Modélisation statistique non paramétrique et reconnaissance du mouvement

Ronan Fablet, Patrick Bouthemy

Modélisation statistique non paramétrique et reconnaissance du mouvement
Statistical non parametric modeling and motion recognition

R. Fablet[1] et P. Bouthemy[2]
[1]IRISA/CNRS [2]IRISA/INRIA
Campus universitaire de Beaulieu 35042 Rennes Cedex, France
e-mail: {rfablet,bouthemy}@irisa.fr

Résumé

Nous présentons une méthode originale d'analyse non paramétrique du mouvement dans des séquences d'images. Elle repose sur une modélisation statistique de distributions de mesures locales (partielles) de mouvement directement évaluées à partir des intensités des images. La définition de modèles de Gibbs temporels multi-échelles permet de prendre en compte simultanément des propriétés spatiales et temporelles du mouvement. La caractéristique essentielle de ces modèles réside dans le calcul simple et direct de leur fonction de vraisemblance. Ceci rend possible d'une part l'estimation des modèles au sens du maximum de vraisemblance, et, d'autre part, la formulation de la reconnaissance du mouvement comme un problème d'inférence statistique. Nous avons ainsi mené des expériences de reconnaissance du mouvement sur un ensemble de séquences d'images contenant des situations dynamiques réelles variées.

Mots Clef

Analyse et reconnaissance du mouvement, modèles statistiques, modèles de Gibbs, cooccurrences.

Abstract

We present an original approach for non parametric motion analysis in image sequences. It relies on the statistical modeling of distributions of local motion-related measurements computed over image sequences. The use of temporal multiscale Gibbs models allows us to handle in a unified statistical framework both spatial and temporal properties of motion content. The important feature of our probabilistic scheme is to make the exact computation of conditional likelihood functions feasible and simple. It enables us to straightforwardly achieve model estimation according to ML criterion and to benefit from a statistical point of view for classification issues. We have conducted motion recognition experiments over a large set of real image sequences comprising various motion types.

Keywords

Motion analysis and recognition, statistical modeling, Gibbs models, cooccurrences.

1 Introduction

L'interprétation d'informations de nature dynamique est au coeur du processus de perception visuelle [2]. L'analyse du mouvement dans des séquences d'images pour l'interprétation ou la classification de scènes dynamiques constitue ainsi une thématique de recherche importante en vision par ordinateur. Dans ce domaine, les travaux se sont initialement concentrés sur le calcul de champs de vitesses à partir de séquences d'images, qui est connu pour être un problème inverse mal posé [1, 12]. Toutefois, comme il est souligné dans [8], il n'est pas toujours nécessaire de disposer de ces informations de mouvement complètes pour effectuer une analyse qualitative du contenu dynamique dans des séquences d'images. Pour certaines applications comme la classification du mouvement [14] ou la reconnaissance d'activités [4], il s'avère suffisant d'extraire des images des représentations spatio-temporelles éventuellement partielles. Dans cet article, nous adoptons ce point de vue et nous considérons le problème de la reconnaissance du mouvement sans connaissance *a priori* sur la scène observée. Notre objectif est de proposer un schéma générique de caractérisation globale du mouvement dans des séquences d'images.

Dans ce contexte, il apparaît nécessaire d'introduire des alternatives, dites "non paramétriques", aux méthodes basées sur des modèles de mouvement 2D paramétriques. Les travaux précurseurs [14] de Nelson et Polana dans ce domaine ont introduit la notion de texture temporelle qui regroupe des scènes dynamiques complexes telles que des mouvements de feuillage, des scènes de rivière. Ils ont plus particulièrement exploité des techniques développées initialement pour l'analyse de texture spatiale, pour décrire les distributions de mesures locales de mouvement dans des séquences d'images. La caractérisation des scènes dynamiques extraite de cette manière porte sur des informations générales d'activité de mouvement. Dans le prolongement de ces travaux, de nouveaux développements ont été proposés pour des applications en indexation vidéo par le contenu [6, 16].

Nous explorons plus avant ce type de méthodes. Nous introduisons des modèles probabilistes non paramétriques du mouvement et spécifions le problème de reconnaissance dans un cadre statistique. Des modèles de Gibbs temporels

multi-échelles sont considérés pour représenter les distributions de mesures locales de mouvement. Nous pouvons ainsi considérer une large gamme de situations dynamiques (mouvement rigide, texture temporelle, entités dynamiques uniques ou multiples, ...). On pourra parler plus généralement de modèles d'activité de mouvement. Modèle statistique de mouvement sera aussi employé de manière équivalente. Cet article est organisé de la manière suivante. Le paragraphe 2 présente les idées directrices de ces travaux. Les mesures locales de mouvement utilisés pour la modélisation non paramétrique du mouvement sont décrites au paragraphe 3. La modélisation statistique du mouvement est introduite au paragraphe 4. Enfin, le paragraphe 5 présente l'application de ces modèles probabilistes à la reconnaissance du mouvement et le paragraphe 6 conclut cet article.

2 Contexte de l'étude

L'analyse non paramétrique du mouvement vise à caractériser globalement la distribution du mouvement dans des séquences d'images. Il faut alors distinguer les propriétés spatiales et temporelles de l'information de mouvement. La figure 1 fournit une illustration de ces deux types de caractéristiques du contenu dynamique pour deux séquences différentes : la première est une séquence de plateaux de journal télévisé et la seconde une scène de rivière correspondant à une forte activité de mouvement. Outre la première image de ces deux séquences, nous présentons pour chacune, d'une part, une carte de mesures locales de mouvement, dont le mode de calcul est présenté en détails dans le paragraphe suivant, et d'autre part, une courbe représentant l'évolution temporelle, sur 25 images successives, de la quantité locale de mouvement calculée au centre de l'image. Les cartes des quantités locales de mouvement fournissent un aperçu de l'organisation spatiale du mouvement dans la scène. De manière complémentaire, l'étude en un point donné de l'évolution temporelle de la quantité de mouvement permet d'appréhender la variabilité temporelle du mouvement suivant le type de phénomènes dynamiques considérés.

Les travaux dédiés à la caractérisation globale du mouvement se sont initialement concentrés sur la caractérisation de l'organisation spatiale du mouvement et repose sur des techniques développées dans le cadre de l'analyse de texture spatiale. Dans [14], des attributs globaux de mouvement sont extraits de distributions de cooccurrences spatiales de champs de vitesses normales et sont exploités pour classer des séquences d'images soit comme instances de mouvements simples (translation, rotation, divergence) soit comme des textures temporelles. Dans [16], l'ajout de nouveaux descripteurs toujours calculés à partir des vitesses normales est proposé selon d'autres méthodes (spectre de Fourier, statistiques des différences locales,...). Dans les deux cas, les attributs extraits conduisent uniquement à une caractérisation de la configuration spatiale du mouvement dans une image donnée (c.a.d., à un instant donné). Afin de décrire des propriétés temporelles de l'information de

