

A VLSI-ORIENTED AND POWER-EFFICIENT APPROACH FOR DYNAMIC TEXTURE RECOGNITION APPLIED TO SMOKE DETECTION

Jorge Fernández-Berni, Ricardo Carmona-Galán, Luis Carranza-González

Institute of Microelectronics of Seville (IMSE-CNM)

Consejo Superior de Investigaciones Científicas y Universidad de Sevilla

Avda. Reina Mercedes s/n 41012, Seville, Spain

berni@imse.cnm.es

Keywords: Dynamic texture recognition, power-efficient VLSI implementation, smoke detection, forest fire detection

Abstract: The recognition of dynamic textures is fundamental in processing image sequences as they are very common in natural scenes. The computation of the optic flow is the most popular method to detect, segment and analyse dynamic textures. For weak dynamic textures, this method is specially adequate. However, for strong dynamic textures, it implies heavy computational load and therefore an important energy consumption. In this paper, we propose a novel approach intended to be implemented by very low-power integrated vision devices. It is based on a simple and flexible computation at the focal plane implemented by power-efficient hardware. The first stages of the processing are dedicated to remove redundant spatial information in order to obtain a simplified representation of the original scene. This simplified representation can be used by subsequent digital processing stages to finally decide about the presence and evolution of a certain dynamic texture in the scene. As an application of the proposed approach, we present the preliminary results of smoke detection for the development of a forest fire detection system based on a wireless vision sensor network.

1 INTRODUCTION

A temporal texture or dynamic texture (DT) is a spatially-repetitive time-varying visual pattern whose temporal variation presents certain stationarity (Nelson and Polana, 1992). An additional feature of a DT is its indeterminate spatial and temporal extent. Smoke, waves, a flock of birds or tree leaves swaying in the wind are some examples of DTs.

The recognition of DTs plays an essential role in image processing as they are very common in natural scenes. Different methods have been proposed to realize this process of recognition, as described in a recent review (Chetverikov and Péteri, 2005). The methods based on optic flow are currently the most popular. Optic flow is a computationally efficient and natural way to characterise the local dynamics of a temporal texture. Specially, this is the case for weak dynamic textures, that is, textures defined by a local moving coordinate system in which they become static. However, the recognition of strong dynamic textures implies a much greater computational effort. For these textures, possessing intrinsic dynamics, the

brightness constancy assumption associated to standard optical flow algorithms cannot be applied. More complex approaches must be considered in order to overcome this problem. Recently, interesting results have been achieved by applying the so-called brightness conservation assumption (Amiaz et al., 2007). However, this method means heavy computational load and the subsequent high energy consumption.

There are certain systems where a power-efficient implementation of DT recognition is totally mandatory. Wireless multimedia sensor networks (Akyildiz et al., 2007) is an obvious example. These networks are composed of a large number of low-power sensors that are densely deployed throughout a region of interest in order to capture and analyse video, audio and environmental data from their surroundings. The massive and scattered deployment of these sensors makes them quite difficult to service and maintain. Therefore, energy efficiency must be a major design goal of this kind of systems in order to prolong the lifetime of the batteries as much as possible.

In this paper, we propose a novel approach to detect DTs in scenes surveilled by very low-power inte-

grated vision devices. It is based on a signal processing architecture where redundant spatial information is removed at the very first stages of the processing by means of a simple, flexible and power-efficient computation at the focal plane. The final decision about the presence of a certain DT in the scene is realized by analysing the reduced representation of the original data obtained from the focal plane. Thus, the computational load is greatly alleviated.

As an application of this approach, we study the smoke detection within a wireless vision sensor network devoted to forest fire detection. The results point to a very high reliability and robustness in the process of detection.

