# Spatiotemporal Gabor filters: a new method for dynamic texture recognition

Wesley Nunes Gonçalves, Bruno Brandoli Machado, Odemir Martinez Bruno

Physics Institute of São Carlos (IFSC)

University of São Paulo (USP)

Av. Trabalhador São-carlense, 400

Cx. Postal 369 - São Carlos - SP - Brasil

wnunes@ursa.ifsc.usp.br, brandoli@icmc.usp.br, bruno@ifsc.usp.br

## Abstract

*This paper presents a new method for dynamic texture recognition based on spatiotemporal Gabor filters. Dynamic textures have emerged as a new field of investigation that extends the concept of self-similarity of texture image to the spatiotemporal domain. To model a dynamic texture, we convolve the sequence of images to a bank of spatiotemporal Gabor filters. For each response, a feature vector is built by calculating the energy statistic. As far as the authors know, this paper is the first to report an effective method for dynamic texture recognition using spatiotemporal Gabor filters. We evaluate the proposed method on two challenging databases and the experimental results indicate that the proposed method is a robust approach for dynamic texture recognition.*

## 1. Introduction

The vision of animals provides a large amount of information that improves the perception of the world. This information is processed into different dimensions, including color, shape, illumination, and motion. While most of the features provide information about the static world, the motion provides essential information for interaction with external environment. In recent decades, the perception and interpretation of motion have attracted a significant interest in computer vision community [14, 16, 1, 9] motivated by the importance in both scientific and industrial communities. Despite significant advances, the motion characterization is still an open problem.

For modeling image sequences, three classes of motion patterns have been suggested [13]: dynamic texture, activities and events. The main difference between them relies on the temporal and spatial regularity of the motion field. In this work, we aim at modeling dynamic textures, also called temporal textures. They are basically texture in motion which is an extension of image texture to the spatiotemporal domain. Examples of dynamic texture includes real world scenes of fire, flag blowing, sea waves, moving escalator, boiling water, grass, and steam.

Existing methods for dynamic texture can be classified into four categories according to how they model the sequence of images. Due to the efficient estimation of features based on motion (e.g. optical flow), the motion based methods (i) are the most popular ones. These methods model dynamic textures based on a sequence of motion patterns [9, 13, 8]. For modeling dynamic texture at different scales in space and time, the spatiotemporal filtering based methods (ii) use spatiotemporal filters such as wavelet transform [7, 17, 5]. Model based methods (iii) are generally based on linear dynamical systems, which provides a model that can be used in applications of segmentation, synthesis, and classification [6, 3, 15]. Based on properties of moving contour surfaces, spatiotemporal geometric property based methods (iv) extract motion and appearance features from the tangent plane distribution [10]. The reader may consult [4] for a review of dynamic texture methods.

In this paper, we propose a new approach for dynamic texture modeling based on spatiotemporal Gabor filters [12]. As far as the authors know, the present paper is the first one to model dynamic texture using spatiotemporal Gabor filters. These filters are basically built using two parameters: the speed $v$ and direction $\theta$. To model a dynamic texture, we convolve the sequence of images to a bank of spatiotemporal Gabor filter built with different values of speed and direction. For each response, a feature vector is built by calculating the energy statistic.

We evaluate the proposed method by classifying dynamic texture from two challenging databases: dyntex [11] and traffic database [2]. Experimental results in both databases indicate that the proposed method is an effective approach for dynamic texture recognition. For the dyntex database, filter with low speeds (e.g. 0.1 pix-

els/frame) achieved better results than high speeds. In fact, the dynamic texture in this database presents low motion patterns. On the other hand, for the traffic database, high speeds (e.g. 1.5 pixels/frame) achieved the best correct classification rate. In this database, vehicles are moving at a speed that matches the filter's speed.

