HAL
open science

Indexation de Textures Dynamiques à l'aide de Décompositions Multi-échelles

Sloven Dubois, Renaud Péteri, Ménard Michel

To cite this version:

Sloven Dubois, Renaud Péteri, Ménard Michel. Indexation de Textures Dynamiques à l'aide de Décompositions Multi-échelles. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3. hal-00656501

HAL Id: hal-00656501

https://hal.science/hal-00656501v1

Submitted on 17 Jan 2012

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

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

Indexation de Textures Dynamiques à l'aide de Décompositions Multi-échelles

Sloven Dubois[1] Renaud Péteri[2] Michel Ménard[3]

[1] Université de Lyon, F-42023, Saint-Étienne, France
CNRS, UMR5516, Laboratoire Hubert Curien, F-42000, Saint-Étienne, France
Université de Saint-Étienne, Jean Monnet, F-42000, Saint-Étienne, France
[2] Laboratoire Mathématiques, Image et Applications, La Rochelle, France
[3] Laboratoire Informatique, Image et Interaction, La Rochelle, France

sloven.dubois@univ-st-etienne.fr, renaud.peteri@univ-lr.fr, michel.menard@univ-lr.fr

Résumé

Ce papier présente six algorithmes de décomposition multi-échelle spatio-temporelle pour la caractérisation de textures dynamiques. L'objectif est de comparer leur comportement et leur performance sur un problème d'indexation. Ce travail présente notamment une comparaison entre la seule méthode existante [23] dans ce contexte d'étude et cinq nouvelles approches de décomposition spatio-temporelles.

Les algorithmes sont présentés et appliqués avec succès sur trois bases conséquentes de textures dynamiques disponibles en ligne. La construction et la pertinence des vecteurs caractéristiques sont étudiées. La performance des méthodes d'analyse est ensuite discutée. Enfin, des perspectives de recherche sont évoquées.

Mots Clef

Indexation, textures dynamiques, décompositions multi-échelles spatio-temporelles

Abstract

This paper presents six spatio-temporal multiscale decompositions for characterizing dynamic textures. The main goal of this work is to compare the influence of these analysis methods on the indexation of dynamic textures. These methods are grouped into three approaches and the novelty of this work is to establish a comparison between the only existing method [23] and five other spatio-temporal decompositions.

The three approaches are presented and successfully applied on three large dynamic texture databases. Feature vectors are described as well as their relevance, and performances of the methods are discussed.
Finally, future prospects are exposed.

Keywords

Indexing, dynamic textures, Spatio-temporal multiscale decompositions

1 Introduction

L'analyse des textures dynamiques est un thème encore émergeant dans le domaine de l'analyse de texture. Il s'agit de l'extension au domaine temporel de la notion de texture pour le domaine de l'image. Un champ d'herbe ondulant, les vagues de la mer, un drapeau dans le vent, le mouvement d'arbres dans une forêt, de la fumée, du feu, ou bien la séquence d'images de la figure 1, sont autant d'exemples de textures dynamiques présentés dans la littérature.

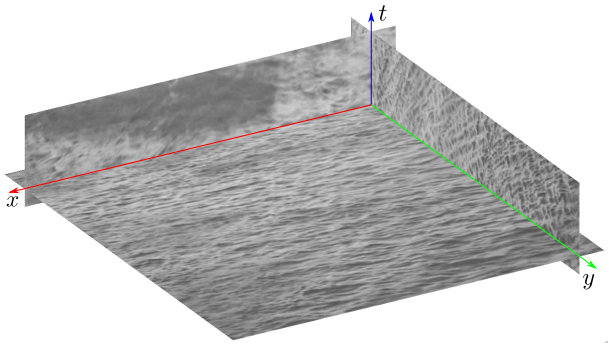

FIG. 1 – *Coupes 2D+T d'une texture dynamique : la surface d'un lac.*

Une texture dynamique peut se définir comme un phénomène variant spatio-temporellement et possédant une répétitivité à la fois spatiale et temporelle. Elles représentent un thème de recherche actuel ; le nombre de publications portant sur la caractérisation de textures dynamiques est en forte augmentation ces dernières années. La première publication sur ce domaine apparait en 1992 avec l'article de Nelson et Polana [17], dans lequel les auteurs mentionnent les textures temporelles, et définissent un premier cadre formel d'étude. Par la suite, à partir de l'année 2002, le nombre de publications proposant des méthodes pour les caractériser augmente considérablement.

L'émergence de ce sujet de recherche peut s'expliquer, d'une part, par la démocratisation du support vidéo et des performances des ordinateurs pouvant les traiter, et d'autre part, par un vaste champ applicatif : indexation vidéo, vidéo surveillance, segmentation spatio-temporelle, soustraction de fonds dynamiques, suivi de textures dynamiques, synthèse de vidéos, *etc*.

Notre contexte de recherche est la reconnaissance et la définition de textures dynamiques [20, 29]. Un aperçu de ce domaine est donné dans [6].

Dans le domaine de la caractérisation de textures dynamiques, les travaux précédents peuvent être organisés selon la taxonomie suivante : les méthodes s'appuyant sur le flot optique [17, 22, 14], celles se basant sur un filtrage spatio-temporel [25], les approches calculant des propriétés géométriques dans le domaine spatio-temporel [18, 30], et les méthodes utilisant des transformées multi-échelles [23, 9].