2 BINNING PROCESS

In general, existing research on DT recognition is based on global features computed over the whole scene. A clear sign of this fact is that practically all of the sequences composing the reference database DynTex (Péteri et al., 2006) contain only close-ups of DTs. For these sequences, it does make sense to apply strategies of global feature recognition over the whole scene. However, there are interesting applications of DT recognition, e.g. video-surveillance, where textures can appear at any location of the scene. In this case, a previous detection and subsequent analysis of candidate regions to contain a certain DT could reduce the computational load by progressively reducing the amount of data to process. To detect such candidate regions, we can take advantage of the spatial repeatability of patterns in DTs. In this way, we are going to divide the scene into blocks, or bins, whose size S_B is defined, in pixels, as:

$$S_B = W \times H \quad (1)$$

The fundamental concept of the proposed approach is that each bin of the scene, labelled as (i, j) , can be considered as an independent entity capable of detecting the presence of a determined spatial pattern within it. When a bin detects the spatial pattern, it is marked as a candidate bin to contain a part of a DT whose spatially-repetitive pattern coincides with the detected pattern. A subsequent phase of spatio-temporal analysis of the candidate bins will eventually confirm or dismiss the presence of the DT. Note that, during this phase, the amount of data to be processed is reduced by a factor $W \times H$ with respect to the previous phase of candidate bin detection.

Every bin will be represented by only one value. This value must be defined in such a way that the presence of a certain spatial pattern within it can be eas-

ily detected. We propose to define the representative value of the bins in terms of spatial frequency information. For example, consider the detection of a flock of birds passing along the sky. A flock of birds in contrast with sky entails significant information at spatial frequencies greater than $\mathbf{0}$, which expresses the origin of coordinates in the Fourier space. Let us suppose that each bin (i, j) of the scene is represented by the following value:

$$B_{ij} = \frac{P_{ij}(\mathbf{k} > \mathbf{0})}{P_{ij}(\mathbf{k})} \quad (2)$$

where $P_{ij}(\mathbf{k} > \mathbf{0})$ is the power at the bin (i, j) for all of the spatial frequencies other than $\mathbf{k} = \mathbf{0}$ and $P_{ij}(\mathbf{k})$ is the power at the bin (i, j) for all the spatial frequencies including $\mathbf{k} = \mathbf{0}$. The bidimensional vector \mathbf{k} represents the 2π -normalized wavenumber vector. Applying this expression, the recognition of a spatial pattern similar to that of a flock of birds is straightforward. In Fig. 1 a grayscale binned image of a flock of birds is depicted along with its representation using Eq. 2. It can be seen that even a recognition of zones with different density of birds can be carried out from the value of B_{ij} . A global pixel-based analysis is not necessary.

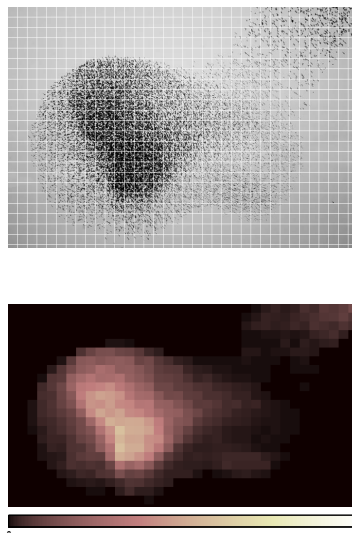


Figure 1: Example of processing to detect candidate bins.

This example allows to understand how the detection of DTs can be simplified by means of a binning process. However, a thorough analysis to establish an optimum value of S_B has not been realized. It is an essential aspect of the proposed approach. The value of S_B must be adequate to detect a certain spatial pattern as well as to track its temporal dynamics across the

scene. Therefore, it will depend not only on the spatial pattern but also on the spatio-temporal dynamics of the texture. Later in this paper we describe the tuning of S_B , along with other parameters, for smoke detection within the framework of forest fire detection.

3 SIGNAL PROCESSING ARCHITECTURE

The only element missing now is an efficient computation of the spatial frequency components present in each and every bin. Thus, moving part of the heaviest computational effort to the focal plane would result in a flexible and power-efficient architecture. This approach is inherent to biological vision systems, like the human retina, where visual information is captured and preprocessed, alleviating the data flow sent to the visual cortex (Roska and Werblin, 2001). In this way, low-level tasks, according to the classification in (Pirsch and Stolberg, 1998), whose requirements of accuracy are not too demanding, are realized in parallel by highly efficient, although moderately coarse, analog hardware at the focal plane. Physical implementations based on this processing architecture achieve higher performance with less cost and power (Carmona et al., 2003). In our case, we are going to convey the detection of candidate bins to the focal plane. A microprocessor would make the final decision about the presence or absence of the DT by analysing the preprocessed and therefore reduced information represented by these candidate bins. Note that this processing architecture simply removes redundant data at the focal plane in order to deliver to the microprocessor just the necessary information to track and detect the texture.