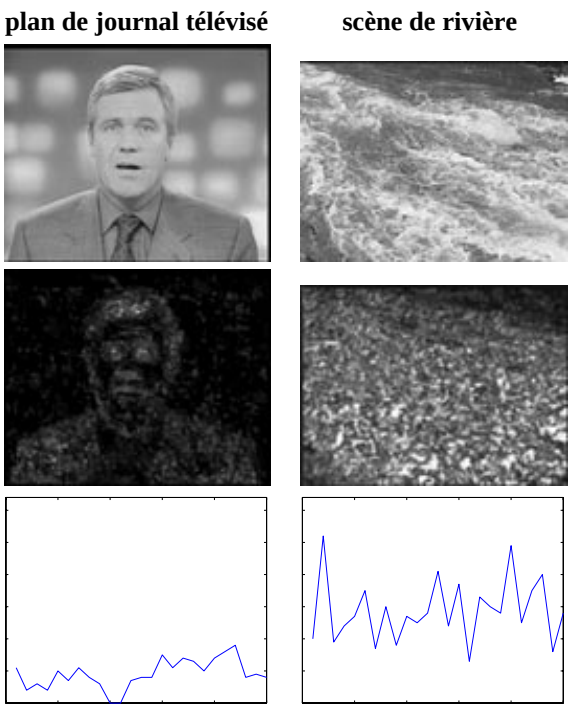


FIG. 1 – *Illustration de la notion de propriétés spatiales et temporelles du mouvement apparent dans des séquences d'images pour une séquence de plateaux de journal télévisé et une scène de rivière. La première ligne contient les premières images des deux séquences traitées, la deuxième les cartes de mesures locales relatives à l'amplitude du mouvement calculée à partir des deux premières images des deux séquences traitées (cf. paragraphe 3), et la troisième l'évolution temporelle de la quantité locale du mouvement calculée au centre des images pour 25 images successives.*

mouvement, nous avons proposé d'extraire des attributs de mouvement calculés à partir de distribution de cooccurrences temporelles de mesures locales de mouvement [6]. Il semble néanmoins préférable de combiner des caractérisations des aspects spatiaux et temporels du contenu dynamique dans une optique de reconnaissance du mouvement. Ceci peut par exemple être effectué à partir de filtres de Gabor spatio-temporels appliqués aux intensités des images comme dans [19].

D'autre part, les modèles probabilistes, tels que les champs de Gibbs [9, 20], ont conduit à des avancées importantes en analyse de texture. Ces modèles constituent des alternatives puissantes à l'extraction de vecteurs de descripteurs. En particulier, ils rendent plus aisée la formulation des problèmes de classification ou d'apprentissage. Dans [18], l'emploi de modèles probabilistes a été envisagé pour la synthèse de texture temporelle. Cependant, les modèles auto-régressifs employés ne peuvent pas être appliqués à la modélisation et à la reconnaissance du mouvement.

Nous exploitons des modèles de Gibbs du fait de la relation explicite entre ces modèles probabilistes et les me-

sures de cooccurrences [10, 20]. Toutefois, l'utilisation directe de modèles de Gibbs généraux pour des problèmes de reconnaissance et de classification se révèle impossible. En effet, les fonctions de vraisemblance qui leur sont associées ne peuvent être évaluées numériquement, du fait du calcul impossible (en pratique) de leurs fonctions de partition. Cela interdit alors la comparaison des vraisemblances conditionnelles d'observations relativement à deux modèles différents. Nous considérons donc des classes de modèles de Gibbs particuliers, pour lesquels il existe une formulation causale équivalente. Dans ce cas, nous pouvons évaluer exactement et simplement leurs fonctions de vraisemblance. Nous introduisons des modèles de Gibbs temporels multi-échelles spécifiés sur des séquences de cartes de mesures locales de mouvement. Cette modélisation multi-échelle nous permet de définir des modèles causaux, tout en appréhendant dans un cadre statistique unifié, des aspects à la fois spatiaux et temporels de l'information de mouvement.

3 Quantités locales de mouvement

3.1 Mesures locales de mouvement

Notre approche pour l'analyse non paramétrique du mouvement repose sur la modélisation statistique de distributions de mesures locales de mouvement. Comme nous l'avons déjà évoqué, l'estimation de champs denses de vitesses est un problème difficile, notamment pour des scènes dynamiques complexes telles que les textures temporelles. Par conséquent, nous préférons considérer des mesures locales partielles du mouvement directement calculées à partir des gradients spatio-temporels des intensités. Sous une hypothèse de conservation de l'intensité sur les trajectoires des points dans l'image, l'Équation de Contrainte du Mouvement Apparent (ECMA) permet d'établir la relation bien connue suivante [12] :

$$\mathbf{w}(p) \cdot \nabla I(p) + I_t(p) = 0 \qquad (1)$$

où ∇I est le gradient spatial de la fonction intensité I et I_t sa dérivée temporelle, $\mathbf{w}(p)$ le vecteur de vitesse au point p. De la relation (1) se déduit l'expression de la vitesse normale, $v_n(p) = -I_t(p)/\|\nabla I(p)\|$, qui est exploitée dans [14, 16]. Cependant, cette quantité est connue pour être peu robuste aux bruits de mesure du gradient de l'intensité ∇I. Pour pallier ce problème, nous considérons une moyenne pondérée des vitesses normales sur une fenêtre locale. Les pondérations sont données par la norme des gradients spatiaux d'intensité [15]. Ainsi, nous évaluons une mesure locale de mouvement plus robuste que la vitesse normale et définie par :

$$v_{obs}(p) = \frac{\sum_{q \in \mathcal{F}(p)} \|\nabla I(q)\| \cdot |I_t(q)|}{\max\left(\eta^2, \sum_{q \in \mathcal{F}(p)} \|\nabla I(q)\|^2\right)} \qquad (2)$$

où $\mathcal{F}(p)$ est une fenêtre de taille 3×3 centrée au point p. η^2 est une constante prédéfinie liée au niveau de bruit dans les images (typiquement, $\eta = 5$).

De manière évidente, l'emploi de cette mesure locale de mouvement ne nous permet d'accéder à aucune information relative à la direction du mouvement. Il nous est donc par exemple impossible de différencier des translations de directions différentes. Toutefois, notre objectif consiste à caractériser globalement le contenu dynamique en termes généraux d'activité de mouvement. D'autre part, contrairement à [14, 16], nous n'utilisons pas les directions des vitesses normales, car il s'agit en fait d'informations liées à la texture spatiale (gradients spatiaux d'intensité) de la scène et non à son contenu dynamique intrinsèque. Or, nous cherchons à accéder à une description globale du mouvement indépendante de la disposition spatiale de la scène.

Une autre propriété intéressante des mesures locales de mouvement considérées réside dans l'existence de bornes d'interprétation de ces quantités. Étant donné un niveau de détection de l'amplitude du mouvement δ dans les images, il existe deux bornes $l_\delta(p)$ et $L_\delta(p)$ qui vérifient les propriétés suivantes. Si la valeur de la mesure $v_{obs}(p)$ est inférieure à $l_\delta(p)$, l'amplitude du déplacement réel (inconnu) $\|\mathbf{w}(p)\|$ au point p est inférieure à δ. Au contraire, si $v_{obs}(p)$ est supérieure à $L_\delta(p)$, $\|\mathbf{w}(p)\|$ est supérieure à δ. $l_\delta(p)$ et $L_\delta(p)$ sont directement calculables à partir des dérivées spatiales de la fonction intensité sur la fenêtre $\mathcal{F}(p)$. Nous invitons le lecteur à se référer à [15] pour davantage de détails sur les expressions de ces bornes.

L'ECMA (relation (1)) est connue pour présenter plusieurs limites. En premier lieu, elle permet seulement d'appréhender des mouvements de faible amplitude. D'autre part, elle n'est pas valide dans des zones d'occultations ou en présence de changements d'illumination. Afin de prendre en compte ces limites, nous exploitons une procédure multi-échelle basée sur le test statistique décrit dans [11] afin d'évaluer la validité de l'ECMA. Nous construisons tout d'abord une pyramide gaussienne pour la paire d'images successives traitées. Étant donné un point p, nous sélectionnons alors le niveau d'échelle le plus fin pour lequel l'ECMA est valide. Nous évaluons à ce niveau la mesure locale de mouvement $v_{obs}(p)$ et les bornes $l_\delta(p)$ et $L_\delta(p)$. Si l'ECMA reste invalide à toutes les échelles, nous n'évaluons aucune mesure de mouvement au point p.