Comme détaillé dans [11], une texture dynamique est composée de phénomènes évoluant à différentes échelles spatio-temporelles. Une méthode efficace doit permettre d'extraire de tels comportements. Ainsi, la transformée en ondelettes apparaît comme un outil adapté pour l'analyse des textures dynamiques. En effet, dans le cadre de l'analyse de textures statiques, la transformée en ondelettes a été employée avec succès. Par exemple, les ondelettes de Gabor sont utilisées pour calculer les descripteurs de textures dans la norme MPEG-7 [27]. Une idée naturelle est donc d'étendre ces décompositions multi-échelles au domaine temporel afin de caractériser les textures dynamiques.

Dans nos travaux précédents dans le domaine de la texture dynamique, nous nous sommes attachés dans un premier temps à définir et modéliser les textures dynamiques [11, 12]. Puis dans un second temps, nous avons utilisé ce modèle formel de textures dynamiques dans plusieurs applications, notamment l'estimation du mouvement global d'une texture dynamique [11, 10] ou la recherche par le contenu [12]. Dans ce papier, nous nous focaliserons sur l'apport de ce modèle formel au domaine de l'indexation de textures dynamiques.

A notre connaissance, le travail de J.R. Smith *et al.* [23] est la seule méthode multi-échelle utilisée pour caractériser les textures dynamiques. Dans cet article, cette approche sera considérée comme la méthode de référence. Dans le futur, les meilleures approches multi-échelles pour l'indexation de textures dynamiques seront comparées aux approches, par exemple celles basées sur le flot optique, faisant figurent de référence. Dans la section 2, nous présentons trois approches différentes pour analyser et caractériser les textures dynamiques :

- **les méthodes s'appuyant sur des transformées en ondelettes utilisant le produit tensoriel** : quatre décompositions multi-échelles, incluant celle de J.R. Smith *et al.* [23] sont présentées.
- **une méthode fondée sur une décomposition multi-échelle géométrique (transformée en curvelet 2D+T)** : cette décomposition est creuse pour la représentation des fronts d'ondes présents dans les textures dynamiques [5, 8].
- **une méthode décomposant une séquence d'images en composantes morphologiques** : un modèle formel de textures dynamiques est utilisée pour guider cette décomposition.

Dans l'article de J.R. Smith *et al.* [23], les auteurs utilisent pour leurs expérimentations une base de données de petite taille et non publiquement accessible. Ceci est une limitation pour l'étude de la pertinence d'autres vecteurs caractéristiques. Dans la section 3, après la construction de nos vecteurs caractéristiques, nous comparons les différentes approches sur trois bases conséquentes de données disponibles en ligne [21].

2 Analyses spatio-temporelles de textures dynamiques

Dans cette section, différentes approches multi-échelles ainsi que leurs caractéristiques pour l'analyse de textures dynamiques sont étudiées. Chaque méthode représente différemment l'information spatio-temporelle de ce phénomène complexe.

2.1 Transformées en ondelettes utilisant le produit tensoriel

Une texture dynamique est composée de phénomènes existants à différentes échelles spatio-temporelles. Pour capturer ces comportements multi-échelles, la transformée en ondelettes [15] apparaît comme un outil naturel.

Notre première approche est composée de quatre décompositions spatio-temporelles utilisant le produit tensoriel. Elles varient dans la façon dont les variables temporelles et spatiales sont prises en compte dans l'analyse multi-résolution.

- **Transformée en ondelettes spatiale** : une approche intuitive est d'utiliser la transformée en ondelettes image par image. Dans ce cas, la corrélation temporelle entre deux images successives n'est pas prise en compte. Pour chaque image et pour chaque échelle de décomposition, une sous-bande d'approximation et trois sous-bandes de détails sont obtenues.
- **Transformée en ondelettes temporelle** : une seconde approche intuitive est d'effectuer une transformée en ondelettes uniquement dans la direction temporelle. Pour chaque pixel d'une séquence d'images, le profil temporel est extrait et une transformée en ondelettes mono-dimensionnel est effectuée.
- **Transformée en ondelettes 2D+T** : la troisième méthode multi-échelle va effectuer une décomposition à la fois spatiale et temporelle de la séquence d'images. Cette extension au domaine temporel de la transformée en ondelettes 2D est effectuée en utilisant un banc de filtre séparable. Comme dans le cas 2D, une convolution à trois dimensions est décomposée en convolutions mono-dimensionnelles selon les lignes, les colonnes et

les index des images. Pour une vidéo, sept sous-bandes de détails et une sous-bande d'approximation sont obtenues pour chaque échelle.

– **Transformée en ondelettes de J.R. Smith *et al.* [23]** : cette transformée est similaire à la transformée en ondelettes 2D+T exceptée que le filtre temporel est appliqué deux fois à chaque résolution. Ainsi la séquence d'images est décimée deux fois spatialement et temporellement. Les auteurs obtiennent à chaque échelle quinze sous-bandes de détails et une sous-bande d'approximation.

Pour plus d'informations sur ces transformées en ondelettes, nous invitons le lecteur à se référer à [9, 23].

2.2 Décomposition géométrique multi-échelle

La transformée en ondelettes a eu un impact majeur dans de nombreuses applications de traitement du signal ou d'image. Cependant, concernant les images, celle-ci n'est pas optimale pour la représentation et la détection d'objets composés d'éléments anisotropiques, comme les lignes et les courbes. Pour cette raison, ces dernières années ont vu l'apparition de plusieurs transformées multi-échelles géométriques : la transformée en bandelet [16], la transformée en ridgelet [2], la transformée en curvelet [7], etc..