At this point, we need to define a flexible hardware structure at the focal plane in order to be able to detect any DT. Such a structure must satisfy two conditions:

1. The size of the bins S_B can take any value
2. Information about any particular band of spatial frequencies can be extracted at every bin

A structure fulfilling both conditions is depicted in Fig. 2. It consists of a $M \times N$ grid where the value of each pixel is stored in a capacitor. These capacitors are 4-connected to the neighboring capacitors by means of analog switches. These switches are controlled by the corresponding row or column selection signal. When selected, i. e. the control signal is high, the switch behaves as a resistor connecting the two nodes, whose utility will be explained in short. If the control signal is low the switch is in the boundary of

a bin. Thus the particular distribution of 0's and 1's in the set of row and column selection signals establishes the size and amount of bins in which the image plane is divided.

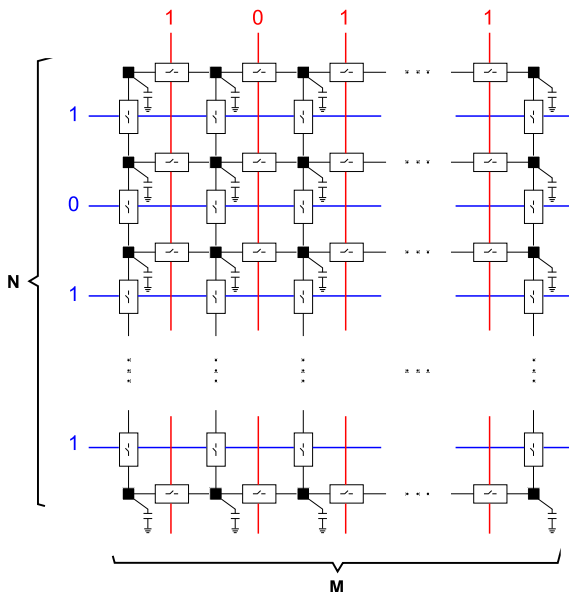


Figure 2: Processing structure at the focal plane.

Once S_B is determined, a resistive grid is established within every bin taking into account the resistance of the switches when they are on. Such a resistive grid can carry a linear diffusion of the pixel values within the bin as long as the switches remain ON, let us say a period of time Δt . Consider $y_{ij}(x, y)$ as a function defining the values of the pixels within the bin (i, j) . The linear diffusion can be expressed as (Jahne et al., 1999):

$$Y_{ij}(\mathbf{k}, \Delta t) = Y_{ij}(\mathbf{k}, 0) e^{-4\pi^2 D \Delta t |\mathbf{k}|^2} \quad (3)$$

where $Y_{ij}(\mathbf{k}, \Delta t)$ is the spatial Fourier transform of the subimage contained in the bin after Δt seconds and $Y_{ij}(\mathbf{k}, 0)$ is the transform at time $t = 0$, that is, just before starting the diffusion. The constant D is the diffusion coefficient.

From Eq. 3, the power of each frequency component during the process of diffusion is:

$$|Y_{ij}(\mathbf{k}, \Delta t)|^2 = |Y_{ij}(\mathbf{k}, 0)|^2 e^{-8\pi^2 D \Delta t |\mathbf{k}|^2} \quad (4)$$

which can be expressed as:

$$|Y_{ij}(\mathbf{k}, \Delta t)|^2 = [1 - \alpha(\mathbf{k}, \Delta t)] |Y_{ij}(\mathbf{k}, 0)|^2 \quad (5)$$

where $\alpha(\mathbf{k}, \Delta t)$ is the attenuation undergone by each component at frequency \mathbf{k} after Δt seconds of linear

diffusion. This attenuation increases along time and is more significant for higher frequencies.