This paper is organized as follows. Section 2 briefly describes spatiotemporal Gabor filters. In Section 3, we present the proposed method for dynamic texture recognition based on spatiotemporal Gabor filters. An analysis of the proposed method with respect to the speed and direction parameters is present in Section 4. Experimental results are given in Section 5, which is followed by the conclusion of this work in Section 6.

## 2. Spatiotemporal Gabor Filters

Gabor filters are based on the important finding made by Hubel and Wiesel in the beginning of the 1960s. They found that the neurons of the primary visual cortex respond to lines or edges of a certain orientation in different positions of the visual field. Following this discovery, computational models were proposed for modeling the function of this neurons and the Gabor functions proved to be suited for this purpose in many works.

Initially, the researches aimed at studying spatial properties of the receptive field. However, some posterior studies revealed that cortical cells change in time and some of them are inseparable functions of space and time. Therefore, these cells are essentially spatiotemporal filters and they combine information over space and time, which makes a great model for dynamic texture analysis.

In this work, the spatiotemporal receptive field is modeled by a family of 3D Gabor filters [12] described in Equation 1.

$$
\begin{aligned}
g_{(v,\theta,\varphi)}(x,y,t) = {} & \frac{\gamma}{2\pi\sigma^2} \exp(\frac{-((\overline{x}+v_c t)^2 + \gamma^2 \overline{y}^2)}{2\sigma^2}) \\
& . \cos(\frac{2\pi}{\lambda}(\overline{x}+vt)+\varphi) \\
& . \frac{1}{\sqrt{2\pi}\tau} \exp(\frac{-(t-\mu_t)^2}{2\tau^2}) \\
& \overline{x} = x\cos\theta + y\sin\theta \\
& \overline{y} = -x\sin\theta + y\cos\theta
\end{aligned}
\tag{1}
$$

We now discuss the parameters of the spatiotemporal Gabor filter. Some parameters were empirically found based on studies of response in the receptive visual field [12]. The size of the receptive field is determined by the standard deviation $\sigma$ of the Gaussian factor. The parameter $\gamma$ is the rate that specifies the ellipticity of the Gaussian envelope in the spatial domain. This parameter is set to $\gamma = 0.5$ for match
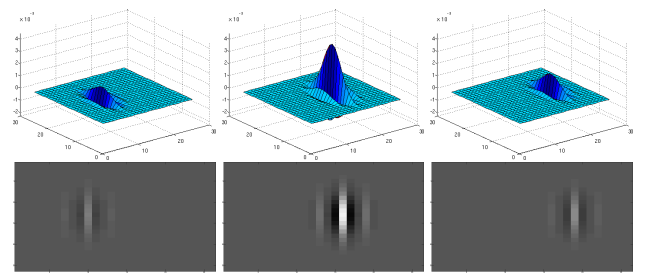


**Figure 1. Example of spatiotemporal Gabor filter for $v_c = 1$.**

to the elongated receptive field along $\overline{y}$ axis. The speed $v$ is the phase speed of the cosine factor, which determines the speed of motion. The speed which the center of the spatial Gaussian moves along the $\overline{x}$ axis is specified by the parameter $v_c$. When $v_c = 0$, the center of the Gaussian envelope is stationary. On the other hand, a moving envelope is obtained when $v_c = v$. Figure 1 presents a moving envelope with $v_c = v = 1$.

The parameter $\lambda$ is the wavelength of the cosine factor. It is obtained through the relation $\lambda = \lambda_0\sqrt{1+v^2}$, where the constant $\lambda_0 = 2$. The angle $\theta \in [0, 2\pi]$ determines the direction of motion and the spatial orientation of the filter. The phase offset $\varphi \in [-\pi, \pi]$ determines the symmetry in the spatial domain. The Gaussian distribution with mean $\mu_t$ and standard deviation $\tau$ is used to model the change in intensities. The mean $\mu_t = 1.75$ and standard deviation $\tau = 2.75$, both parameters are fixed based on the mean duration of most receptive fields.