La transformée en curvelet a été conçue pour combler les limitations de la transformée en ondelettes : tandis que les ondelettes capturent les singularités 1D, les curvelets peuvent détecter les structures de dimension supérieure (de co-dimension 1, *i.e.* les courbes dans les images).

Cette transformée géométrique multi-échelle a été récemment étendue à la troisième dimension [28, 4]. Dans des volumes 3D, la transformée en curvelet représente de manière creuse les surfaces.

La transformée en curvelet est originellement dédiée à des données spatiales à trois dimensions (x, y et z). Pour l'appliquer aux séquences d'images, nous introduisons $z = c.t$, où c est une constante permettant de garder l'homogénéité entre les variables spatiales et temporelle. La constante c est homogène à une vitesse et doit être adaptée selon la vidéo considérée.

Dans [8, 3, 11], il a été montré que la transformée en curvelet 2D+T était pertinente pour l'extraction des phénomènes non-locaux se propageant spatio-temporellement. Ainsi, elle semble particulièrement intéressante pour la modélisation des ondes porteuses présentes dans de nombreuses textures dynamiques.

Pour plus d'informations sur la transformée en curvelet 3D, le lecteur peut se référer à [28, 4].

2.3 Décomposition basée sur un modèle formel de textures dynamiques

Comme mentionné précédemment, une texture dynamique est souvent décrite comme un phénomène temporel avec une certaine répétitivité à la fois spatiale et temporelle. Dans [11], inspiré par des travaux en synthèse vidéo et par l'observation de la base de données DynTex [21], nous présentons un modèle formel prenant en compte les caractéristiques d'une famille importante de textures dynamiques. Une texture dynamique Υ_i peut être modélisée comme la superposition d'ondes porteuses à grandes échelles et de phénomènes locaux oscillants :

$$\forall i, \Upsilon_i^{\Omega_i}(\mathbf{x}) = \mathcal{P}_i(\mathbf{x}) + \mathcal{L}_i(\mathbf{x}) \tag{1}$$

où $\mathbf{x} = (x, y, t)^T$ représente les coordonnées d'un voxel dans le cube vidéo, Ω_i le support spatio-temporel de la texture dynamique Υ_i, $\mathcal{P}_i$ et $\mathcal{L}_i$ sont deux fonctions qui décrivent respectivement l'onde porteuse et les phénomènes locaux. Pour plus d'information sur ce modèle, nous invitons le lecteur à se référer à [11].

L'analyse d'une texture dynamique représentée par ce modèle revient donc à la décomposer en phénomènes locaux oscillants et en ondes porteuses. Des travaux récents dans le domaine de la décomposition d'images et de vidéos [1, 24, 26] semblent pertinents pour l'extraction de ces composantes. De part la richesse de son dictionnaire, l'analyse en composantes morphologiques a été retenue. Pour une description complète de cette approche, le lecteur peut se référer à [24].

Un point crucial de l'analyse en composantes morphologiques est le choix du dictionnaire d'analyse. Un choix non judicieux des bases conduit à une décomposition non creuse et non pertinente des différents phénomènes spatio-temporels présents dans la séquence d'images. Il est donc primordial d'associer à chaque composante de notre modèle la base la plus représentative.

Dans [8], nous montrons que la transformée en curvelet 2D+T [4] est représentative des phénomènes non-locaux se propageant temporellement. La seconde partie de notre modèle (phénomènes locaux) peut être extraite à l'aide de la transformée en cosinus locale 2D+T. Le dictionnaire de l'analyse en composantes morphologiques est donc composé de ces deux bases.

Nos expérimentations ont été faites sur les séquences d'images de la base DynTex [21]. Les séquences utilisées ont une durée de 5 secondes (128 images) et une taille de 648×540 pixels. Sur des volumes de cette taille, le temps de calcul des différentes bases de notre dictionnaire est non négligeable. Pour réduire ce temps, deux nouvelles stratégies de seuillage sont proposées dans [11] et utilisées pour obtenir le résultat de la figure 2.

Les composantes obtenues à l'aide de l'analyse en composantes morphologiques peuvent être utilisées pour extraire des caractéristiques particulières : certaines relatives à la géométrie de la texture dynamique (direction du mouvement principal, uniformité du mouvement, *etc.*) et d'autres relatives aux phénomènes locaux (vitesse, vortex locaux, *etc.*).

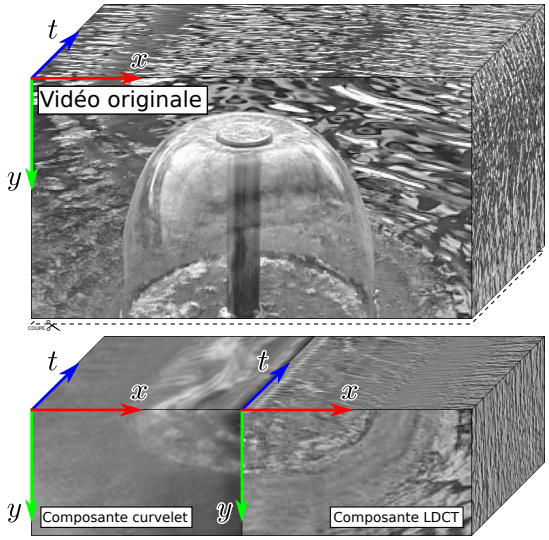

FIG. 2 – *Résultat de la décomposition d'une texture dynamique en utilisant l'analyse en composantes morphologiques. L'onde porteuse est extraite à l'aide de la transformée en curvelet 2D+T tandis que les phénomènes locaux sont obtenus avec la transformée en cosinus locale 2D+T.*

3 Indexation de textures dynamiques

La principale contribution de cet article concerne l'indexation de textures dynamiques. Différentes approches (présentées dans la section 2) et différents vecteurs de descripteurs sont comparés sur des grandes bases de textures dynamiques disponibles en ligne. L'objectif est d'évaluer ces approches et d'identifier les descripteurs les plus pertinents.