Therefore, by controlling Δt we can obtain information at different bands of spatial frequencies. Consider the simplest case, a lowpass filter. We can specify a maximum attenuation for the passband α_P , with a cut-off spatial frequency of $|\mathbf{k}_P|$. At the same time, for the stopband, a minimum attenuation α_S is required starting at frequency $|\mathbf{k}_S|$. Then, the period of time after which the linear diffusion must be stopped in order to carry out the so defined filtering is between these two bonds:

$$\frac{|\ln(1 - \alpha_S)|}{8\pi^2 D |\mathbf{k}_S|^2} \leq \Delta t \leq \frac{|\ln(1 - \alpha_P)|}{8\pi^2 D |\mathbf{k}_P|^2} \quad (6)$$

Then, providing the means for storing and combining two consecutive samples of the diffused image, taken after Δt_1 and Δt_2 from the starting point of the diffusion, we can compute a bandpass, a highpass or a bandreject filter. Keep in mind, however, that the highest frequency feature that can be considered is determined by the size of the pixel, and therefore a real highpass filter is actually a bandpass filter.

A crucial property of the structure just defined is that the diffusion operation is really a simple charge redistribution which does not consume energy, i. e. it is realized by a passive network. That is to say, the signal processing within the structure just described is massively parallel and ultra power-efficient.

4 AN APPLICATION TO SMOKE DETECTION

As an application of the approach previously explained, we present a vision algorithm suitable for a forest fire detection system based on a wireless vision sensor network. The vision sensors will be placed on top of poles in order to focus the canopy of small vegetation areas. Each sensor of the network will run on-site a vision algorithm in order to detect smoke arising among the vegetation. When a sensor detects smoke, a warning message is sent to a control center by multihopping. This structure of the system, based on a careful placing of the sensors, reduces significantly the sources of false alarms with respect to its counterparts based on lookout towers with automatic surveillance. On the contrary, because of the necessary dense deployment of vision sensors, energy efficiency is a key point impacting the system cost.

The smoke detection algorithm is based on the analysis of the sequence of images captured by each sensor. We define the time interval between two consecutively captured images as T_C . These images are

compared to a reference image, the background, in order to detect changes generated by smoke dynamics. This method of motion detection, called background subtraction (Hu et al., 2004), is suitable for scenes with a relatively static background. In the case of the proposed system, the visual field is basically composed of vegetation, and therefore the background will hardly suffer significant sudden changes. It will experience, however, gradual illumination changes throughout the day. Taking this fact into account, the reference image will be updated every time interval T_R .

Regarding the binning process, note that the appearance of smoke in any scene means the equalization of the R, G and B components in the pixels affected by smoke (Chen et al., 2006). Therefore, it could be a criterion to detect candidate bins. However, we propose a more data-efficient option based only on the detection of sudden increases in the B component with respect to the background. The use of the B component is owing to its greater sensitivity in natural scenarios with vegetation to the changes generated by smoke when compared to the R and G components and combined luminance. This will also discard changes introduced by sources different than smoke, like the motion of tree leaves.

Consider Fig. 3. We have highlighted different zones in three scenes where the background is mainly constituted by vegetation. These frames correspond to different parts of the associated sequences, before and after a trace of smoke has appeared in the field of view. Table 1 shows the normalized average increase, referred to the image without smoke, undergone by each RGB component and the combined luminance of the pixels within the marked zones in presence of smoke. It can be seen that, in all cases, the appearance of smoke among vegetation conveys a greater increase in the B component than that observed in the R and G components and the combined luminance.

Therefore, each bin will be represented by the average value of the B component of its pixels. In terms of spatial frequencies, it is equivalent to say that we are interested in the information contained at $\mathbf{k} = \mathbf{0}$. The value of S_B will depend on the average variation in the number of smoke pixels during their appearance in the scene: the less the average variation, the smaller the necessary size of the bins to track the dynamics. Consider a sequence of k consecutive images containing smoke captured by a sensor. The average variation in the number of smoke pixels will be:

