utilisant la distance de Rabiner D_s entre MMC.

5. Expérimentations

5.1. Trajectoires vidéos réelles

Nous avons traité des trajectoires réelles extraites de vidéos correspondant à des programmes TV de course de Formule 1 et de ski, chacune filmée par plusieurs caméras. Les trajectoires ont été obtenues à l'aide d'une méthode de suivi basée sur le calcul

du déplacement associé à des points d'intérêt extraits sur l'objet suivi [13]. Le mouvement de l'arrière plan dû à un mouvement de la caméra (panoramique ou zoom) est compensé (Fig. 6 et 8). Ainsi, dans le cas de la Formule 1, les trajectoires obtenues (Fig. 7) par cette méthode sont visuellement assez similaires aux trajectoires 3D réelles (à une homographie près, le mouvement 3D étant quasi-planaire).

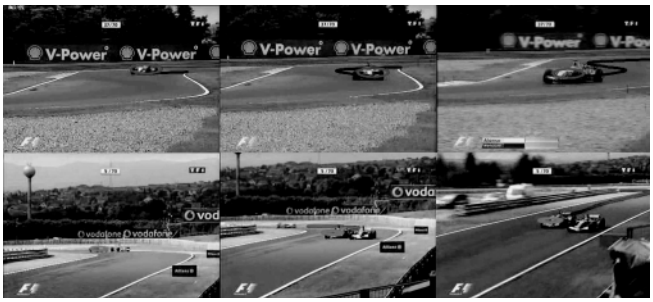


Figure 6. Images de plans vidéo d'un programme TV de Formule 1 acquis par deux caméras différentes placées à deux endroits différents du circuit.

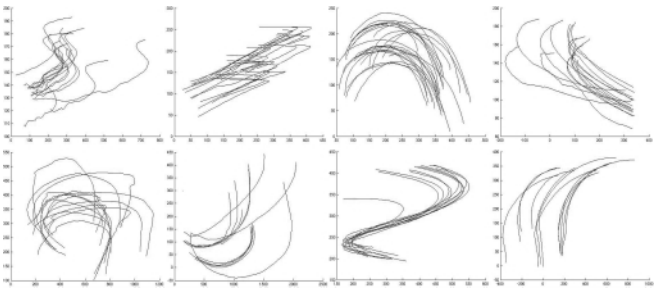


Figure 7. Tracé des 8 classes de trajectoires (125 trajectoires) extraites d'une vidéo de Formule 1. Une classe est composée de trajectoires extraites de plans filmés par une même caméra (placée à un endroit donné du circuit).

Figure 8. Images de plans vidéo d'un programme TV de ski acquis par deux caméras différentes dans deux courses différentes (la première correspond à une trajectoire de descente et la deuxième à une trajectoire de slalom).

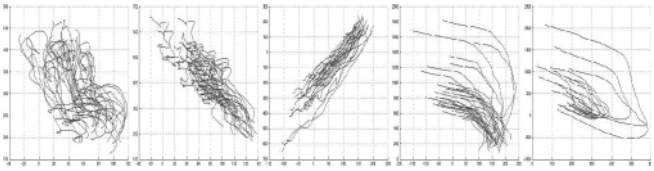


Figure 9. Tracé de 5 classes de trajectoires (134 trajectoires) extraites de vidéos de ski. Une classe est composée de trajectoires extraites de plans filmés par une même caméra. Les trois classes sur la gauche correspondent à des trajectoires issues de vidéos de slalom, les deux de droite correspondant à des vidéos de descente.

5.2. Résultats expérimentaux

Pour évaluer les performances de reconnaissance, nous avons adopté la technique de validation croisée « leave-one-out ». Expérimentalement, la valeur de P (pourcentage de données considéré pour définir l'intervalle $[B_1, B_2]$) a été fixée à $P = 95\ \%$.

La tâche de reconnaissance a été appliquée aux ensembles de trajectoires de ski et de Formule 1. Deux cas ont été considérés : 6 et 8 classes de trajectoires, le groupe de 6 classes étant composé des 3 premières classes de chaque rangée présentées à la Fig. 7). Une classe est composée de trajectoires extraites de plans filmés par une même caméra. Les différentes classes de trajectoires de Formule 1 correspondent à des caméras placées à différents virages, la forme de ces virages induisant celle de la trajectoire des véhicules. De même, pour les trajectoires de ski, les formes des trajectoires appartenant à différentes classes sont induites par les parcours imposés sur la piste. Tab. 1 et Tab. 2 contiennent les résultats de classification obtenus en considérant respectivement les trajectoires de Formule 1 et les trajectoires de skieurs.

