A Comparison of Wavelet Based Spatio-temporal Decomposition Methods for Dynamic Texture Recognition

Sloven Dubois, Renaud Péteri, Michel Menard

A Comparison of Wavelet Based Spatio-temporal Decomposition Methods for Dynamic Texture Recognition

Sloven Dubois[1,2], Renaud Péteri[2], and Michel Ménard[1]

[1] L3i - Laboratoire Informatique Image et Interaction

[2] MIA - Mathématiques, Image et Applications

Avenue Michel Crépeau
17042 La Rochelle, France
{sloven.dubois01, renaud.peteri, michel.menard} @univ-lr.fr

Abstract. This paper presents four spatio-temporal wavelet decompositions for characterizing dynamic textures. The main goal of this work is to compare the influence of spatial and temporal variables in the wavelet decomposition scheme. Its novelty is to establish a comparison between the only existing method [11] and three other spatio-temporal decompositions.
The four decomposition schemes are presented and successfully applied on a large dynamic texture database. Construction of feature descriptors are tackled as well their relevance, and performances of the methods are discussed. Finally, future prospects are exposed.

Key words: Spatio-temporal wavelets, dynamic textures, feature descriptors, video indexing

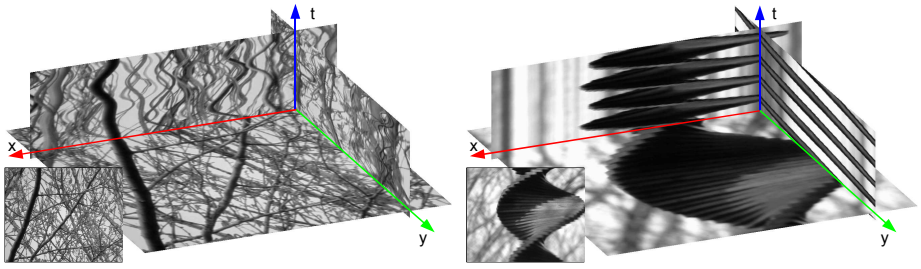
1 Introduction

The last fifteen years have seen the rising of a new issue in texture analysis which its extension to the spatio-temporal domain, called dynamic textures. Dynamic textures are spatially and temporally repetitive patterns like trees waving in the wind, water flows, fire, smoke phenomena, rotational motions ...

The study of dynamic textures has several fields of applications. One of the main topics is the synthesis of dynamic textures [6, 12]. Other fields of research are dynamic texture detection [4] or dynamic texture segmentation [2, 5]. Our research context is the recognition and description of dynamic textures [8, 14]. For a brief survey on this topic, one could refer to [3].

A dynamic texture is composed of motions occurring at several spatio-temporal scales: on figure 1 (a), one can observe the low spatio-temporal motion of the trunk and the high spatio-temporal motion of small branches. An efficient method should then be able to capture this spatio-temporal behavior.

A natural tool for multiscale analysis is the wavelet transform. In the field of image processing, the wavelet transform has been successfully used for characterizing static textures. For instance, gabor wavelets have been used for computing the texture descriptor of the MPEG-7 norm [13].



(a) (b)

Fig. 1. 2D+t slices of two dynamic textures. Here, a dynamic texture is seen as a data cube and is cut at pixel $O(x, y, t)$ for obtaining three planes $(\overrightarrow{x}O\overrightarrow{y})$, $(\overrightarrow{x}O\overrightarrow{t})$ and $(\overrightarrow{y}O\overrightarrow{t})$.

A natural idea is hence to extend the wavelet decomposition to the temporal domain. To our knowledge, the work of [11] has been so far the only spatio-temporal wavelet decomposition method for characterizing dynamic textures. Yet, the authors use a small offline database, which can be a limitation when it comes to study feature relevance and to compare different methods.

In this paper, we present a comparison between this method [11] and three other spatio-temporal wavelet decomposition schemes. Our goal is to study the influence of spatial or temporal variables in the decomposition scheme. For relevant testing purposes, we also aim at computing results on a large online dynamic texture dataset [9].

This article is organized as follows : section 2 recalls how to compute the discrete wavelet transform using filter banks. Section 3 presents four spatio-temporal wavelet decomposition methods which differ from the way space and time variables are treated. In section 4 the construction of texture feature descriptors for each decomposition scheme is presented. Finally, the relevance of each decomposition scheme is tested on a large dataset of dynamic textures [9] and numerical results are presented and discussed.