## 3. Dynamic Textures Modeling based on Spatiotemporal Gabor Filters

In this section, we describe the proposed method for dynamic texture modeling based on spatiotemporal Gabor filters. Briefly, the sequence of images is convolved with a bank of spatiotemporal Gabor filters and a feature vector is constructed with the energy of the responses as components.

The response $r_{(v,\theta,\varphi)}(x,y,t)$ of a spatiotemporal Gabor filter $g_{(v,\theta,\varphi)}$ to a sequence of images $I(x,y,t)$ is computed by convolution:

$$
r_{(v,\theta,\varphi)}(x,y,t) = I(x,y,t) * g_{(v,\theta,\varphi)}
\tag{2}
$$

Spatiotemporal Gabor filters are phase sensitive because its response to a moving pattern depends on the exact position within the sequence of images. To overcome this drawback, a response that is phase insensitive can be obtained

by:

$$R_{(v,\theta)} = \sqrt{r^2_{(v,\theta,\varphi)}(x,y,t) + r^2_{(v,\theta,\varphi/2)}(x,y,t)} \quad (3)$$

To characterize the Gabor space resulting from the convolution, the energy of the response is computed according to Equation 4.

$$E_{(v,\theta)} = \sum_{x,y,t} R^2_{(v,\theta)}(x,y,t) \quad (4)$$

A central issue in applying spatiotemporal Gabor filters is the determination of the filter parameters that covers the spatiotemporal frequency space, and captures dynamic texture information as much as possible. Each spatiotemporal Gabor filter is determined by two main parameters: the direction $\theta$ and the speed of motion $v$. In order to cover a wide range of dynamic textures, we design a bank of spatiotemporal Gabor filter using a set of values for speed $V = [v_1, v_2, \ldots, v_n]$ and direction $\Theta = [\theta_1, \theta_2, \ldots, \theta_m]$. The feature vector that characterizes the dynamic texture is composed by the energy of the response for each combination of velocities $V$ and direction $\Theta$ (Equation 5).

$$\varphi = [E_{(v_1,\theta_1)}, \ldots, E_{(v_n,\theta_m)}] \quad (5)$$

The proposed method is summarized in Figure 2. First, we design a bank of spatiotemporal Gabor filters composed by filters with different directions and speeds. Then, the sequence of images is convolved with the bank of filters. For each convolved sequence of images, we calculate the energy to compose a feature vector.

## 4. Response Analysis of Spatiotemporal Gabor Filters

Here, we analyze the speed and direction properties of the spatiotemporal Gabor filters in synthetic sequence of images. In Figure 3, we present the response of spatiotemporal Gabor filters to bars moving at the same speed but in different direction $\theta$. The filters and the moving bars have preference for the same speed $v = 1$. The response has the highest magnitude when the direction of the filter $\theta$ matches the direction of the moving bar. For instance, when $\theta = 0$, a vertical bar moving rightwards evokes higher response than bars with other direction of movement.

The speed property is evaluated in Figure 4. We analyze the response of spatiotemporal Gabor filters to edges drifting rightward at different speeds. The filters and the synthetic sequence of images have for the same preference direction $\theta = 0$. The highest response is achieved by filters which the speed matches the speed of the moving edge.