3.2 Quantification markovienne robuste

Notre approche peut être assimilée à une extension des modèles de texture des images en niveaux de gris, où les mesures locales de mouvement jouent un rôle équivalent aux niveaux de gris. Une des principales différences entre ces deux quantités tient dans la nature continue des mesures de mouvement considérées. Différentes raisons motivent la mise en oeuvre d'une quantification de ces quantités de mouvement. Tout d'abord, même si nous spécifions les modèles dans un cadre continu, nous exploiterions en pratique des états discrets pour les aspects d'estimation et de sto-

ckage des modèles. D'autre part, dans le contexte de la reconnaissance du mouvement, l'introduction d'un espace de quantification commun à toutes les séquences d'images traitées se révèle nécessaire pour évaluer des similarités entre ces séquences. Enfin, nous pouvons tirer parti des bornes d'interprétation des mesures locales de mouvement pour proposer un schéma de quantification efficace.

La quantification des mesures locales de mouvement est formulée comme un problème d'étiquetage markovien. Comparée à une simple procédure de quantification linéaire, cette technique markovienne présente plusieurs intérêts. En premier lieu, les mesures de mouvement quantifiées peuvent être vues comme des approximations des amplitudes des déplacements réels (inconnues). Soit Λ l'ensemble des valeurs discrètes des mesures de mouvement quantifiées. Nous posons $\Lambda = \{\nu_0 = 0, \nu_1, \nu_2, \ldots, \nu_{|\Lambda|}\}$ avec $0 < \nu_1 < \ldots < \nu_{|\Lambda|}$. Étant donné un point p, la quantification markovienne vise à déterminer l'intervalle du type $[\nu_{i-1}, \nu_i]$ auquel appartient le plus probablement l'amplitude du déplacement réel (inconnue) en p. Ceci est évalué à travers un terme d'attache aux données fonction de la mesure locale de mouvement $v_{obs}(p)$ et des bornes d'interprétation $\{(l_{\nu_i}(p), L_{\nu_i}(p))\}$ décrites précédemment. Par ailleurs, l'utilisation d'une technique d'étiquetage contextuel permet de rejeter les observation locales erronées. De plus, des expériences menées dans [5] pour des mouvements simples connus (translation, rotation, mouvement divergent) ont démontré que cette quantification markovienne fournissait des mesures locales de mouvement quantifiées plus proches des amplitudes des déplacements réels, par rapport à une simple quantification linéaire. Ces comparaisons ont été évaluées entre la carte des mesures locales de mouvement quantifiées obtenue et la carte des amplitudes quantifiées des déplacements réels (vérité de terrain), en termes d'erreur quadratique moyenne et de distance L_1 entre les histogrammes des amplitudes quantifiées.

Soit $\mathcal{R}$ le support spatial de l'image, $e = (e_p)_{p \in \mathcal{R}}$ le champ des étiquettes où chaque étiquette prend une valeur dans Λ, et $o = (v_{obs}(p))_{p \in \mathcal{R}}$ le champ des observations formées par les mesures locales de mouvement. La quantification markovienne repose sur le critère du Maximum A Posteriori (MAP) et revient à minimiser une fonction d'énergie globale U [9] :

$$\begin{aligned} \widehat{e} &= \arg \min_{e \in \Lambda^{|\mathcal{R}|}} U(e, o) \\ &= \arg \min_{e \in \Lambda^{|\mathcal{R}|}} [U_1(e, o) + U_2(e)] \end{aligned} \tag{3}$$

où la fonction d'énergie $U(e, o)$ est scindée en un terme d'attache aux données $U_1(e, o)$ et un terme de régularisation contextuelle $U_2(e)$. De plus, U_1 et U_2 sont exprimés comme des sommes de potentiels V_1 et V_2 :

$$\left\{ \begin{array}{rcl} U_1(e, o) &=& \displaystyle\sum_{p \in \mathcal{R}} V_1(e_p, v_{obs}(p)) \\ U_2(e) &=& \displaystyle\sum_{(p,q) \in \mathcal{C}} \beta \cdot \rho(e_p - e_q) \end{array} \right. \tag{4}$$

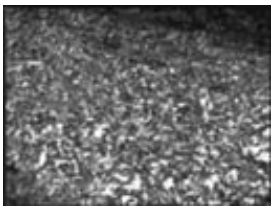
plan de journal télévisé

image originale

carte des mesures locales de mouvement quantifiées

scène de rivière

image originale



carte des mesures locales de mouvement quantifiées

FIG. 2 – *Exemples de cartes de mesures locales de mouvement quantifiées. Nous utilisons une quantification markovienne sur 64 niveaux dans l'intervalle* $[0, 8[$. *Nous présentons des exemples correspondant aux deux premières images de deux séquences : une séquence de plateaux de journal télévisé et une scène de rivière. Les cartes de mesures locales de mouvement quantifiées sont visualisées sur 64 niveaux entre 0 et 255.*

où $\mathcal{C}$ est l'ensemble des cliques binaires du 4-voisinage. β est un coefficient positif qui pondère l'influence relative de la régularisation (en pratique, $\beta = 2.0$). ρ est un M-estimateur fortement redescendant, ici la fonction "biweight" de Tukey. Nous pouvons ainsi préserver les discontinuités présentes dans le champ des déplacements réels. Le potentiel V_1 évalue la pertinence d'une étiquette pour décrire une mesure locale de mouvement donnée. Soit ν_i un niveau de quantification avec $i \in [\![1, |\Lambda|]\!]$, où $[\![1, |\Lambda|]\!]$ est l'intervalle des valeurs discrètes comprises entre 1 et $|\Lambda|$. Le potentiel $V_1(\nu_i, v_{obs}(p))$ quantifie la vraisemblance que l'amplitude du déplacement réel (inconnu) au point p soit dans l'intervalle $[\nu_{i-1}, \nu_i]$. Il est défini par :

$$\begin{aligned} V_1(\nu_i, v_{obs}(p)) &= Sup_{L_{\nu_{i-1}(p)}}(v_{obs}(p)) \\ &+ Inf_{l_{\nu_i(p)}}(v_{obs}(p)) \end{aligned} \tag{5}$$

Sup_L est un échelon continu centré en L et Inf_l est l'opposé d'une fonction échelon centrée en l et translatée sur l'intervalle $[0, 1]$.

La minimisation du critère (3) est effectuée au moyen d'un algorithme ICM modifié et l'initialisation résulte de la seule prise en compte du terme d'attache aux données dans la minimisation. La figure 2 présente deux exemples de cartes de mesures locales de mouvement quantifiées pour une séquence de plateaux de journal télévisé et une scène de rivière. Nous utilisons une quantification markovienne sur

64 niveaux dans l'intervalle $[0, 8[$. Ces deux cartes montrent que le calcul des quantités locales de mouvement fournit des informations sur la présence et la distribution du mouvement que nous pouvons exploiter directement. L'étude de séquences de ces cartes de mesures locales de mouvement nous semble donc appropriée pour accéder à une caractérisation de séquences d'images en termes d'activité de mouvement.

4 Modélisation statistique d'activité de mouvement

4.1 Modèles de Gibbs temporels multi-échelles

Afin de prendre en compte à la fois les aspects spatiaux et temporels du contenu dynamique dans des séquences d'images, nous avons développé un cadre statistique multi-échelle. Étant donné une séquence de cartes de mesures locales de mouvement quantifiées, nous considérons en chaque point non pas une seule quantité scalaire mais un vecteur de mesures évaluées à des échelles successives. Les modèles de Gibbs sont alors spécifiés sur une séquence de cartes de vecteurs multi-échelles de mesures de mouvement. La spécificité des modèles introduits tient dans l'évaluation simple et directe de leur fonction de vraisemblance. Nous pouvons également adopter un schéma direct d'estimation des modèles au sens du Maximum de Vraisemblance (MV).