Chaque expérimentation se déroule de la manière suivante : (1) analyse de la séquence d'images en utilisant une décomposition spatio-temporelle, (2) calcul du descripteur et construction d'un vecteur caractéristique, (3) obtention d'une matrice de confusion à l'aide de la méthode du *leave-one-out*, (4) calcul d'un taux de classification à partir de cette matrice.

3.1 Construction des vecteurs caractéristiques

Les vecteurs caractéristiques sont construits à partir des différentes décompositions multi-échelles présentées précédemment : (1) la transformée en ondelettes image par image, (2) la transformée en ondelettes temporelle, (3) la transformée en ondelettes 2D+T, (4) l'approche de J.R. Smith *et al.* [23], (5) la transformée en curvelet 2D+T, (6) et les composantes morphologiques issues de la décomposition en utilisant la transformée en curvelet 2D+T et la transformée en cosinus locale 2D+T. Chaque méthode est identifiée par un indice : $m = \{fpf, t, xyt, xy2t, curv, mca\}$ (les indices sont dans l'ordre de la liste précédente).

Une méthode habituelle pour caractériser les textures 2D en utilisant une transformée en ondelettes est de construire un vecteur caractéristique à partir des sous-bandes de détails. De la même manière, $d_m^{(j,\ell)}$ représente la sous-bande de détails de taille $(t_1^{(j,\ell)} \times t_2^{(j,\ell)} \times t_3^{(j,\ell)})$ pour l'analyse multi-résolution m à l'échelle j pour l'orientation ℓ (horizontale, verticale, *etc.*).

Les descripteurs suivants sont calculés pour chaque sous-bandes $d_m^{(j,\ell)}$ de détails :

– Moyennes des sous-bandes de détails :

$$\mu_m^{(j,\ell)} = \alpha \sum_{k_1=1}^{t_1^{(j,\ell)}} \sum_{k_2=1}^{t_2^{(j,\ell)}} \sum_{k_3=1}^{t_3^{(j,\ell)}} d_m^{(j,\ell)}(k_1,k_2,k_3) \tag{2}$$

avec $\alpha = \dfrac{1}{t_1^{(j,\ell)} t_2^{(j,\ell)} t_3^{(j,\ell)}}$.

– Écarts-types des sous-bandes de détails :

$$\sigma_m^{(j,\ell)} = \sqrt{\alpha \sum_{k_1=1}^{t_1^{(j,\ell)}} \sum_{k_2=1}^{t_2^{(j,\ell)}} \sum_{k_3=1}^{t_3^{(j,\ell)}} \left(d_m^{(j,\ell)}(k_1,k_2,k_3) - \mu_m^{(j,\ell)}\right)^2} \tag{3}$$

– Moyennes des énergies des sous-bandes de détails :

$$E_m^{(j,\ell)} = \alpha \sum_{k_1=1}^{t_1^{(j,\ell)}} \sum_{k_2=1}^{t_2^{(j,\ell)}} \sum_{k_3=1}^{t_3^{(j,\ell)}} \left(d_m^{(j,\ell)}(k_1,k_2,k_3)\right)^2 \tag{4}$$

– Entropies des sous-bandes de détails :

$$H_m^{(j,\ell)} = -\sum_{k_1=1}^{t_1^{(j,\ell)}} \sum_{k_2=1}^{t_2^{(j,\ell)}} \sum_{k_3=1}^{t_3^{(j,\ell)}} \rho \log(\rho) \tag{5}$$

avec $\rho = p(d_m^{(j,\ell)}(k_1,k_2,k_3))$ la fonction de densité de probabilité de $d_m^{(j,k)}$ pour une sous-bande de détails ($d_m^{(j,k)}$ est considéré comme une variable aléatoire).

Dans le cas de la transformée en cosinus locale 2D+T utilisée dans l'analyse en composantes morphologiques, il n'existe pas de sous-bandes de détails. Comme illustré dans la figure 3, les coefficients de cette transformée sont divisés en plusieurs sous-bandes. Chaque sous-bande correspond ainsi à un ensemble de fréquences d'échelle similaire. Les descripteurs sont alors calculés de la même façon que lors d'une analyse multi-résolution.

Dans notre objectif d'indexation, nous obtenons cinq signatures différentes, quatre construites directement à partir d'un seul descripteur, et une construite à partir de la concaténation des quatre descripteurs :

– Vecteur de descripteurs s'appuyant sur la moyenne des sous-bandes de détails :

$$\mathbf{S}_m^{\mu} = \left(\mu_m^{(1,1)}, \cdots, \mu_m^{(1,N_\ell^1)}, \cdots, \mu_m^{(N_j,1)}, \cdots, \mu_m^{(N_j,N_\ell^j)}\right) \tag{6}$$

avec N_ℓ^j le nombre d'orientations à l'échelle j et N_j le nombre d'échelles.