2 Discrete Wavelet Transform and filter banks

For a better understanding of the different decomposition schemes, the discrete wavelet transform (DWT) and its associated filter bank is recalled. For any level j and for any n, we denote:

$$a^j[n] = \langle f, \phi_n^{2^j} \rangle \quad \text{and} \quad d^j[n] = \langle f, \psi_n^{2^j} \rangle \tag{1}$$

with ϕ the scaling function and ψ is the mother wavelet. $a^j[n]$ is the approximation of the signal f and $d^j[n]$ its details at resolution 2^j.

A fast discrete wavelet transform implementation is performed using filter bank [7]. Relations between the decomposition scales are:

$$a^{j+1}[n] = a^j[2n] \otimes \overline{h} \quad \text{and} \quad d^{j+1}[n] = a^j[2n] \otimes \overline{g} \tag{2}$$

where $\otimes$ denotes the convolution product, h and g are respectively one-dimensional low-pass and high-pass decomposition filters associated to the scaling function and the mother wavelet, and $\overline{g}[m] = g[-m]$. This filter bank is represented on figure 2. Thereafter in the article, the symbol $\boxed{\text{WT}}$ represents the filter bank.

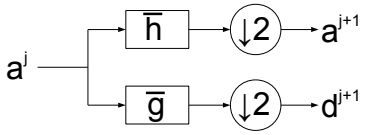


Fig. 2. Scheme of discrete wavelet transform for one level of decomposition where $\boxed{F}$ is the convolution by F and $\downarrow 2$ is the decimation by 2

For digital images, the fast wavelet transform is extended to the 2D case using one scaling function and three mother wavelets (see [7]).

At each scale one approximation subband and three detail subbands are obtained using the following filter relations:

$$\begin{array}{ll} a^{j+1}[n,m] = a^j[2n,2m] \otimes \overline{h_x}\,\overline{h_y} & \text{and} \quad d_v^{j+1}[n,m] = a^j[2n,2m] \otimes \overline{h_x}\,\overline{g_y} \\ d_h^{j+1}[n,m] = a^j[2n,2m] \otimes \overline{g_x}\,\overline{h_y} & \qquad\quad d_d^{j+1}[n,m] = a^j[2n,2m] \otimes \overline{g_x}\,\overline{g_y} \end{array} \quad (3)$$

with the notation $hg[n_1,n_2] = h[n_1]g[n_2]$ (hence the associated 2D filters are the matrix product of 1D filters).

3 Video Analysis using wavelet decomposition

In this section, four spatio-temporal decomposition schemes using wavelets are presented. They vary in the way they consider space and time variables in the multiresolution analysis.

3.1 Spatial Wavelet Decomposition

A straightforward approach is to use the wavelet transform image per image. In this case, there is no consideration on the temporal correlation between two successive frames. The first method is summarized figure 3. For each image and for each scale of multiresolution analysis, the approximation subband and three details subbands are computed. The feature descriptors use wavelet subband energies and will be detailed in section 4.

3.2 Temporal Wavelet Decomposition

The first method considers a video image per image, and is thus a purely spatial method. The second natural approach is to perform the multiresolution analysis in the time direction.

Figure 4 shows how the feature descriptors is extracted. For each pixel (x,y)

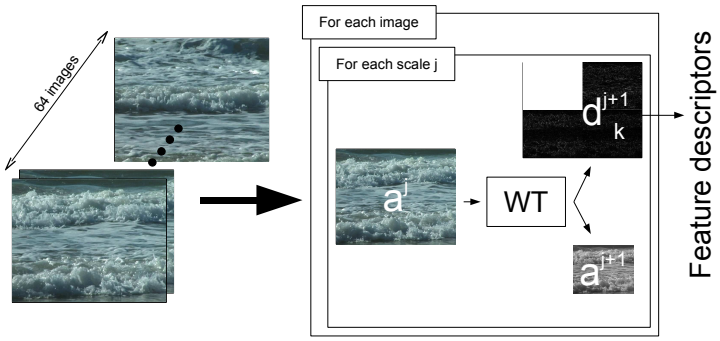


Fig. 3. Spatial wavelet decomposition applied to a sequence of 64 images for obtaining feature descriptors. WT is the filter bank.