$$\bar{v} = \frac{\sum_{i=1}^k |U_T[i] - U_T[i-1]|}{k} \quad (7)$$

where $U_T[i]$ is the total number of smoke pixels of the

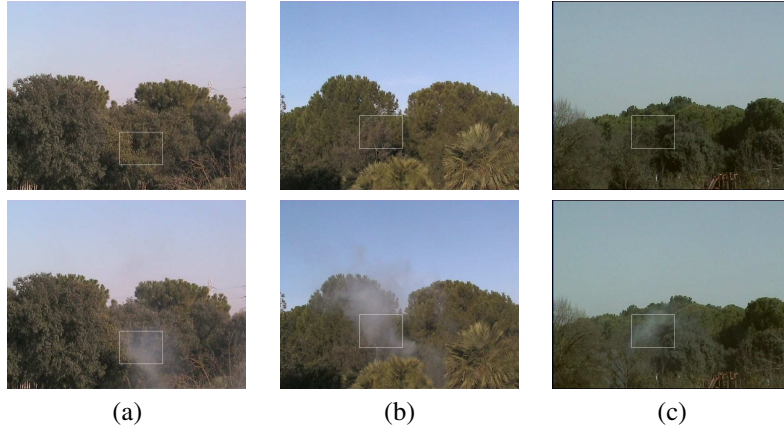


Figure 3: Three typical scenes with vegetation background without smoke and in presence of smoke.

Table 1: Normalized average increase, referred to the case without smoke, suffered by each component in the zones marked in Fig. 3.

	R component	G component	B component	Luminance
Scene (a)	13.8%	14.5%	20%	14.9%
Scene (b)	12.5%	13.4%	19.5%	13.8%
Scene (c)	11%	12.8%	16.4%	12.7%

i – th image of the sequence (suppose $U_T[0] = 0$). According to this expression, the smoke dynamics represented by these images can be correctly tracked whenever S_B fulfills the following condition:

$$S_B \leq \bar{V} \quad (8)$$

Finally, the condition which must be satisfied in order to consider a foreground bin B_{Fij} as a candidate bin to contain smoke is:

$$B_{Fij} - B_{Bij} > \frac{p_{th}}{100}(B_{MAX} - B_{MIN}) \quad (9)$$

where p_{th} represents the threshold percentage of the B component signal range, set by $(B_{MAX} - B_{MIN})$, which the B component of a foreground bin B_{Fij} must rise with respect to the corresponding background bin B_{Bij} .

Once the candidate bins are detected, the algorithm analyses them looking for the spatio-temporal dynamics which is characteristic of smoke. This stage of the algorithm is divided into two phases: the detection and the confirmation phase. The detection phase starts when the first candidate bins are discovered, an instant denoted as t_0 , and finishes at $t = t_D$. Then the confirmation phase is started, and will finish, if the result is positive, at time $t = t_F$, by sending an alarm message. The internal processing at both phases is described next.

First of all, in order to consider that smoke is present at the scene, a minimum number of candi-

date bins must exist. Let us define $N(t)$ as the number of candidate bins at time instant t . This parameter changes every T_C , that is, with every captured image. During the confirmation phase, the following expression must be fulfilled:

$$N(t) \geq N_{MIN} \quad \{t \in [t_D, t_F]\} \quad (10)$$

where N_{MIN} represents the minimum necessary number of candidate bins to consider smoke.

Another important element of the smoke dynamics is its gradual appearance into the scene. Once the first candidate bins are detected, new candidate bins must gradually appear until reaching at least N_{MIN} at $t = t_D$. This fact can be described by means of two conditions. The first one is:

$$t_D - t_0 \leq T_{D_{MAX}} \quad (11)$$

where $T_{D_{MAX}}$ represents the maximum time interval within which smoke must appear into the scene once the first candidate bins are detected. The second condition is:

$$N(t) - N(t - T_C) \leq G_{MAX} \quad \{t \in [t_0, t_F]\} \quad (12)$$

where G_{MAX} expresses the maximum permitted growth of candidate bins between two consecutive captured images during the smoke dynamics.

Finally, smoke does not appear as candidate bins scattered throughout the scene. On the contrary, it

is formed by compact regions of candidate bins. Let us define $Z(t)$ as the number of 8-connected regions of candidate bins. Just like $N(t)$, $Z(t)$ changes with every captured image. The compactness of smoke can be described as:

$$Z(t) \leq Z_{\text{MAX}} \quad \{t \in [t_0, t_F]\} \quad (13)$$

being Z_{MAX} the maximum permitted number of 8-connected regions during the smoke dynamics.