In Figure 5(a), we plot the response of filters to a moving bar with direction $\theta = 0$ at a speed $v = 1$. The response reaches its maximum value to a filter with direction
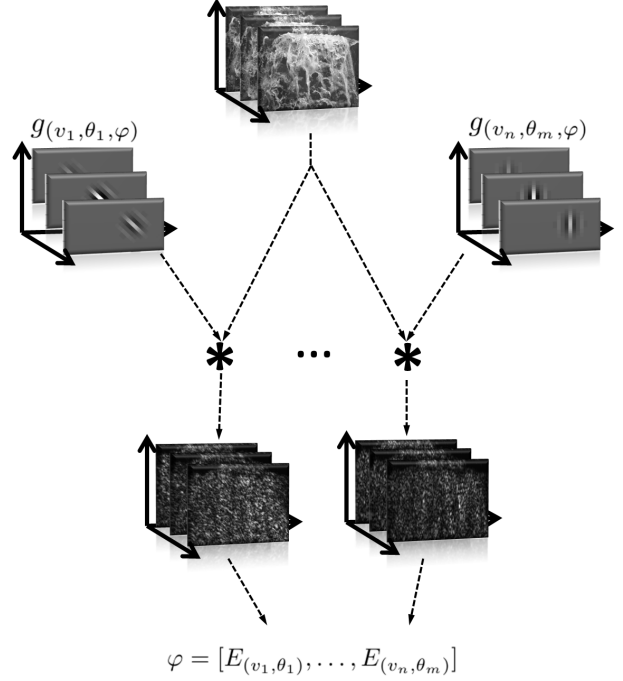


$$\varphi = [E_{(v_1,\theta_1)}, \ldots, E_{(v_n,\theta_m)}]$$

**Figure 2. Proposed method considers the following steps: (i) Design a bank of spatiotemporal Gabor filters using different values of speed $v$ and direction $\theta$; (ii) Convolve the sequence of images with the bank of filters; (iii) Calculate the energy for each response to compose the feature vector.**

$\theta = 0$, which matches to the direction of the moving bar. As we can see, the filter with moving envelope ($v_c = v$) achieved higher response than a filter with stationary envelope ($v_c = 0$). The plot for the speed parameter is shown in Figure 5(b). We convolved filters to an edge drifting rightward in direction $\theta = 0$ at a speed of $v = 2$. The maximum response is achieved for the filter whose speed matches to stimulus' speed. Again, we can conclude that filters with the moving envelope are more selective for both direction and speed than filters with stationary envelope.

## 5. Experimental Results

In this section, we present the experimental results using two databases: (i) dyntex database and (ii) traffic video database. The dyntex database consists of 50 dynamic texture classes each containing 10 samples collected from Dyntex database [11]. The videos are at least 250 frames long with dimension of $400 \times 300$ pixels. Figure 6 shows
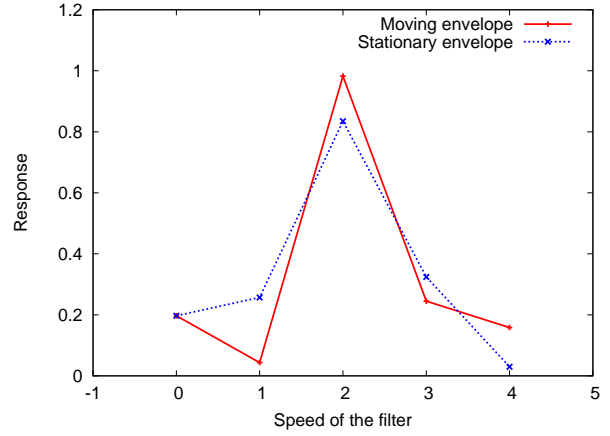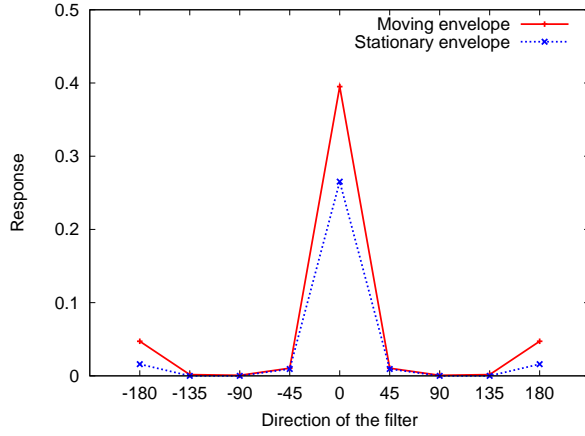
**Figure 5. (a) Response for filters of different direction $\theta$ to a moving bar with $\theta = 0$. (b) Response for filters of different speeds $v$ to a edge moving at speed $v = 2$.**
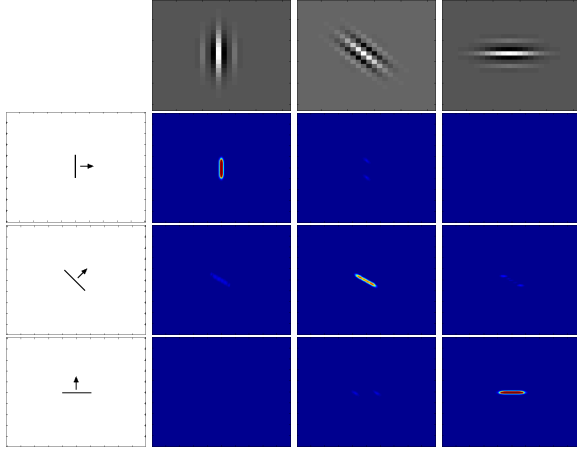