Considérons une séquence de mesures locales de mouvement $v = (v_0, v_1, \ldots, v_K)$. Nous construisons une nouvelle séquence $x = (x_0, x_1, \ldots, x_K)$. À chaque instant k et pour tout point p dans l'image k, $x_k(p)$ est défini comme un vecteur de mesures $x_k(p) = (x_k^0(p), \ldots, x_k^L(p))$ résultant de lissages gaussiens successifs de variance croissante de la carte de mesures locales de mouvement quantifiées v_k pour les échelles de 0 à L.

Les cartes $\{x_k\}$ de mesures multi-échelles de mouvement ainsi obtenues ne doivent pas être confondues avec le mode de calcul des mesures locales de mouvement des cartes de mesures locales de mouvement quantifiées $\{v_k\}$. Dans le paragraphe précédent, nous avons présenté une méthode de calcul de ces mesures locales de mouvement reposant sur un test de validité de l'ECMA à différentes échelles. Ici, le calcul du vecteur $x_k(p)$ de mesures multi-échelles de mouvement vise à traduire indirectement la distribution spatiale du mouvement autour du pont p par la prise en compte d'un certain support spatial à travers l'opération de lissage.

La modélisation statistique considérée repose sur l'hypothèse que la séquence x est la réalisation d'une chaîne de Markov du premier ordre $X = (X_0, \ldots, X_k)$ telle que :

$$P_{\mathcal{M}}(x) = P_{\mathcal{M}}(x_0) \prod_{k=1}^{K} P_{\mathcal{M}}(x_k|x_{k-1}) \tag{6}$$

$\mathcal{M}$ corresponds au modèle statistique de mouvement sous-jacent qui sera explicitement spécifié par la suite. $P_{\mathcal{M}}(x_0)$ représente l'*a priori* sur la distribution pour la première image de la séquence. En pratique, nous n'avons aucun *a priori*, c.a.d, $P_{\mathcal{M}}(x_0)$ est constante. Nous notons $1/Z$ la valeur de cette constante. Afin de définir des modèles purement causaux, nous supposons que les variables aléatoires $\{X_k(p)\}_{p\in\mathcal{R}}$ à l'instant k sont indépendants conditionnellement à X_{k-1}. En outre, étant donné un point p et un instant k, nous faisons l'hypothèse que $X_k(p)$ est également indépendante de $\{X_{k-1}(q)\}_{q\in\mathcal{R}\setminus\{p\}}$ conditionnellement à $X_{k-1}(p)$. Ainsi, $P_{\mathcal{M}}(x_k|x_{k-1})$ est donné par :

$$\begin{aligned} P_{\mathcal{M}}(x_k|x_{k-1}) &= \prod_{p\in\mathcal{R}} P_{\mathcal{M}}\left(x_k(p)|x_{k-1}\right) \\ &= \prod_{p\in\mathcal{R}} P_{\mathcal{M}}\left(x_k(p)|x_{k-1}(p)\right) \end{aligned} \tag{7}$$

Pour $(k,p) \in [\![1, K]\!] \times \mathcal{R}$, nous appliquons la relation de Bayes sachant que $x_k(p) = (x_k^0(p), \ldots, x_k^L(p))$, et nous obtenons l'expression suivante :

$$\begin{aligned} &P_{\mathcal{M}}\left(x_k(p)|x_{k-1}(p)\right) = \\ &P_{\mathcal{M}}\left(x_k^0(p)|x_k^L(p), x_k^{L-1}(p), \ldots, x_k^1(p), x_{k-1}(p)\right) \\ &\times \ldots \times P_{\mathcal{M}}\left(x_k^{L-1}(p)|x_k^L(p), x_{k-1}(p)\right) \\ &\times P_{\mathcal{M}}\left(x_k^L(p)|x_{k-1}(p)\right) \end{aligned} \tag{8}$$

Puisque $\{x_k^0(p), \ldots, x_k^L(p)\}$ sont des quantités locales de mouvement calculées à différentes échelles, les quantités relatives aux niveaux d'échelle les plus fins fournissent des informations précises et très localisées, alors que celles relatives à des niveaux plus grossiers captent du fait des filtrages successifs des caractéristiques un peu plus "étendues". En termes de dépendance conditionnelle, ceci nous amène à postuler que, pour tout point p à tout instant k et tout niveau d'échelle $l \in [\![0, L-2]\!]$, $X_k^l(p)$ est indépendant de $X_k^{l+2}(p), ..., X_k^L(p)$ conditionnellement à $X_k^{l+1}(p)$. De même, pour ce qui est des dépendances conditionnelles de $X_k^l(p)$ sachant $X_{k-1}(p)$, l'information la plus pertinente est associée à la mesure $x_{k-1}^0(p)$ à l'échelle 0. À partir de ces deux hypothèses, la relation (8) se simplifie de la manière suivante :

$$\begin{aligned} &P_{\mathcal{M}}\left(x_k(p)|x_{k-1}(p)\right) = \\ &P_{\mathcal{M}}\left(x_k^0(p)|x_k^1(p), x_{k-1}^0(p)\right) \\ &\times \ldots \times P_{\mathcal{M}}\left(x_k^{L-1}(p)|x_k^L(p), x_{k-1}^0(p)\right) \\ &\times P_{\mathcal{M}}\left(x_k^L(p)|x_{k-1}^0(p)\right) \end{aligned} \tag{9}$$

Cette formulation statistique implique l'évaluation de "trioccurrences", ce qui induit une complexité importante pour spécifier explicitement le modèle $\mathcal{M}$. En outre, nous avons noté en pratique que les cooccurrences en échelle évaluées pour des paires de mesures $\{(x_k^{l-1}(p), x_k^l(p))\}$ à deux échelles successives $l-1$ et l prennent des valeurs d'autant plus grandes qu'il s'agit de termes proches de la diagonale.

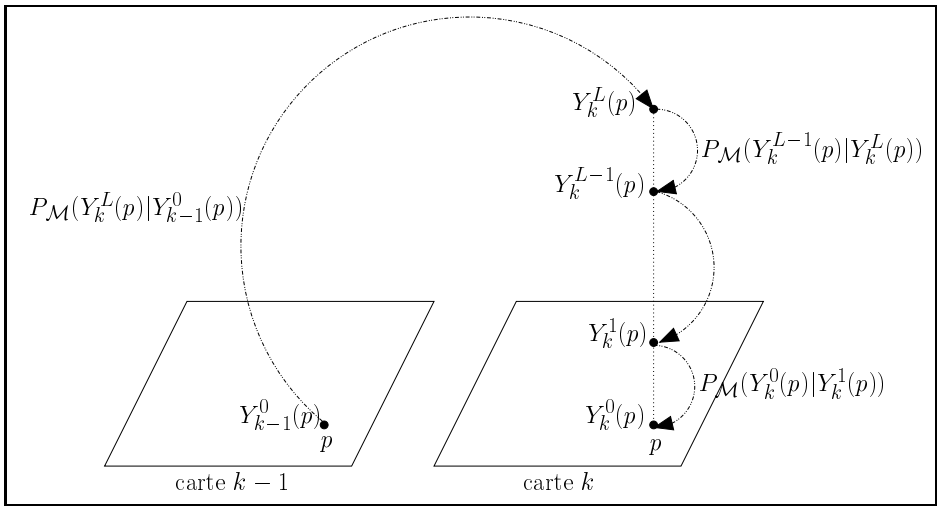


FIG. 3 – *Illustration des dépendances conditionnelles introduites dans les modèles de Gibbs temporels multi-échelles. En chaque point* (p,k), $Y_k(p) = \{Y_k^0(p), \ldots, Y_k^L(p)\}$ *est un vecteur de variables aléatoires correspondant à différentes échelles de 0 à* L. $P_{\mathcal{M}}(Y_k^L(p)|Y_{k-1}^0(p))$ *et* $\{P_{\mathcal{M}}(Y_k^l(p)|Y_k^{l+1}(p))\}_{l\in[\![0,\ldots,L-1]\!]}$ *sont les probabilités conditionnelles de transition prises en compte par le modèle multi-échelle d'activité de mouvement* $\mathcal{M}$.