The duration of the confirmation phase will be fixed, that is:

$$t_F - t_D = T_F \quad (14)$$

Therefore, the spatio-temporal dynamics of smoke can be summarised as follows: a minimum number of candidate bins N_{MIN} must appear once the first candidate bins are discovered at $t = t_0$. The time instant at which N_{MIN} is reached, $t = t_D$, establishes the end of the detection phase and the beginning of the confirmation phase. Besides, t_D must fulfill Eq. (11), accounting in this way for the time scale of the gradual appearance of smoke in the scene. If the limit established by $T_{D_{\text{MAX}}}$ is reached without having accomplished the minimum candidate bins fixed by N_{MIN} , then the process starts again as the detection was triggered by something that was not smoke. If the detection is correct confirmation phase starts. During this phase, whose duration is determined by Eq. (14), the number of candidate bins must be always above N_{MIN} , i. e. they must satisfy Eq. (10). In this way, we take into account that smoke does not disappear suddenly from the scene. Finally, the growth rate of smoke, defined in Eq. (12), and the compactness of the smoke traces in the scene, represented by Eq. (13), are checked during both the detection phase and the confirmation phase. Failure in holding these conditions means tracking some dynamics belonging to a source different from smoke, what stops the process returning to the beginning of the detection cycle.

5 PRELIMINARY TESTS

In order to test the vision algorithm, we realized some recordings in natural scenarios. They were carried out in 9 different locations under different illumination conditions using three different cameras. Approximately 80 minutes were recorded containing 16 sequences of gradual appearance of smoke following its natural evolution in scenes whose background is basically composed of vegetation and numerous sequences without smoke in order to check the false alarm rate. The resolution of the frames is 720×576

px and the frame rate is 25 frames per second in all the recordings. As an example, the images in Fig. 3 correspond to three different tests.

To set the values of the parameters of the algorithm, we analysed 9 of the 16 different sequences where smoke appears gradually (Fernández-Berni, 2008). The other 7 sequences will be employed to test the algorithm later. Specifically, 25 seconds of every sequence were analysed. The starting point is just before smoke begins to appear. Only the B component of the frames was used for the adjustment.

The proposed algorithm presents a basic parameter: T_C . It establishes the temporal scale of the smoke dynamics. In order to determine its value, we first detected the pixels affected by smoke. Indeed, smoke is the most significant foreground motion in the 9 sequences analysed. Therefore, we can consider that a pixel represents smoke if:

$$p_F(x, y) - p_B(x, y) > \varepsilon \quad (15)$$

where $p_F(x, y)$ is the foreground value of the pixel (x, y) , $p_B(x, y)$ is its background value and ε is the pixel noise of the camera.

Once the smoke pixels are detected, the parameter T_C must reflect the temporal scale of its dynamics. At the same time, this parameter must be as large as possible to reduce the processing load. Consider Fig. 4. It shows a magnitude that is very sensitive to the smoke dynamics: the number of smoke pixels per frame of every sequence. If we examine the evolution of these curves in the frequency domain, it can be seen that most of the power is concentrated at very low frequencies. In fact, at $f = 0.5\text{Hz}$ the power of the DFT is approximately three orders of magnitude smaller than the power at $f = 0\text{Hz}$ for all the sequences. This means that the essential characteristics of the dynamics can be tracked by only analysing the frequency interval $[0, 0.5]\text{Hz}$. In terms of T_C , it is translated into $T_C = 1\text{s}$, that is, a sample frequency of 1Hz . Therefore, the adjustment of the rest of parameters will be realized taking into account that $T_C = 1\text{s}$.

To determine S_B , we represent \bar{V} for each sequence in Fig. 5. It can be seen that the minimum value is approximately 200 smoke pixels per second. According to Eq. (8), this is the maximum number of pixels which must contain every bin in order to track the smoke dynamics. Taking into account the resolution of the frames, the size of the bins must be 15×12 pixels (180 pixels). It is important to emphasize that, thanks to the binning process, the initial information composed of 0.4Mpx is reduced to only 2.3Kbins. Besides, in this case, every bin can be represented by only one bit indicating if it is marked as a candidate bin or not.