**Figure 3. Response of spatiotemporal Gabor filters to bars moving in different direction $\theta$. First row corresponds to the filters. Second, third and fourth rows corresponds to the response of a bar moving in direction $\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}$, respectively.**
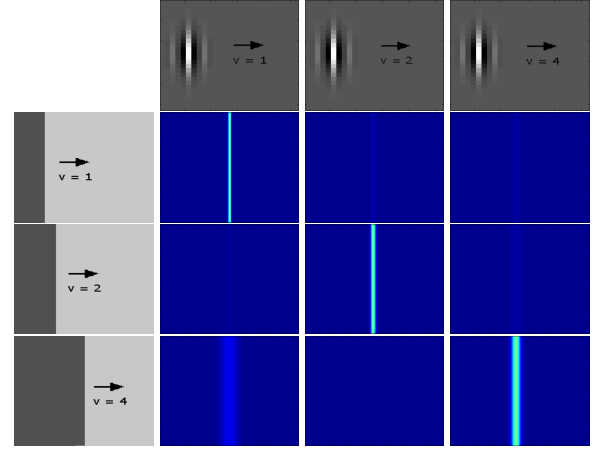


**Figure 4. Response of spatiotemporal Gabor filters to edges moving at different speed $v$. First, second, and third rows corresponds to the response of a edges moving at speed $v = 1, 2, 4$, respectively.**

examples of dynamic textures from the first database. The second database, collected from traffic database [2], consists of 254 videos divided into three classes: light, medium, and heavy traffic. Videos had 42 to 52 frames with a resolution of $320 \times 240$ pixels. The variety of traffic patterns and weather conditions are shown in Figure 7. All the experiments used a k-nearest neighbor classifier with $k = 1$ in a scheme 10-fold cross validation.

Now, we discuss the influence of the direction and speed

parameters in the dynamic texture recognition. Table 1 shows the average and standard deviation of correct classification rate for the traffic database. Columns present the direction parameter evaluation using 4-directions, which is a combination of filters with $\theta = [0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}]$, and 8-directions which is a combination of filters with $\theta = [0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}]$. Rows present the speed evaluation using combination of speed with step of 0.25. As we can see, the bank of filter composed by filters with 8-directions outperformed the bank of filter with 4-directions
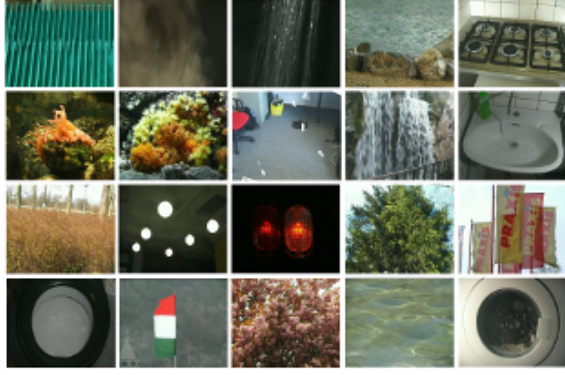
**Figure 6. Examples of dynamic textures from dyntex database [11]. The database is composed by 50 dynamic texture classes each containing 10 samples.**