Les distributions de cooccurrences temporelles calculées pour des paires $\{(x_k^l(p), x_{k-1}^0(p))\}$ se révèlent par contre plus étalées. Ainsi, les dépendances temporelles peuvent être négligées devant les dépendances en échelle. La vraisemblance conditionnelle $P_{\mathcal{M}}\left(x_k(p)|x_{k-1}(p)\right)$ est finalement donnée par :

$$P_{\mathcal{M}}\left(x_k(p)|x_{k-1}(p)\right) = P_{\mathcal{M}}\left(x_k^0(p)|x_k^1(p)\right) \times \ldots \times P_{\mathcal{M}}\left(x_k^{L-1}(p)|x_k^L(p)\right) \times P_{\mathcal{M}}\left(x_k^L(p)|x_{k-1}^0(p)\right) \quad (10)$$

De cette façon, les modèles statistiques de mouvement reposent uniquement sur le calcul de distributions de cooccurrences évaluées, soit à des échelles successives, soit à deux instants successifs entre les échelles 0 et L. La figure 3 en fournit une illustration. Nous pouvons noter au passage que des statistiques évaluées à des niveaux d'échelle successifs apparaissent comme des caractéristiques importantes pour l'analyse et la synthèse de texture [3, 13, 17].
Dans le but de proposer une formulation exponentielle de la vraisemblance $P_{\mathcal{M}}(x)$, nous introduisons les notations suivantes :

$$P_{\mathcal{M}}\left(x_k^L(p)|x_{k-1}^0(p)\right) \propto \exp \Psi_{\mathcal{M}}^L\left(x_k^L(p), x_{k-1}^0(p)\right) \quad (11)$$

et $\forall l \in [\![0, L-1]\!]$:

$$P_{\mathcal{M}}\left(x_k^{L-1}(p)|x_k^L(p)\right) \propto \exp \Psi_{\mathcal{M}}^l\left(x_k^l(p), x_k^{l+1}(p)\right) \quad (12)$$

où $\Psi_{\mathcal{M}} = \left\{\Psi_{\mathcal{M}}^l(\nu,\nu')\right\}_{(l,\nu,\nu')\in[\![0,L]\!]\times\Lambda^2}$ sont les potentiels qui spécifient explicitement le modèle $\mathcal{M}$. Afin de garantir l'unicité des potentiels associés à la loi $P_{\mathcal{M}}$, nous imposons la contrainte de normalisation suivante :

$$\forall (l,\nu') \in [\![0,L]\!] \times \Lambda, \ \sum_{\nu\in\Lambda} \exp \Psi_{\mathcal{M}}^l(\nu,\nu') = 1 \quad (13)$$

En utilisant ces potentiels, la vraisemblance $P_{\mathcal{M}}(x)$ s'écrit :

$$P_{\mathcal{M}}(x) = \frac{1}{Z} \exp\left[\sum_{k=1}^{K} \sum_{p\in\mathcal{R}} \Psi_{\mathcal{M}}\left(x_k(p), x_{k-1}(p)\right)\right] \quad (14)$$

où $\Psi_{\mathcal{M}}\left(x_k(p), x_{k-1}(p)\right)$ est la somme des potentiels temporels et en échelle qui suit :

$$\Psi_{\mathcal{M}}(x_k(p), x_{k-1}(p)) = \Psi_{\mathcal{M}}^L(x_k^L(p), x_{k-1}^0(p)) + \sum_{l=0}^{L-1} \Psi_{\mathcal{M}}^l(x_k^l(p), x_k^{l+1}(p)) \quad (15)$$

La spécification de $\Psi_{\mathcal{M}}$ fournit une connaissance complète de la loi $P_{\mathcal{M}}$. Ceci nous permet de proposer un cadre statistique général pour formuler le problème de reconnaissance du mouvement. De plus, la relation (14) montre que le modèle d'activité de mouvement introduit $\mathcal{M}$ est un modèle de Gibbs pour lequel la fonction de partition est connue et vaut Z. Cette constante est indépendante du modèle $\mathcal{M}$ spécifié.
D'autre part, nous pouvons fournir une formulation exponentielle de l'expression (14) à partir de distributions de cooccurrences temporelles ou en échelle. La vraisemblance $P_{\mathcal{M}}(x)$ se déduit simplement du calcul du produit scalaire $\Psi_{\mathcal{M}} \bullet \Gamma(x)$ entre les potentiels associés au modèle $\mathcal{M}$ et l'ensemble des distributions de cooccurrences temporelles ou en échelle évaluées sur la séquence x de cartes de vecteurs multi-échelles de mesure de mouvement. Nous obtenons en fait :

$$P_{\mathcal{M}}(x) = \frac{1}{Z} \cdot \exp\Big[\Psi_{\mathcal{M}} \bullet \Gamma(x)\Big] \quad \text{avec } \Psi_{\mathcal{M}} \bullet \Gamma(x) = \sum_{l=0}^{l=L} \Psi_{\mathcal{M}}^l \bullet \Gamma^l(x) \quad (16)$$

où $\Psi^l_{\mathcal{M}} \bullet \Gamma^l(x)$ est le produit scalaire entre la distribution de cooccurrences temporelles (par convention pour $l = L$, d'après la relation (11)) ou en échelle (pour $l \in [\![0, L-1]\!]$, d'après la relation (12)) et les potentiels correspondant $\Psi^l_{\mathcal{M}}$. La distribution de cooccurrences temporelles est définie par : $\forall(\nu, \nu') \in \Lambda^2$,

$$\Gamma^L(\nu, \nu'|x) = \sum_{k=1}^{K} \sum_{p \in \mathcal{R}} \delta(\nu - x_k^L(p))\delta(\nu' - x_{k-1}^0(p)) \quad (17)$$

où δ est le symbole de Kronecker. La distribution de cooccurrences en échelle $l \in [\![0, L-1]\!]$ est donnée par : $\forall(\nu, \nu') \in \Lambda^2$,

$$\Gamma^l(\nu, \nu'|x) = \sum_{k=1}^{K} \sum_{p \in \mathcal{R}} \delta(\nu - x_k^l(p))\delta(\nu' - x_k^{l+1}(p)) \quad (18)$$

Étant donné $l \in [\![0, L]\!]$, le produit scalaire $\Psi^l_{\mathcal{M}} \bullet \Gamma^l(x)$ s'exprime comme suit :

$$\Psi^l_{\mathcal{M}} \bullet \Gamma^l(x) = \sum_{(\nu,\nu') \in \Lambda^2} \Psi^l_{\mathcal{M}}(\nu, \nu') \cdot \Gamma^l(\nu, \nu'|x) \quad (19)$$

Cette expression exponentielle de la vraisemblance $P_{\mathcal{M}}(x)$ est intéressante à plusieurs titres. Tout d'abord, elle montre *in fine* que le calcul de cette vraisemblance pour tout modèle $\mathcal{M}$ et toute séquence x est immédiat et simple en pratique. L'utilisation des modèles statistiques est alors directe pour des problèmes de reconnaissance ou de classification du mouvement selon des critères MV ou MAP. Ensuite, l'ensemble de l'information de mouvement exploitée par ces modèles est portée par les distributions de cooccurrences temporelles et en échelle. En particulier, s'il est nécessaire de calculer la vraisemblance $P_{\mathcal{M}}(x)$ d'une séquence donnée x pour plusieurs modèles $\{\mathcal{M}_i\}$, il n'est pas nécessaire de stocker cette séquence. Il nous suffit de déterminer et de stocker l'ensemble des distributions de cooccurrences $\Gamma(x)$. Le calcul des vraisemblances $\{P_{\mathcal{M}_i}(x)\}$ se ramènent alors simplement à l'évaluation des produits $\{\Psi_{\mathcal{M}_i} \bullet \Gamma(x)\}$ selon la relation (16).