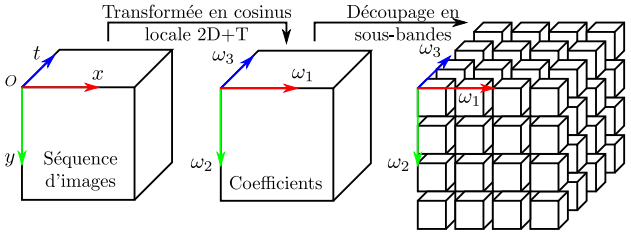

FIG. 3 – *Découpage des coefficients de la transformée en cosinus locale 2D+T pour obtenir différentes sous-bandes de détails.*

– Vecteur de descripteurs fondé sur l'écart-type des sous-bandes de détails :

$$\mathbf{S}_m^{\sigma} = \left(\sigma_m^{(1,1)}, \cdots, \sigma_m^{(1,N_\ell^1)}, \cdots, \sigma_m^{(N_j,1)}, \cdots, \sigma_m^{(N_j,N_\ell^j)}\right) \tag{7}$$

– Vecteur de descripteurs s'appuyant sur l'énergie des sous-bandes de détails :

$$\mathbf{S}_m^{E} = \left(E_m^{(1,1)}, \cdots, E_m^{(1,N_\ell^1)}, \cdots, E_m^{(N_j,1)}, \cdots, E_m^{(N_j,N_\ell^j)}\right) \tag{8}$$

– Vecteur de descripteurs calculé à partir de l'entropie des sous-bandes de détails :

$$\mathbf{S}_m^{H} = \left(H_m^{(1,1)}, \cdots, H_m^{(1,N_\ell^1)}, \cdots, H_m^{(N_j,1)}, \cdots, H_m^{(N_j,N_\ell^j)}\right) \tag{9}$$

– Vecteur de descripteurs s'appuyant sur différentes caractéristiques des sous-bandes de détails :

$$\mathbf{S}_m^{A} = \left(\mathbf{S}_m^{\mu}, \mathbf{S}_m^{\sigma}, \mathbf{S}_m^{E}, \mathbf{S}_m^{H}\right) \tag{10}$$

Ainsi pour une base de données vidéos, nous obtenons un ensemble de signatures $\mathbf{S}_{m,c,i}^d$ avec m représentant la méthode spatio-temporelle utilisée, $d = \{\mu, \sigma, E, H, A\}$ étant le(s) descripteur(s) utilisé(s) et i le i-ème échantillon de la classe c de la base.

Dans notre cas, nous normalisons les vecteurs caractéristiques de la manière suivante :

$$\forall n, \forall r, \forall j, \;\; \mathbf{S}_{m,r,j}^d(n) = \frac{\mathbf{S}_{m,r,j}^d(n) - \min\limits_{c,i} \mathbf{S}_{m,c,i}^d(n)}{\max\limits_{c,i} \mathbf{S}_{m,c,i}^d(n) - \min\limits_{c,i} \mathbf{S}_{m,c,i}^d(n)} \tag{11}$$

avec n représentant l'indice du paramètre dans le vecteur caractéristique.

Les expérimentations ont également été menées avec d'autres paramètres (variation du nombre d'échelles de décomposition, différentes normalisations, *etc.*). Les meilleurs résultats sont présentés dans la section 3.3. Les expérimentations sont effectuées en prenant en compte différents ensembles de paramètres.

3.2 Bases de données utilisées

Chaque décomposition multi-échelle a été testée pour tous les descripteurs de la section 3.1 sur trois bases de textures dynamiques de complexité croissante.

Les trois bases de données utilisées pour nos expérimentations sont constituées des vidéos de DynTex [21]. Un exemple de chaque classe de chaque base est visible sur la figure 4.

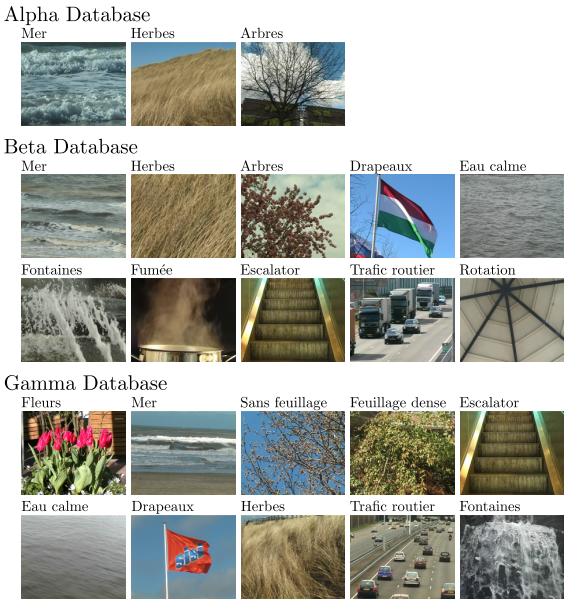

FIG. 4 – *Exemple de chaque classe de chaque base de textures dynamiques.*

Ces bases diffèrent par leur difficulté, le nombre de classes et le nombre d'éléments :

– **Base de données Alpha** : 60 séquences d'images de textures dynamiques regroupées en 3 classes relativement disctinctes : Mer, Herbes et Arbres.
– **Base de données Beta** : 162 séquences d'images de textures dynamiques regroupées en 10 classes : Mer, Herbes, Arbres, Drapeaux, Eau calme, Fontaines, Fumée, Escalator, Trafic routier, Rotation. Ici, des phénomènes plus complexes sont présents par rapport à la base de donnée Alpha.
– **Base de données Gamma** : 275 séquences d'images de textures dynamiques regroupées en 10 classes : Fleurs, Mer, Arbres sans feuillage, Feuillage dense, Escalator, Eau calme, Drapeaux, Herbes, Trafic routier, Fontaines. Dans cette base de données, les classes sont composées de nombreux échantillons couvrant de nombreux cas (changement d'échelles, d'orientations). Il s'agit d'un jeu de données très complexe.