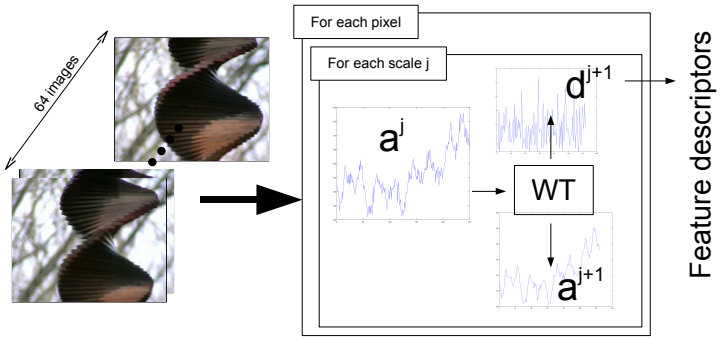


Fig. 4. Temporal wavelet decomposition applied to sequence of 64 images for obtaining feature descriptors. WT is the filter bank.

of a dynamic texture video, the temporal profile is extracted and its one dimensional wavelet transform is performed.

The associated filter bank can be written in the same way than in section 2, filters $\overline{h_t}$ and $\overline{g_t}$ being applied in the time direction.

3.3 2D+T Wavelet Decomposition using filter product

Whereas the first method is a purely spatial decomposition and the second one is a temporal decomposition, the third method performs decomposition spatially and temporally.

This extension to the temporal domain of the 2D discrete wavelet transform is done using separable filter banks. As in the 2D case, a separable 3 dimensional convolution can be factored into one-dimensional convolution along rows, columns and image indexes of the video. A cascade of three filters is obtained for each decomposition level and different subbands can be obtained with the

following relations:

$$\begin{array}{ll} a^{j+1}[n,m,p] = a^j[2n,2m,2p] \otimes \overline{h_x}\ \overline{h_y}\ \overline{h_t} & d_t^{j+1}[n,m,p] = a^j[2n,2m,2p] \otimes \overline{h_x}\ \overline{h_y}\ \overline{g_t} \\ d_h^{j+1}[n,m,p] = a^j[2n,2m,2p] \otimes \overline{g_x}\ \overline{h_y}\ \overline{h_t} & d_{ht}^{j+1}[n,m,p] = a^j[2n,2m,2p] \otimes \overline{g_x}\ \overline{h_y}\ \overline{g_t} \\ d_v^{j+1}[n,m,p] = a^j[2n,2m,2p] \otimes \overline{h_x}\ \overline{g_y}\ \overline{h_t} & d_{vt}^{j+1}[n,m,p] = a^j[2n,2m,2p] \otimes \overline{h_x}\ \overline{g_y}\ \overline{g_t} \\ d_d^{j+1}[n,m,p] = a^j[2n,2m,2p] \otimes \overline{g_x}\ \overline{g_y}\ \overline{h_t} & d_{dt}^{j+1}[n,m,p] = a^j[2n,2m,2p] \otimes \overline{g_x}\ \overline{g_y}\ \overline{g_t} \end{array} \quad (4)$$

Figure 5 shows the third method. For a video of 64 images, seven detail subcubes and one approximation subcube are computed for each scale.

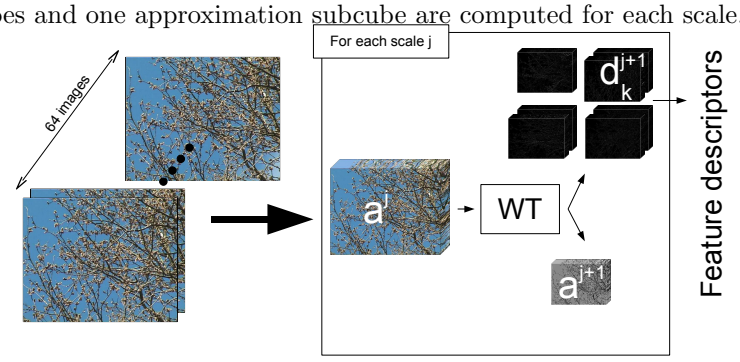


Fig. 5. Simple 3D wavelet decomposition applied to a sequence of 64 images for obtaining feature descriptors. WT is the filter bank.

3.4 J. R. Smith *et al.* Wavelet Decomposition

The last decomposition method is the one of Smith *et al.* [11]. This transform is similar to the 2D+T wavelet decomposition, except that the temporal filter is applied two times at each multiresolution step so that the video is decimated twice spatially and twice temporally. For a video of 64 images, they then obtained fifteen detail subcubes and one approximation subcube (see [11] for more details).