4.2 Estimation des modèles au sens du maximum de vraisemblance

Nous présentons dans ce paragraphe la méthode d'estimation du modèle statistique de mouvement associé à une séquence d'images. Étant donné une séquence de cartes de vecteurs multi-échelles de quantités de mouvement, nous estimons les potentiels $\{\Psi^l_{\widehat{\mathcal{M}}}(\nu, \nu')\}_{(l,\nu,\nu') \in [\![0,L]\!] \times \Lambda^2}$ du modèle $\widehat{\mathcal{M}}$ qui décrit le mieux la séquence x. Nous considérons le critère du Maximum de Vraisemblance (MV), qui nous amène à résoudre le problème suivant :

$$\widehat{\mathcal{M}} = \arg\max_{\mathcal{M}} P_{\mathcal{M}}(x) \quad (20)$$

Comme la modélisation statistique que nous avons introduite n'implique que des produits de vraisemblances conditionnelles comme le montre la relation (10), l'estimation au sens du MV consiste simplement à évaluer empiriquement ces vraisemblances conditionnelles (ou transitions). L'estimé au sens du MV des potentiels du modèle $\mathcal{M}$ est donné par : $\forall(l, \nu, \nu') \in [\![0, L]\!] \times \Lambda^2$,

$$\Psi^l_{\widehat{\mathcal{M}}}(\nu, \nu') = \log\left(\Gamma^l(\nu, \nu'|x) / \sum_{\nu'' \in \Lambda} \Gamma^l(\nu'', \nu'|x)\right) \quad (21)$$

Ainsi, l'estimation au sens du MV du modèle d'activité de mouvement $\mathcal{M}$ relatif à une séquence x se déduit directement de l'ensemble des cooccurrences temporelles ou en échelle $\Gamma(x)$. De plus, nous pouvons envisager de réduire la complexité des modèles dans le but de fournir une représentation plus parcimonieuse du mouvement. Pour ce faire, la sélection des potentiels les plus informatifs est basée sur un calcul de rapports de vraisemblance de manière analogue à la technique décrite dans [7].

5 Reconnaissance du mouvement

Pour démontrer la capacité des modèles statistiques non paramétriques de mouvement à appréhender et discriminer des formes de mouvement variées, nous avons mené des tests de reconnaissance sur un ensemble de séquences d'images associées à une large gamme de contenus dynamiques.

5.1 Base de séquences d'images

La base de séquences d'images considérées comprend diverses situations de texture temporelle, des exemples de mouvement rigide et des déplacements de piétons. Plus précisément, elle contient quatre types de texture temporelle : des mouvements d'herbe (A), des scènes de mer calme (B), des scènes de rivière (C), des scènes d'arbre en présence de vent (D). D'autre part, une classe de séquences de plateaux de journaux télévisées (E), et deux classes de situations de mouvement plutôt rigide, des escaliers mécaniques (F) et des séquences de trafic routier (G), sont également incluses. La dernière classe (H) comprend des exemples de piétons marchant soit de la gauche vers la droite, soit de la droite vers la gauche. Nous avons ainsi une base de test comprenant huit classes différentes.

Chaque classe de mouvement, exceptée la classe (H), est représentée par trois séquences de cent images. La classe (H) contient dix séquences de trente images (cinq exemples de déplacement de piétons de la gauche vers la droite, et cinq de piétons marchant de la droite vers la gauche). La figure 4 présente une image pour chaque séquence des classes (A) à (G). Pour la classe (H), nous avons sélectionné des images de trois séquences.

5.2 Méthodes d'apprentissage et de reconnaissance

À partir des huit classes de mouvement, nous réalisons dans un premier temps une phase d'apprentissage sur un ensemble de séquences d'images. Ensuite, nous menons des

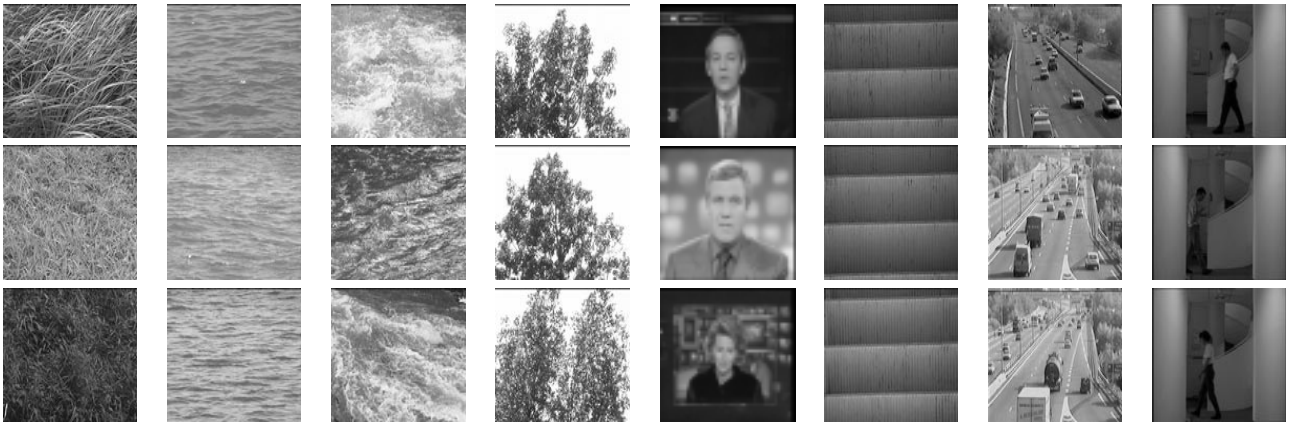


FIG. 4 – *Base test de séquences d'images. Nous fournissons trois images représentatives des séquences de chaque classe de mouvement (A) à (H). Ces classes correspondent aux contenus dynamiques qui suivent : (A) mouvement d'herbe, (B) scènes de mer calme, (C) scènes de rivière, (D) scènes d'arbre en présence de vent, (E) scènes de plateaux de journaux télévisés, (F) scènes d'escalier mécanique, (G) scènes de trafic autoroutier, (H) déplacements de piétons.*

expériences de reconnaissance du mouvement sur un ensemble test différent de l'ensemble d'apprentissage. Ces deux ensembles de séquences d'images sont définis de la manière suivante.

Chaque séquence d'images de la base est subdivisée en "micro-séquences" de 6 images. Nous obtenons ainsi un échantillon de 57 éléments pour représenter chaque classe. Nous disposons donc globalement d'un ensemble de 456 micro-séquences. Les dix premières micro-séquences de la première séquence des classes de (A) à (G) sont utilisées comme données d'apprentissage. Pour la classe (H), comme les séquences ne contiennent que trente images, nous considérons les cinq premières micro-séquences de deux séquences de cette classe. Finalement, nous disposons d'une base d'apprentissage de 80 éléments et d'un ensemble de test comprenant 376 micro-séquences. Nous notons $\mathcal{C}$ l'ensemble des huit classes de mouvement, $\mathcal{A}_c$ l'ensemble d'apprentissage associé à une classe donnée $c \in \mathcal{C}$ et $\mathcal{T}$ l'ensemble de test.