3.3 Résultats

Dans cette section, les résultats de l'indexation de textures dynamiques sont présentés. Toutes les décompositions multi-échelles sont calculées sur 5 niveaux de réso-

lution. Seule l'approche de J.R. Smith *et al.* est calculée sur 3 niveaux de résolution de par sa construction.

Le tableau 1 montre les taux de reconnaissance obtenus avec les vecteurs caractéristiques précédents (*cf.* 3.1).

	Base	$\mathbf{S}_m^{\mu}$	$\mathbf{S}_m^{\sigma}$	$\mathbf{S}_m^{E}$	$\mathbf{S}_m^{H}$	$\mathbf{S}_m^{A}$
(1)	Alpha	68	82	78	88 †⋆	88 †⋆
	Beta	41 †	50	51	57	66
	Gamma	38	60	60	55	65
(2)	Alpha	37	75	67	67	73
	Beta	15	43	37	28	46
	Gamma	15	34	31	26	40
(3)	Alpha	72 †	85 †	85 †	87	85
	Beta	33	62	61	65	65
	Gamma	36	65 †	64 †	61 †	68 †⋆
(4)	Alpha	65	83	80	82	83
	Beta	35	65	59	65	67
	Gamma	43 †	63	56	59	65
(5)	Alpha	47	85 †	83	85	85
	Beta	19	65	61	62	67
	Gamma	18	62	60	56	63
(6)	Alpha	37	83	83	83	85
	Beta	23	68 †	64 †	66 †	70 †⋆
	Gamma	19	61	62	59	63

TAB. 1 – Taux de reconnaissance (en %) des textures dynamiques selon les différents descripteurs et sur trois bases de données. † représente les meilleurs taux de reconnaissance pour un descripteur sur une base de données. ⋆ représente le meilleur taux de reconnaissance pour une base de données. (1) : transformée en ondelettes image par image. (2) : transformée en ondelettes temporelle. (3) : transformée en ondelettes 2D+T. (4) : approche de J.R. Smith *et al.* [23]. (5) : transformée en curvelet 2D+T. (6) : composantes morphologiques issus de la décomposition en utilisant la transformée en curvelet 2D+T et la transformée en cosinus locale 2D+T.

Pour chaque base de données, les taux de reconnaissance permettent d'observer :

- Alpha : la méthode la plus discriminante est celle de la transformée en ondelettes image par image. Au regard des éléments constituant la base de données, nous constatons clairement que la distinction entre les classes peut s'effectuer uniquement à partir des propriétés spatiales de chaque séquence. Dans ce cas, la dimension temporelle n'ajoute que peu d'information intéressante.

  Les autres méthodes d'analyse (exceptée la transformée en ondelettes temporelle) restent performantes puisqu'elles obtiennent un taux de reconnaissance proche (5% d'écart).
- Beta : pour cette base, l'approche la plus discriminante est celle fondée sur l'analyse en composantes morphologiques. Celle-ci obtient le meilleur taux de reconnaissance pour 4 des vecteurs caractéristiques proposés (S_m^{σ}, S_m^{E}, S_m^{H} et S_m^{A}).
- Gamma : la transformée en ondelettes 2D+T est la méthode obtenant les taux de reconnaissance les plus élevés pour cette base. Comme pour les autres bases de données, les différentes méthodes se démarquent peu et les taux de reconnaissance restent relativement proches.

Les mêmes tests effectués sur les bases de données Beta et Gamma conduisent à un taux de reconnaissance de 15% moins élevé que la base Alpha ; ceci est du à la complexité des classes ainsi qu'au nombre d'échantillons.

Les différentes approches multi-échelles, exceptée la transformée en ondelettes temporelle, apportent des taux de reconnaissance acceptables sur l'ensemble des bases de données. Les résultats obtenus sont toujours dans un intervalle de valeurs relativement proches, alors qu'ils sont bien inférieurs lorsque seule l'information temporelle est considérée.

3.4 Discussion

Les taux de reconnaissance obtenus sont donc satisfaisants : ils sont de l'ordre de 70% pour des bases de données de dimension relativement importante (bases Beta et Gamma) et ce en utilisant seulement l'information fréquetielle. Ces performances peuvent sûrement être améliorées. Nous proposons ici plusieurs pistes de réflexion.

Lorsque l'on regarde les dimensions des vecteurs caractéristiques, fonction de la méthode multi-échelle employée (cf. tableau 2), on observe que pour certaines méthodes, le nombre de descripteurs est plus grand que le nombre d'échantillons à classer. Ceci pose plusieurs difficultés :

- l'information permettant de classer les échantillons est trop redondante et peut détériorer la classification.
- les remarques apportées sur les approches multi-échelles peuvent être discutées puisqu'elles ne s'établissent pas dans les mêmes conditions. En effet, la classification de 10 classes dans un espace à 3 dimensions n'est pas de la même difficulté que dans un espace à 5508 dimensions, et l'approche de classification utilisée n'est pas nécessairement adaptée.