For a given resolution 2^j, the filter bank can be written as follows :

$$\begin{array}{ll} a^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{h_x}\ \overline{h_y}\ \overline{h_t}\ \overline{h_t} & d_t^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{h_x}\ \overline{h_y}\ \overline{h_t}\ \overline{g_t} \\ d_h^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{g_x}\ \overline{h_y}\ \overline{h_t}\ \overline{h_t} & d_{ht}^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{g_x}\ \overline{h_y}\ \overline{h_t}\ \overline{g_t} \\ d_v^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{h_x}\ \overline{g_y}\ \overline{h_t}\ \overline{h_t} & d_{vt}^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{h_x}\ \overline{g_y}\ \overline{h_t}\ \overline{g_t} \\ d_d^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{g_x}\ \overline{g_y}\ \overline{h_t}\ \overline{h_t} & d_{dt}^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{g_x}\ \overline{g_y}\ \overline{h_t}\ \overline{g_t} \\ d_t^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{h_x}\ \overline{h_y}\ \overline{g_t}\ \overline{h_t} & d_{tt}^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{h_x}\ \overline{h_y}\ \overline{g_t}\ \overline{g_t} \\ d_{ht}^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{g_x}\ \overline{h_y}\ \overline{g_t}\ \overline{h_t} & d_{htt}^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{g_x}\ \overline{h_y}\ \overline{g_t}\ \overline{g_t} \\ d_{vt}^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{h_x}\ \overline{g_y}\ \overline{g_t}\ \overline{h_t} & d_{vtt}^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{h_x}\ \overline{g_y}\ \overline{g_t}\ \overline{g_t} \\ d_{dt}^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{g_x}\ \overline{g_y}\ \overline{g_t}\ \overline{h_t} & d_{dtt}^{j+1}[n,m,p] = a^j[2n,2m,4p] \otimes \overline{g_x}\ \overline{g_y}\ \overline{g_t}\ \overline{g_t} \end{array} \quad (5)$$

For all four presented methods, feature vectors based on wavelet energies are computed and tested in the next sections.

4 Computing dynamic texture descriptors

In order to characterize a given dynamic texture, subband energies are used. For each decomposition, a family of wavelet coefficients $\{d_k^j\}$ is obtained, k being

the direction that depends on the wavelet decomposition scheme (h, v, d, and t). A classical way for establishing feature descriptors is to sum the energy of each subband as follows [11]:

$$E_k^j = \frac{1}{XYT} \sum_x^X \sum_y^Y \sum_t^T |d_k^j|^2 \tag{6}$$

where j denotes the decomposition level, k the direction, X, Y, T are the spatio-temporal dimensions.

The size of feature descriptors depends on the decomposition method. Table 1 gives this size with respect to the method. This table shows that three methods are computed with three decomposition levels and one with five levels.

Methods	Spatial Wavelet Decomposition	Temporal Wavelet Decomposition	2D+T Wavelet Decomposition	J. R. Smith *et al.* Wavelet Decomposition
Size	9	5	21	45

Table 1. Size of feature descriptors with respect to the wavelet decomposition method

5 Results

The feature vectors of the four methods are tested using 83 videos from DynTex [9], a database that provides high-quality videos of dynamic textures. These videos are divided into five classes, representative of different spatio-temporal dynamics. An example for each class can be seen on figure 6.

Dynamic textures of class (a) are composed of low and high frequency motions (for instance a wave motion and sea foam). Class (b) is composed of oscillating motions with different velocities. For example a tree waving in the wind is composed of high oscillations for small branches or leaves and low oscillations for big branches or the trunk. Class (c) is a high frequency motion only, like fountains or wavy water. The class (d) represents a fluid motion, for instance smoke. This is a difficult class of dynamic textures because they can be sometimes partially transparent. The last class (e) is composed of simple translation events like an escalator motion.

(a) (b) (c) (d) (e)

Fig. 6. Sample for each class.

In order to be in the same testing conditions than [11], the low pass filters and the high pass filters used for all wavelet decomposition methods are the Haar filters.

For studying features pertinence, the "leave-one-out" test is used (see [8] for more details). Feature descriptors are computed on 83 videos and results are summarized in table 2.

Methods	Spatial Wavelet Decomposition	Temporal Wavelet Decomposition	2D+t Wavelet Decomposition	J. R. Smith *et al.* Wavelet Decomposition
(a) 20	**20**	16	17	16
(b) 20	15	11	**17**	13
(c) 20	**17**	13	13	8
(d) 16	**16**	**16**	12	13
(e) 7	5	4	5	5
Success rate	88%	73%	77%	67%

Table 2. Table of results. The first column, presents the class label and its number of videos. The intersection class/method represents the number of well classified videos for each class.