Étant donné une classe $c \in \mathcal{C}$, la phase d'apprentissage consiste à déterminer le modèle statistique de mouvement associé $\mathcal{M}_c$. Pour chaque élément $a \in \mathcal{A}_c$, nous calculons la séquence de cartes de vecteurs multi-échelles de mesures locales de mouvement x^a et l'ensemble correspondant des distributions de cooccurrences temporelles et en échelle $\Gamma(x^a)$. Nous estimons alors le modèle $\mathcal{M}_c$ associé à l'ensemble d'observations $\{x^a\}_{a\in\mathcal{A}_c}$ au sens du MV. Nous considérons donc le critère suivant :

$$\mathcal{M}_c = \arg\max_{\mathcal{M}} \left[\prod_{a\in\mathcal{A}_c} P_{\mathcal{M}}(x^a) \right] \tag{22}$$

En utilisant la forme exponentielle de la loi $P_{\mathcal{M}}(x^a)$, donnée par la relation (16), nous obtenons :

$$\mathcal{M}_c = \arg\max_{\mathcal{M}} \left[\sum_{a\in\mathcal{A}_c} \Psi_{\mathcal{M}} \bullet \Gamma(x^a) \right] \tag{23}$$

Comme le produit scalaire $\Psi_{\mathcal{M}} \bullet \Gamma(x^a)$ est linéaire vis à vis des distributions de cooccurrences, ce critère équivaut à :

$$\mathcal{M}_c = \arg\max_{\mathcal{M}} \left[\Psi_{\mathcal{M}} \bullet \sum_{a\in\mathcal{A}_c} \Gamma(x^a) \right] \tag{24}$$

Ainsi, la résolution du critère (22) revient à effectuer l'estimation au sens du MV du modèle correspondant à la distribution de cooccurrences moyenne Γ_c sur l'ensemble des distributions de cooccurrences $\{\Gamma(x^a)\}_{a\in\mathcal{A}_c}$:

$$\mathcal{M}_c = \arg\max_{\mathcal{M}} [\Psi_{\mathcal{M}} \bullet \Gamma_c] \tag{25}$$

avec : $\forall (l,\nu,\nu') \in [\![0,L]\!] \times \Lambda^2$,

$$\Gamma_c^l(\nu,\nu') = \sum_{a\in\mathcal{A}_c} \Gamma^l(\nu,\nu'|x^a) \tag{26}$$

Les potentiels $\Psi_{\mathcal{M}_c}$ sont alors directement déduits de Γ_c à partir de la relation (21).
Nous utilisons cet ensemble de modèles statistiques de mouvement $\{\mathcal{M}_c\}_{c\in\mathcal{C}}$ pour formuler la reconnaissance du mouvement comme un problème d'inférence statistique selon le critère du MV. Étant donné un élément t de l'ensemble de test $\mathcal{T}$, nous calculons la séquence correspondante de cartes de mesures multi-échelles de mouvement x^t et les distributions de cooccurrences temporelles et en échelles associées $\Gamma(x^t)$. Pour déterminer la classe de mouvement c^t de l'élément t, nous exploitons le critère du MV de la manière suivante :

$$\begin{aligned} c^t &= \arg\max_{c\in\mathcal{C}} P_{\mathcal{M}_c}(x^t) \\ &= \arg\max_{c\in\mathcal{C}} \left[\Psi_{\mathcal{M}_c} \bullet \Gamma(x^t) \right] \end{aligned} \tag{27}$$

Il suffit donc d'évaluer huit produits scalaires $\Psi_{\mathcal{M}_c} \bullet \Gamma(x^t)$ entre les potentiels des modèles $\{\Psi_{\mathcal{M}_c}\}_{c\in\mathcal{C}}$ relatifs à chaque classe de mouvement et les distributions de cooccurrences $\Gamma(x^t)$.

5.3 Résultats expérimentaux

Les évaluations expérimentales ont été menées avec les valeurs de paramètres suivantes. La quantification des mesures locales de mouvement est effectuée sur 64 niveaux dans l'intervalle $[0, 8]$. Nous considérons des valeurs du nombre L de niveaux d'échelle de 0 à 4. La technique de réduction de la complexité des modèles permet de ne conserver que 10% à 20% de potentiels informatifs (c.a.d., de l'ordre de 1000 potentiels pour spécifier chaque modèle). De manière générale, les potentiels retenus correspondent à des niveaux de mouvement peu élevés et sont souvent associés à des mesures de cooccurrences proche de la diagonale de la matrice de cooccurrences.

Dans le cas où $L = 0$, la modélisation se réduit à une version mono-échelle. Il n'y a donc alors aucune information spatiale qui soit explicitement intégrée et les modèles sont spécifiés uniquement à partir de distributions de cooccurrences temporelles. Ces modèles sont des modèles de Gibbs Temporels (GT), alors que, pour $L \geq 1$, il s'agit de modèles de Gibbs Temporels Multi-Echelles (GTME). Dans la suite, les méthodes de reconnaissance du mouvement associées à chaque type de modèles sont respectivement dénommées la méthode GT et la méthode GTME. La comparaison de ces deux méthodes nous permettra d'évaluer l'intérêt d'une caractérisation simultanée des aspects spatiaux et temporels du mouvement par le biais d'une modélisation multi-échelle.

La figure 5 présente la moyenne τ et l'écart-type $\Delta\tau$, sur les huit classes de mouvement, du taux de reconnaissance pour les éléments de l'ensemble de test $\mathcal{T}$. Nous considérons les méthodes GT et GTME avec un à quatre niveaux d'échelle. En utilisant les modèles GTME, le taux moyen τ de reconnaissance est toujours supérieur à 95%, alors qu'il n'est que de 92.4% en exploitant les modèles GT. Les meilleurs résultats sont obtenus en considérant les modèles GTME avec trois niveaux d'échelle ($L = 3$). Le taux moyen de reconnaissance est alors de l'ordre de 99% avec un écart-type inférieur à 1%. Par conséquent, la prise en compte des aspects spatiaux et temporels du mouvement par le biais d'une approche multi-échelle se traduit par une précision de caractérisation nettement accrue comparativement aux modèles uniquement temporels. Les moins bons résultats obtenus pour $L = 2$ comparativement aux cas $L = 1$ et $L = 3$ sont surprenants. Ceci peut suggérer qu'il serait pertinent de procéder à une sélection automatique du niveau L en fonction du contenu de chaque séquence, par exemple, au moyen d'une technique de sélection d'ordre de modèle. Par ailleurs, nous avons remarqué que les performances se dégradent au delà de 3 niveaux d'échelle. Ceci est vraisemblablement dû aux deux facteurs suivants. Tout d'abord, les poids des termes proches de la