Des résultats très satisfaisants ont été obtenus avec la méthode proposée. La méthode utilisant les SVM donne également des résultats très intéressants, bien que moins précis avec 8 classes. Cela montre l'importance de l'utilisation de l'algorithme de Viterbi pour calculer la distance de Rabiner, là où la méthode basée sur les SVM utilise seulement les paramètres des MMC. Des résultats satisfaisants mais moins précis ont également été obtenus dans le cas de la méthode basée sur la distance de Bhattacharyya, mettant en valeur l'apport de la prise en compte de la causalité temporelle (Tab. 1). La méthode basée sur les MMC/MMG montre l'inaptitude d'un modèle trop compliqué à modéliser des données en faible quantité. Notre méthode basée sur les MMC est utilisable pour la comparaison de trajectoires ayant des tailles très différentes. De plus, elle est plus flexible que la méthode utilisant les SVM. Il est notamment possible avec notre méthode de détecter des trajectoires anormales, de procéder à un clustering multi-classes ou d'ajouter des classes avec un apprentissage complémentaire relatif uniquement aux classes ajoutées.

La méthode de sélection d'un nombre d'états dans les MMC (sous-section 3.2) s'est avérée efficace. En effet, en comparant les résultats de classification obtenus avec la sélection automatique de $\widetilde{N}'$ aux résultats atteignable en considérant une large palette de N' possibles, on constate que $\widetilde{N}'$ correspond à un choix proche du choix optimal pour N' dans la plupart des cas de classification testés. Nous avons ainsi une méthode automatique d'analyse et de comparaison des trajectoires.

Il est également important de préciser qu'une classification parfaite a été obtenue en considérant deux classes pour les trajectoires de skieurs issues de vidéos de « slalom » et « descente »: la classe « slalom » regroupant les trois classes à gauche dans Fig. 9 et la classe « descente » composée des deux classes de droite, et ce avec toutes les méthodes testées. Ceci met en valeur la pertinence de la représentation à l'aide des descripteurs $\dot{\gamma}$ pour distinguer des trajectoires issues de vidéos. Rappelons que les classes considérées pour ces dernier tests sont composées de trajectoires issues de caméras différentes, d'où l'importance des invariances de représentation des trajectoires de ces classes.

Tableau 1. Comparaison des résultats de reconnaissance pour les trajectoires extraites de vidéo de Formule 1. Les pourcentages ont été obtenus en considérant une technique de validation croisée « leave-one-out ».

	Pourcentage de bonne classification	
# classes	6	8
Notre méthode	99	94.4
Bhattacharyya	96	93.6
SVM	99	92.8
MMC/MMG	96	80

Tableau 2. Comparaison des résultats de reconnaissance pour les trajectoires extraites de vidéo de ski. Les pourcentages ont été obtenus en considérant une technique de validation croisée « leave-one-out ».

	Pourcentage de bonne classification
Notre méthode	91.7
Bhattacharyya	91.7
SVM	91
MMC/MMG	78.2

6. Conclusion

Nous avons proposé une méthode de reconnaissance d'événements dans des vidéos exploitant les trajectoires des objets mobiles dans l'image. Nous avons introduit une représentation appropriée des trajectoires, invariante à la translation, à la rotation et au facteur d'échelle, et proposé une méthode d'extraction robuste. La causalité temporelle de cette représentation a ensuite été exploitée à l'aide d'une modélisation originale par MMC, permettant l'analyse efficace de trajectoires de toutes tailles. Un soin particulier a également été porté au choix du nombre d'états associés à ces MMC. En comparant la méthode développée avec d'autres méthodes (incluant notamment une méthode par SVM), nous avons justifié le choix de cette représentation ainsi que le schéma de classification par MMC. Des résultats très encourageants ont été obtenus sur données réelles.

Références

[1] F. BASHIR, A. KHOKHAR et D. SCHONFELD. *Real-time motion trajectory-based indexing and retrieval of video sequences*. IEEE Trans. on Multimedia, vol.9, no.1, pp. 58-65, 2007.

[2] C. BURGES. *A tutorial on support vector machines for pattern recognition*. Data Mining and Knowledge Discovery, Springer, no.2, pp. 121-167, 1998.

[3] W. FELLER. *An Introduction to Probability Theory and Its Applications*. Vol. 2, 3rd ed, Wiley, New York, 1971.