Results show that the first method "Spatial Wavelet Decomposition" gives the best classification rate. The main reason is that the four spatio-temporal wavelet decompositions are simple 3D extension of 2D wavelet transform: these decompositions do not fully take benefit of temporal dynamics. These transforms are indeed not adapted to the local motion and create class ambiguities. However for class (b), the 2D+t methods performs better than other methods. This is due to the fact dynamic textures of this class are really 3D textures occurring in all 3 directions of the wavelet transform.

The method "Temporal Wavelet Decomposition" obtains a good classification rate for class (d). Indeed, for most of class (d) videos, the main information is for the time dimension rather than the spatial one.

An observation of all results shows that the number of well classified videos is over 50% except for class (c) with the method of [11]. It could be explained by the fact that their method applies filtering twice in the temporal direction. As this class (c) is composed of high-frequency phenomena, too much temporal information is filtered at each step.

6 Conclusion

This paper presents four methods for characterizing dynamic textures. Our goal is to study the influence of space and time variables in a spatio-temporal wavelet decomposition scheme and to test the extracted features on a significant video database of dynamic texture.

The three wavelet decomposition methods we have proposed already perform better than the only existing method published so far [11]. On a 83 videos database, success rate ranges between 73% and 88% depending on the proposed decomposition method.

When tested on approximately 300 videos where class ambiguity highly increases, the success rate declines of about 20% for each method. It emphasizes

the need for the development of more adaptive spatio-temporal wavelet transforms that are able to take into account the motion dynamics. Many dynamic textures are indeed non-stationary and have a privileged motion orientation. Like geometrical wavelets ([1, 10]) which are more adapted to the geometrical content of an image than the classical wavelet transform, we are currently working on developing motion tuned wavelet transforms.

References

1. E. Candes, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5:861–899, March 2006.
2. A.B. Chan and N. Vasconcelos. Mixtures of dynamic textures. In *Proceedings of Tenth IEEE International Conference on Computer Vision (ICCV'05)*, volume 1, pages 641–647, 2005.
3. D. Chetverikov and R. Péteri. A brief survey of dynamic texture description and recognition. In *Proceedings of 4th International Conference on Computer Recognition Systems (CORES'05)*, "Advances in Soft Computing", pages pp. 17–26, Rydzyna, Poland, 2005. Springer-Verlag.
4. Y. Dedeoglu, B. U. Toreyin, U. Gudukbay, and A. E. Cetin. Real-time fire and flame detection in video. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*, volume II, pages 669–673, Philadelphia, PA, March 2005.
5. G. Doretto, D. Cremers, P. Favaro, and S. Soatto. Dynamic texture segmentation. In *Proceedings of Ninth IEEE International Conference on Computer Vision (ICCV'03)*, volume 2, pages 1236–1242, 2003.
6. J. Filip, M. Haindl, and D. Chetverikov. Fast synthesis of dynamic colour textures. In *Proceedings of the 18th IAPR Int. Conf. on Pattern Recognition (ICPR'06)*, pages 25–28, Hong Kong, 2006.
7. S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence journal (TPAMI)*, 11(7):674–693, July 1989.
8. R. Péteri and D. Chetverikov. Dynamic texture recognition using normal flow and texture regularity. In *Proceedings of 2nd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'05)*, volume 3523 of *Lecture Notes in Computer Science*, pages pp. 223–230, Estoril, Portugal, 2005. Springer.
9. R. Péteri, M. Huiskes, and S. Fazekas. Dyntex : A comprehensive database of dynamic textures. http://www.cwi.nl/projects/dyntex/.
10. G. Peyré. *Géométrie multi-echelles pour les images et les textures*. PhD thesis, Ecole Polytechnique, dec 2005, 148 pages.
11. J. R. Smith, C. Y. Lin, and M. Naphade. Video texture indexing using spatio-temporal wavelets. In *Proceedings of IEEE International Conference on Image Processing (ICIP'02)*, volume II, pages 437–440, 2002.
12. M. Szummer and R. W. Picard. Temporal texture modeling. In *Proceedings of IEEE International Conference on Image Processing (ICIP'96)*, volume 3, pages 823–826, 1996.
13. P Wu, Y. M. Ro, C. S. Won, and Y. Choi. Texture descriptors in MPEG-7. *Computer Analysis of Images and Patterns (CAIP'01)*, pages 21–28, 2001.
14. G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence journal (TPAMI'07)*, 6(29):915–928, 2007.