Regarding the percentage p_{th} , consider Fig. 6. It represents the average increase, normalized to the signal range, of the pixels affected by smoke with regard to the background. It can be seen that the average increase exceeds 10% for most of the time. We are going to extrapolate this result to the candidate bins by setting $p_{th} = 10\%$.

Applying Eq. (9) with the values of S_B and p_{th} just set, we obtain the number of candidate bins of the sequences, as depicted in Fig. 7. According to this representation, the minimum number of candidate bins reached in the sequences is 17. We set $N_{MIN} = 14$ in order to concede a margin of three candidate bins. This choice implicitly sets $T_{D_{MAX}} = 20s$ and $T_F = 4s$. Therefore, the detection of smoke will be carried out within a maximum interval of 24s after the discovery of the first candidate bins.

At this point, it is easy to adjust the value of G_{MAX} . Consider Fig. 8, where the maximum growth rate of candidate bins is represented for each sequence. Among these maximum rates, the greatest is 47 bins per second. We set $G_{MAX} = 50$, adding a small margin of three bins.

There are two parameters left: Z_{MAX} and T_R . To determine Z_{MAX} , consider the Fig. 9 where the number of 8-connected regions of candidate bins is depicted throughout the sequences. We can see that the maximum value is 6. Besides, it is only reached once in one of the sequences. Therefore, we set $Z_{MAX} = 6$, without additional margin.

Finally, with respect to T_R , longer recordings are mandatory in order to estimate an approximate value of this parameter. Even for the longest sequence analysed, whose duration is 458s (around 8 minutes), smoke detection was possible without updating the initial reference image.

Once the parameters were set, we applied the algorithm, to all the recordings. Smoke was detected in the 16 smoke sequences, the 9 employed to set the parameters, obviously, and the other 7. Besides, in the rest of sequences where smoke was not present, no false alarm was detected. Therefore, the algorithm achieves the highest reliability when applied to all our recordings.

6 CONCLUSIONS

A new generic approach to detect dynamic textures has been presented. It is specially suitable for real-time applications where the power requirements demand very low energy consumption. As an application, we propose the smoke detection within a forest fire detection system based on a wireless vision sen-

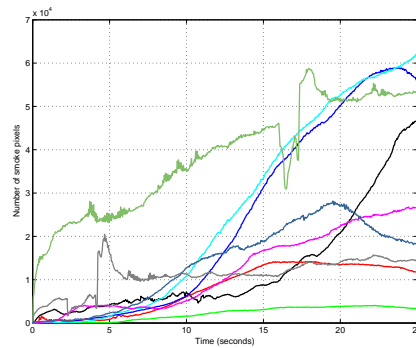


Figure 4: Number of smoke pixels per frame.

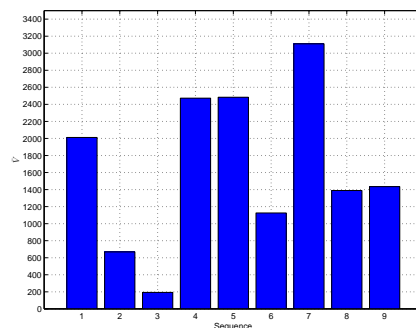


Figure 5: Average variation in the number of smoke pixels for each sequence.

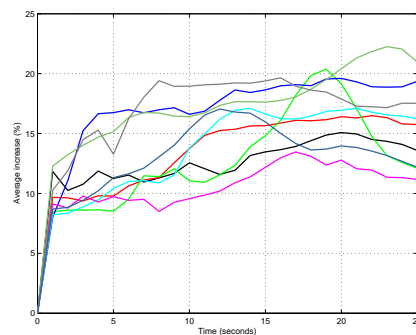


Figure 6: Normalized average increase of the pixels affected by smoke.

sor network. A set of recordings was carried out in natural scenarios in order to validate the feasibility of the proposed approach for this application. The results point to a very high reliability and robustness in the process of detection. Our primary next objective is the physical implementation of a vision chip whose focal plane includes the structure proposed in this paper. We also intend to extend the applicability of the proposed approach by defining new vision algorithms which can take significant advantage of including both binning and spatial filtering within each

bin.