| Speed Combination | 4-directions | 8-directions |
|---|---|---|
| $[0.50 - 0.75]$ | 90.06(5.53) | 90.18(5.62) |
| $[0.50 - 1.00]$ | 90.17(5.70) | 90.57(5.15) |
| $[0.50 - 1.25]$ | 89.07(5.91) | 89.90(5.06) |
| $[0.50 - 1.50]$ | 89.68(5.99) | 89.74(5.80) |
| $[0.50 - 1.75]$ | 89.69(6.18) | 90.56(5.58) |
| $[0.50 - 2.00]$ | 89.45(5.77) | **91.50(5.20)** |
| $[0.50 - 2.25]$ | 90.24(5.51) | 90.87(5.07) |
| $[0.50 - 2.50]$ | 89.33(5.60) | 90.75(5.26) |
| $[0.50 - 2.75]$ | 89.76(5.44) | 90.35(5.44) |
| $[0.50 - 3.00]$ | 89.60(5.16) | 90.63(5.21) |

**Table 1. Correct classification rate and standard deviation for different combinations of speed and direction on the traffic database.**

| Speed Combination | 4-directions | 8-directions |
|---|---|---|
| $[0.1 - 0.2]$ | 92.50(3.40) | 94.92(3.09) |
| $[0.1 - 0.5]$ | 96.00(2.40) | 96.82(2.41) |
| $[0.1 - 1.0]$ | 96.56(2.11) | 98.02(1.65) |
| $[0.1 - 1.5]$ | 96.92(2.32) | **98.60(1.60)** |
| $[0.1 - 2.0]$ | 97.24(2.18) | 97.84(1.92) |
| $[0.1 - 2.5]$ | 96.92(2.49) | 97.34(2.36) |
| $[0.1 - 3.0]$ | 96.37(2.98) | 96.94(2.73) |

**Table 2. Correct classification rate and standard deviation for different combinations of speed and direction on the dyntex database.**
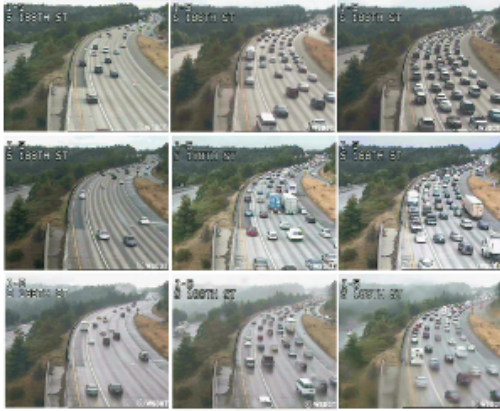


**Figure 7. Examples of dynamic textures from traffic database [2]. The database consists of 254 videos split into three classes of light, medium and heavy traffic.**

for all combination of speeds. However, very little improvement can be appreciated as the combination of direction is increased. The improvement in correct classification rate was on average less than $1\%$ when direction combination rises from 4 to 8. This is because the cars on the highway move always on the same direction, which can be modeled by 4-direction. With respect to the speed parameter, the best results were achieved for high speeds, such as 1.5 pixels/frame and 2.0 pixel/frame. A correct classification rate of $91.50\%$ was achieved by a bank of filter composed by 8-directions and speeds $[0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0]$. The high speeds match to speed of the cars in the sequence of images and then the traffic jam can be modeled using these parameters.

In Table 2, we present the experimental results obtained on the dyntex database. The same combination of directions of the early experiment was used to evaluate the proposed method. However, as the dynamic textures in this database present low speeds, the combination of speed started in $v = 0.1$ pixels/frame and taken step of 0.1. As the previous results, the 8-direction bank of filters achieved higher values of correct classification rate compared to the 4-direction bank of filters. In this case, a correct classification rate of $98.60\%$ was obtained, which clearly shows the effectiveness of the proposed method on the dynamic texture recognition.