[4] W. HÄRDLE, M. MULLER, S. SPERLICH et A. WERWATZ. *Nonparametric and semiparametric models*. Springer, Springer series in statistics, Berlin, Germany, 2004.

[5] W. HU, X. XIAO, Z. FU, D. XIE, T. TAN et S. MAYBANK. *A system for learning statistical motion patterns*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.28, no.9, pp 1450-1464, sept. 2006.

[6] T. IZO et W.E L. GRIMSON. *Unsupervised modeling of object tracks for fast anomaly detection*. IEEE Int. Conf. on Image processing, ICIP'07, San Antonio, US, sept. 2007.

[7] A. KOKARAM, N. REA, R. DAHYOT, M. TEKALP, P. BOUTHEMY, P. GROS, et I. SEZAN. *Browsing sports video (Trends in sports-related indexing and retrieval work)*. IEEE Signal Processing Magazine, vol.23, no.2, pp 47-58, mars 2006.

[8] G. PIRIOU, P. BOUTHEMY et J.-F. YAO. *Recognition of dynamic video contents with global probabilistic models of visual motion*. IEEE Trans. on Image Processing, vol.15, no.11, pp 3417-3430, 2006.

[9] F. PORIKLI. *Trajectory distance metric using hidden Markov model based representation*. PETS Workshop, Prague, mai 2004.

[10] J. FORD et J. MOORE. *Adaptive estimation of HMM transition probabilities*. IEEE Trans. on Signal Processing, vol. 46, no. 5, mai 1998.

[11] L. RABINER. *A tutorial on hidden Markov models and selected applications in speech recognition*. Proc. IEEE, vol. 77, no. 2, pp. 121-167, 1989.

[12] X. WANG, K. TIEU et E. GRIMSON. *Learning semantic scene models by trajectory analysis*. Europ. Conf. on Computer Vision, ECCV'06, Graz, Autriche, Mai 2006.

[13] N. GENGEMBRE et P. PÉREZ. *Probabilistic color-based multi-object tracking with application to team sports*. Technical report, INRIA, RR-6555, mai 2008.

Alexandre **Hervieu**

Alexandre Hervieu est ingénieur de l'INSA Rouen, département Mathématiques. Aprés un DEA en Mathématiques Fondamentales et Appliquées (MFA) en 2005 à l'université de Rouen, il a effectué une thèse, sous l'encadrement de Patrick Bouthemy et Jean-Pierre Le Cadre, à l'INRIA - Rennes Bretagne Atlantique soutenue en mars 2009 et intitulée « Analyse de trajectoires vidéos à l'aide de représentations markoviennes pour l'interprétation de contenus ». Ses centres d'intérêts se situent dans le domaîne de la compréhension de contenus vidéos : reconnaissance d'évènements et détection d'évènements rares, reconnaissance de gestes, d'actions et d'activités, reconnaissance de formes.

Jean-Pierre **Le Cadre**

Après des études de mathématiques (Maîtrise math., DEA math. 1977), J.P. Le Cadre a obtenu un DEA en traitement de l'information, puis soutenu une thèse de 3ème cycle (INPG Grenoble, 1982) et enfin une thèse d'Etat (INPG Grenoble, 1987). De 1980 à 1989, il effectue des recherches au GERDSM (Groupe d'Etudes et de Recherche en Détection Sous-Marines, Toulon), essentiellement dans le domaine du traitement d'antenne et de la détection sous-marine. Depuis 1989, il mène ses travaux au sein de l'IRISA (Rennes), où il est Directeur de Recherche CNRS. Ses centres d'interêts ont considérablement évolués et se situent dans les domaines de l'analyse des sytèmes de détection, de poursuite et d'associations de données.

Patrick **Boutemy**

Patrick Bouthemy est ingénieur Télécom-Paris, Docteur-Ingénieur (1982) et Habilité à Diriger des Recherches (1989) de l'Université de Rennes 1. Il est actuellement Directeur de Recherche Inria à l'Irisa. Il a créé en 1997 le projet Vista qu'il a dirigé jusqu'en 2007. Il a été président du Comité des projets de l'Irisa de 1998 à 2002. Il est depuis le 1er juillet 2007 directeur du centre de recherche Inria Rennes – Bretagne Altantique. Ses axes principaux de recherche sont les suivants : modèles statistiques pour le traitement de séquences d'images, analyse du mouvement, reconnaissance de contenus dynamiques et apprentissage.