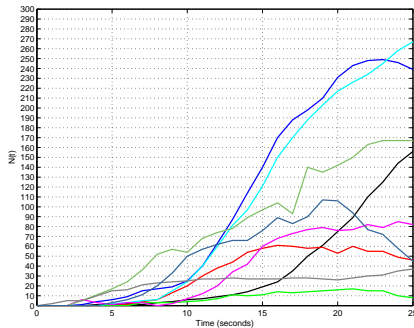


Figure 7: Number of candidate bins at every second of the sequences.

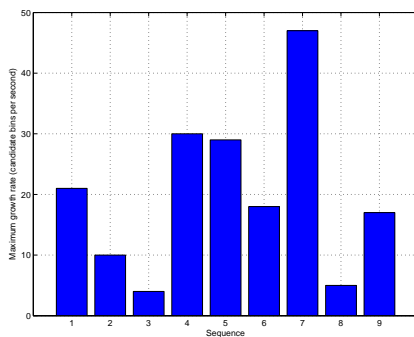


Figure 8: Maximum growth rate of candidate bins in each sequence.

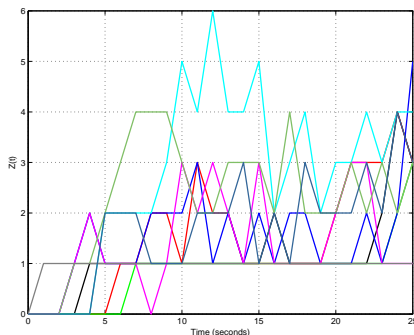


Figure 9: Number of 8-connected regions of candidate bins during the sequences

ACKNOWLEDGEMENTS

This work is funded by Junta de Andalucía (CICE) through project 2006-TIC-2352.

REFERENCES

- Akyildiz, I., Melodia, T., and Chowdhury, K. (2007). A survey on wireless multimedia sensor networks. *Computer Networks*, 51(4):921–960.
- Amiaz, T., Fazekas, S., Chetverikov, D., and Kiryati, N. (2007). Detecting regions of dynamic texture. In *International Conference on Scale Space and Variational Methods in Computer Vision (SSVM'07)*, pages 848–859.
- Carmona, R., Jiménez-Garrido, F., Domínguez-Castro, R., Espejo, S., Roska, T., Rekeczky, C., and Rodríguez-Vázquez, A. (2003). A bio-inspired flexible programmable chip for early vision. *IEEE Transactions on Neural Networks*, 14(5):1313–1336.
- Chen, T., Yin, Y., Huang, S., and Ye, Y. (2006). The smoke detection for early fire-alarming system based on video processing. In *IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'06)*, pages 427–430, California, USA.
- Chetverikov, D. and Péteri, R. (2005). A brief survey of dynamic texture description and recognition. In *International Conference on Computer Recognition Systems (CORES'05)*, pages 17–26, Rydzyna Castle, Poland.
- Fernández-Berni, J. (2008). Video database: Smoke propagation in natural scenarios. <http://www.imse.cnm.es/~berni/>.
- Hu, W., Tan, T., Wang, L., and Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics*, 34(3):334–352.
- Jahne, B., Haußecker, H., and Geißler, P. (1999). *Handbook of Computer Vision and Applications*, volume 2, chapter 4. Academic Press.
- Nelson, R. and Polana, R. (1992). Qualitative recognition of motion using temporal texture. *CVGIP: Image Understanding*, 56(1):78–89.
- Péteri, R., Huskies, M., and Fazekas, S. (2006). Dyn-tex: A comprehensive database of dynamic textures. <http://www.cwi.nl/projects/dyntex/>.
- Pirsch, P. and Stolberg, H. (1998). VLSI implementations of image and video multimedia processing systems. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(7):878–891.
- Roska, B. and Werblin, F. (2001). Vertical interactions across ten parallel, stacked representations in the mammalian retina. *Nature*, 410:583–587.