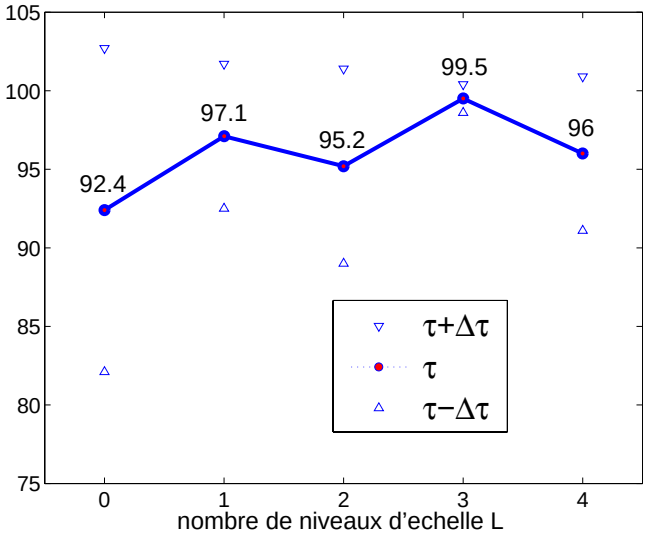


FIG. 5 – *Résultats de reconnaissance du mouvement pour la base de séquences d'images présentées en figure 4. Il s'agit des résultats obtenus avec les modèles de Gibbs Temporels (GT) ($L = 0$) et avec les modèles de Gibbs Temporels Multi-Echelles (GTME) avec $L \in \{1, 4\}$. Nous présentons la moyenne et l'écart-type du taux de reconnaissance calculés sur les huit classes de mouvement.*

diagonale augmentent avec le niveau d'échelle dans les distributions de cooccurrences en échelle $\Gamma^l(x)$. Au delà d'un certain niveau d'échelle, les distributions de cooccurrences en échelle n'apportent donc plus d'information supplémentaire. Parallèlement, plus le nombre de niveaux d'échelle L augmente, plus les informations captées par la distribution de cooccurrences temporelles $\Gamma^L(x)$ sont lissées, ce qui tend à réduire le pouvoir discriminant de ces statistiques.

Le tableau 1 présente en détails les résultats de reconnaissance pour les méthodes GT et GTME avec $L = 3$. Dans les deux cas, nous donnons les taux de bonne et mauvaise classification pour chaque classe de mouvement. La comparaison de ces méthodes montrent que la seconde est toujours la plus performante. En fait, en utilisant les modèles GTME, le taux de reconnaissance est toujours supérieur à 97%, alors qu'il est compris entre 69.6% et 100% pour les modèles GT. Les améliorations les plus significatives sont obtenues pour les classes (A) et (E). Les taux de reconnaissance passent respectivement de 83% à 97.9% et de 69.6% à 100%. Dans ce deuxième cas, 28.3% des séquences de test sont attribuées à la classe (D) par la méthode GT. En fait, les scènes de journaux télévisés de la classe (E) contiennent une faible activité de mouvement avec des déplacements peu importants des présentateurs. Les séquences de la classe (D) incluent des mouvements de feuilles de faible amplitude. La prise en compte des aspects spatiaux et temporels de la distribution du mouvement nous permet de discriminer parfaitement ces deux classes.

	A	B	C	D	E	F	G	H
A	**97.9**		**2.1**					
	83.0	*4.3*						*12.7*
B		**100.**						
		100.						
C			**100.**					
			100.					
D				**97.9**				**2.1**
				91.5	*2.1*		*6.4*	
E					**100.0**			
	2.1			*28.3*	*69.6*			
F						**100.**		
		2.1				*97.9*		
G							**100.**	
							100.0	
H								**100.0**
					2.4			*97.6*

TAB. 1 – *Pourcentage de bonne et mauvaise classification pour les huit classes de mouvement. Nous présentons les résultats obtenus à partir de la méthode GT et de la méthode GTME avec $L = 3$. Pour chaque classe, la première ligne (en gras) correspond à la méthode GTME (par exemple, pour la classe (A), le pourcentage de séquences de test attribuées aux classes (A) et (C) est respectivement de 97.9% et 2.1%) alors que la seconde ligne (en italique) est relative à la méthode GT.*

6 Conclusion

Nous avons présenté une méthode de modélisation statistique non paramétrique du mouvement dans des séquences d'images. Elle appréhende simultanément des aspects spatiaux et temporels du mouvement. Elle est basée sur des modèles de Gibbs temporels multi-échelles. La nature causale de ces modèles rend possible l'évaluation exacte et simple de leur fonction de vraisemblance. L'estimation des modèles au sens du maximum de vraisemblance est alors directe. De plus, nous pouvons exploiter ces modèles pour la reconnaissance du mouvement spécifiée comme un problème d'inférence statistique.
Cette technique d'analyse non paramétrique du mouvement permet de considérer une large gamme de situations dynamiques (des mouvements rigides aux textures temporelles). Nous avons obtenu des résultats très satisfaisants en reconnaissance du mouvement.

Références

[1] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proc. of the IEEE*, 76(8):869–890, 1988.

[2] A. Bobick. Movement, activity, and action: The role of knowledge in the perception of motion. *Phil. Trans. Royal Society London B*, pages 1257–1265, 1997.

[3] J.S. De Bonet and P. Viola. Texture recognition using a non-parametric multi-scale statistical model. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'98*, pages 641–647, Santa-Barbara, June 1998.

[4] J.W. Davis and A. Bobick. The representation and recognition of human movement using temporal templates. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'97*, pages 928–934, Porto-Rico, June 1997.

[5] R. Fablet. Modélisation statistique non paramétrique et reconnaissance du mouvement dans des séquences d'images ; application à l'indexation vidéo. *Thèse Université de Rennes 1, Irisa No. 2526*, 2001.

[6] R. Fablet and P. Bouthemy. Motion-based feature extraction and ascendant hierarchical classification for video indexing and retrieval. In *Proc. of 3rd Int. Conf. on Visual Information Systems, VISUAL'99*, LNCS Vol 1614, pages 221–228, Amsterdam, June 1999. Springer.

[7] R. Fablet, P. Bouthemy, and P. Pérez. Statistical motion-based video indexing and retrieval. In *Proc. of 6th Int. Conf. on Content-Based Multimedia Information Access, RIAO'2000*, pages 602–619, Paris, Apr. 2000.

[8] C. Fermuller and Y. Aloimonos. Vision and action. *Image and Vision Computing*, 13(10):725–744, 1995.

[9] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. on PAMI*, 6(6):721–741, 1984.

[10] G.L. Gimel'Farb. Texture modeling by multiple pairwise pixel interactions. *IEEE Trans. on PAMI*, 18(11):1110–1114, 1996.

[11] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using Markov random fields. *IEEE Trans. on PAMI*, 15(2):1217–1232, 1993.

[12] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, 1981.

[13] J. Huang and D. Mumford. Statistics of natural images and models. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'99*, pages 541–547, Fort Collins, June 1999.

[14] R. Nelson and R. Polana. Qualitative recognition of motion using temporal texture. *CVGIP*, 56(1):78–99, 1992.

[15] J.M. Odobez and P. Bouthemy. Separation of moving regions from background in an image sequence acquired with a mobile camera. In *Video Data Compression for Multimedia Computing*, chapter 8, pages 295–311. H. H. Li, S. Sun, and H. Derin, eds, Kluwer, 1997.

[16] C.-H. Peh and L.-F. Cheong. Exploring video content in extended spatio-temporal textures. In *Workshop on Content-Based Multimedia Indexing, CBMI'99*, pages 147–153, Toulouse, France, Oct. 1999.

[17] J. Portilla and E. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. Jal of Comp. Vis.*, 40(1):49–70, 2000.

[18] M. Szummer and R.W. Picard. Temporal texture modeling. In *Proc. of 3rd IEEE Int. Conf. on Image Processing, ICIP'96*, pages 823–826, Lausanne, Sept. 1996.

[19] R.P. Wildes and J.R. Bergen. Qualitative spatiotemporal analysis using an oriented energy representation. In *Proc. of 6th Eur. Conf. on Computer Vision, ECCV'2000*, pages 768–784, Dublin, June 2000.

[20] S.C. Zhu, T. Wu, and D. Mumford. Filters, random fields and maximum entropy (FRAME) : towards a unified theory for texture modeling. *Int. Jal of Comp. Vis.*, 27(2):107–126, 1998.