## 6. Conclusion

In this paper, we proposed a new method for dynamic texture recognition based on spatiotemporal Gabor filters. First, it convolves a sequence of images to a bank of filters

and then extracts energy statistic from each response. Basically, the spatiotemporal Gabor filters are built using speed $v$ and direction $\theta$. The bank of filter is composed by filters with different values of speed and direction.

Promising results considering different combinations of $v$ and $\theta$ were achieved on two important databases: dyntex and traffic database. On the traffic database, our method achieved a correct classification rate of $91.50\%$ using combination of high speeds. On the other hand, a correct classification rate of $98.60\%$ was obtained on the dyntex database using a combination of low speeds.

# References

[1] C. Cedras and M. Shah. Motion-based recognition: A survey. *Image and Vision Computing*, 13(2):129–155, mar 1995.

[2] A. B. Chan and N. Vasconcelos. Classification and retrieval of traffic video using auto-regressive stochastic processes. In *IEEE Intelligent Vehicles Symposium*, pages 771–776, June 2005.

[3] A. B. Chan and N. Vasconcelos. Classifying video with kernel dynamic textures. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–6, 2007.

[4] D. Chetverikov and R. Péteri. A brief survey of dynamic texture description and recognition. In *Computer Recognition Systems, Proceedings of the 4th International Conference on Computer Recognition Systems*, volume 30 of *Advances in Soft Computing*, pages 17–26. Springer, 2005.

[5] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In *IC-CCN '05: Proceedings of the 14th International Conference on Computer Communications and Networks*, pages 65–72, Washington, DC, USA, 2005. IEEE Computer Society.

[6] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto. Dynamic textures. *International Journal of Computer Vision*, 51(2):91–109, February 2003.

[7] S. Dubois, R. Péteri, and M. Ménard. A comparison of wavelet based spatio-temporal decomposition methods for dynamic texture recognition. In *IbPRIA '09: Proceedings of the 4th Iberian Conference on Pattern Recognition and Image Analysis*, pages 314–321, Berlin, Heidelberg, 2009. Springer-Verlag.

[8] R. Fablet and P. Bouthemy. Motion recognition using non-parametric image motion models estimated from temporal and multiscale cooccurrence statistics. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(12):1619–1624, 2003.

[9] S. Fazekas and D. Chetverikov. Analysis and performance evaluation of optical flow features for dynamic texture recognition. *SP:IC*, 22(7-8):680–691, Aug. 2007.

[10] M. Fujii, T. Horikoshi, K. Otsuka, and S. Suzuki. Feature extraction of temporal texture based on spatiotemporal motion trajectory. In *ICPR*, pages Vol II: 1047–1051, 1998.

[11] R. Peteri, S. Fazekas, and M. Huiskes. Dyntex: A comprehensive database of dynamic textures. *Pattern Recognition Letters*, 31(12):1627–1632, September 2010.

[12] N. Petkov and E. Subramanian. Motion detection, noise reduction, texture suppression, and contour enhancement by spatiotemporal gabor filters with surround inhibition. *Biological Cybernetics*, 97:423–439, March 2008.

[13] R. Polana and R. C. Nelson. Temporal texture and activity recognition. In *Motion-Based Recognition*, page Chapter 5, 1997.

[14] H. Qian, Y. Mao, W. Xiang, and Z. Wang. Recognition of human activities using svm multi-class classifier. *Pattern Recognition Letters*, In Press, Corrected Proof:–, 2009.

[15] M. Szummer and R. W. Picard. Temporal texture modeling. In *ICIP*, pages III: 823–826, 1996.

[16] S. Wu and Y. F. Li. Motion trajectory reproduction from generalized signature description. *Pattern Recognition*, 43(1):204–221, 2010.

[17] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 819–826, 2004.