Il convient donc de réduire la dimension des signatures. Plusieurs approches sont envisageables :

- réduire la dimension des signatures à l'aide de l'analyse en composantes principales. Deux tests ont été réalisés en ramenant la dimension des signatures à 50 et 15. Toutefois, les taux de reconnaissance ont grandement chuté (de l'ordre de 30% à 40%).
- effectuer une sélection de variables à l'aide, par exemple, de la méthode *Stepwise Discriminant Analysis* [13]. Géométriquement, il s'agit de trouver le sous-espace de représentation qui permet un écartement maximal entre les centres de gravité des nuages de points.
- modifier la façon de construire les vecteurs descripteurs afin de les rendre plus compacts et plus représentatifs de l'information spatio-temporelle. Plutôt que de calculer l'énergie de toutes les sous-bandes de détails des approches multi-échelles, nous pourrions uniquement calculer celle de chaque échelle et utiliser d'autres descrip-

	Nombre d'échelles	Nombre de subdivisions	Taille des vecteurs	
			$S_m^{\mu,\sigma,E,H}$	S_m^A
(1)	3	×	9	36
	4	×	12	48
	5	×	15	60
(2)	3	×	3	12
	4	×	4	16
	5	×	5	20
(3)	3	×	21	84
	4	×	28	112
	5	×	35	140
(4)	3	×	45	180
(5)	3	2	25	100
	3	4	97	388
	4	2	121	484
	4	4	481	1924
	5	2	217	868
	5	4	865	3460
(6)	×	×	1377	5508

TAB. 2 – Dimension des vecteurs caractéristiques en fonction des méthodes d'analyses. (1) : transformée en ondelettes image par image. (2) : transformée en ondelettes temporelle. (3) : transformée en ondelettes 2D+T. (4) : approche de J.R. Smith *et al.* [23]. (5) : transformée en curvelet 2D+T. (6) : composantes morphologiques issues de la décomposition en utilisant la transformée en curvelet 2D+T et la transformée en cosinus locale 2D+T.

teurs pour qualifier l'information directionelle (comme l'homogénéité de direction [19]).

Par exemple, dans le cadre de la transformée en curvelet 2D+T, au lieu d'un vecteur de 3460 éléments, on obtiendrait un ensemble de 5 éléments représentatifs des différentes échelles additionné des éléments caractéristiques de l'information directionelle.

Dans l'objectif d'améliorer les performances de l'indexation, nous pourrions également regarder plus en détails l'invariance de nos descripteurs. En effet, à l'heure actuelle, nos vecteurs caractéristiques sont par construction invariants en translation, mais non en rotation et en échelle. C'est un problème classiquement rencontré lors de l'utilisation de l'énergie de sous-bandes d'ondelettes dans un vecteur caractéristique.

Pour résoudre ce problème d'invariance, nous envisageons deux pistes :

– la création d'un ordre de parcours des différentes orientations pour chaque approche multi-échelle, permettant à l'aide d'une permutation circulaire, de rendre invariant en rotation la signature. Toutefois, cette méthode ne permet pas la réduction du nombre de dimensions.
– la construction de nouveaux descripteurs comme par exemple l'homogénéité des directions [19], ou l'amplitude maximale intégrant directement la notion d'invariance.

Cette deuxième approche est la piste que nous privilégions pour nos travaux futurs.

Concernant l'analyse en composantes morphologiques, les descripteurs construits sont les mêmes pour les deux composantes. Il pourrait être pertinent de construire des descripteurs plus adaptés à la nature des composantes extraites.

4 Conclusion et perspectives

Cet article présente trois approches pour la caractérisation de textures dynamiques. Une première s'appuyant sur des transformées en ondelettes utilisant le produit tensoriel, une seconde utilisant une transformée multi-échelle géométrique et enfin une dernière se fondant sur la décomposition d'une texture dynamique en deux composantes (ondes porteuses et phénomènes locaux). L'objectif principal était l'étude de l'influence des décompositions spatio-temporelles sur la caractérisation de textures dynamiques.

Après avoir présenté les différentes approches multi-échelles pour la caractérisation de textures dynamiques, nous proposons des vecteurs de descripteurs utilisant les sous-bandes de détails. Ces signatures sont testées sur trois grandes bases de textures dynamiques disponibles en ligne. Enfin, les résultats de l'indexation de textures dynamiques sont présentés et discutés.

L'utilisation des approches multi-échelles pour l'analyse des textures dynamiques est réellement prometteuse car liées aux propriétés physiques présentes dans ces phénomènes complexes.

Les descripteurs utilisés dans cet article peuvent être améliorés ; par exemple, les différentes composantes obtenus en utilisant l'analyse en composantes morphologiques peuvent servir à extraire des caractéristiques propres ; certaines relatives à la géométrie (direction du mouvement principal, uniformité du mouvement global, *etc.*) et certaines relatives aux phénomènes locaux (vitesse, vortex locaux, *etc.*).

Dans une étude futur, l'ensemble des descripteurs obtenue sera comparé aux autres approches de caractérisation de textures dynamiques faisant figurent de référence dans la littérature (par exemple les approches basées sur les caractéristiques du flot optique).

Avec des descripteurs pertinents de textures dynamiques, de nombreuses applications sont envisageables : suivi de textures dynamiques (évolution d'un feu) ou synthèse (rendu réaliste d'une texture dynamique dans un jeu ou un film), *etc.*.

Références

[1] J.F. Aujol and A. Chambolle. Dual Norms and Image Decomposition Models. *Computer Vision*, 63 :85–

104, 2005.

[2] E. Candès. *Ridgelets : Theory and Applications*. PhD thesis, University of Standford, 1998.

[3] E. Candès and L. Demanet. The curvelet Representation of Wave Propagators is Optimally Sparse. *Communications on Pure and Applied Mathematics*, 58 :1472–1528, 2005.

[4] E. Candès, L. Demanet, D.L. Donoho, and L. Ying. Fast Discrete Curvelet Transforms. Technical report, California Institute of Technology, 2005.

[5] E. Candès and D.L. Donoho. Continuous Curvelet Transform : I. Resolution of the Wavefront Set. *Applied and Computational Harmonic Analysis*, 19 :162 – 197, 2005.

[6] D. Chetverikov and R. Péteri. A Brief Survey of Dynamic Texture Description And Recognition. In *Int. Conf. Computer Recognition Systems*, pages 17–26. Springer, 2005.

[7] D.L. Donoho and M.R. Duncan. Digital Curvelet Transform : Strategy, Implementation and Experiments. In *Wavelet Applications VII*, pages 12–29. SPIE, 1999.

[8] S. Dubois, R. Péteri, and M. Ménard. A 3D Discrete Curvelet based Method for Segmenting Dynamic Textures. In *Int. Conf. on Image Processing*, pages 1373–1376, 2009.

[9] S. Dubois, R. Péteri, and M. Ménard. A Comparison of Wavelet Based Spatio-temporal Decomposition Methods for Dynamic Texture Recognition. In *Ib. Conf. on Pattern Recognition and Image Analysis*, pages 314–321, 2009.

[10] S. Dubois, R. Péteri, and M. Ménard. Analyse de Textures Dynamiques par décompositions spatio-temporelles : application à l'estimation du mouvement global. In *CORESA*, 2010.

[11] S. Dubois, R. Péteri, and M. Ménard. Decomposition of Dynamic Textures using Morphological Component Analysis. *IEEE Trans. on Circuits and Systems for Video Technology*, 2011.

[12] S. Dubois, R. Péteri, and M. Ménard. Textures Dynamiques : état de l'art, modélisation, applications. In *23ème colloque du GRETSI*, 2011.

[13] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley-Interscience, 2001.

[14] R. Fablet and P. Bouthemy. Motion Recognition Using Spatio-temporal Random Walks in Sequence of 2D Motion-related Measurements. In *IEEE Int. Conf. on Image Processing*, pages 652–655, 2001.

[15] S. Mallat. A Theory for Multiresolution Signal Decomposition : The Wavelet Representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence journal*, 11 :674–693, 1989.

[16] S. Mallat and G. Peyré. A Review of Bandlet Methods for Geometrical Image Representation. *Numerical Algorithms*, 44(3) :205–234, 2007.

[17] R.C. Nelson and R. Polana. Qualitative Recognition of Motion using Temporal Texture. *Computer Vision and Image Understanding*, 56 :78–89, 1992.

[18] K. Otsuka, T. Horikoshi, S. Suzuki, and M. Fujii. Feature Extraction of Temporal Texture Based on Spatio-temporal Motion Trajectory. In *Int. Conf. on Pattern Recognition*, page 1047, 1998.

[19] R. Péteri and D. Chetverikov. Qualitative Characterization of Dynamic Textures for Video Retrieval. In *Int. Conf. on Computer Vision and Graphics*, pages 33–38, 2004.

[20] R. Péteri and D. Chetverikov. Dynamic Texture Recognition Using Normal Flow and Texture Regularity. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 223–230, 2005.

[21] R. Péteri, S. Fazekas, and M.J. Huiskes. DynTex : A Comprehensive Database of Dynamic Textures. *Pattern Recognition Letters*, 31 :1627–1632, 2010.

[22] P. Saisan, G. Doretto, Y.N. Wu, and S. Soatto. Dynamic Texture Recognition. In *Conf. on Computer Vision and Pattern Recognition*, pages 58–63, 2001.

[23] J.R. Smith, C.Y. Lin, and M. Naphade. Video Texture Indexing using Spatio-Temporal Wavelets. In *IEEE Int. Conf. on Image Processing*, pages 437–440, 2002.

[24] J.L. Starck, M. Elad, and D.L. Donoho. Redundant Multiscale Transforms and their Application for Morphological Component Analysis. *Advances in Imaging and Electron Physics*, 132, 2004.

[25] R.P. Wildes and J.R. Bergen. Qualitative Spatiotemporal Analysis Using an Oriented Energy Representation. In *Europ. Conf. on Computer Vision-Part II*, pages 768–784, 2000.

[26] A. Woiselle, J.L. Starck, and M.J. Fadili. 3d data denoising and inpainting with the fast curvelet transform. *Journal of Mathematical Imaging and Vision*, 39(2) :121–139, 2011.

[27] P. Wu, Y.M. Ro, C.S. Won, and Y. Choi. Texture Descriptors in MPEG-7. In *Int. Conf. on Computer Analysis of Images and Patterns*, pages 21–28, 2001.

[28] L. Ying, L. Demanet, and E. Candès. 3D Discrete Curvelet Transform. In *International Society for Optical Engineering*, 2005.

[29] G. Zhao and M. Pietikäinen. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29 :915–928, 2007.

[30] J. Zhong and S. Scarlaroff. Temporal Texture Recognition Model Using 3D Features. Technical report, 2002.