EARLY SMOKE DETECTION IN FOREST AREAS FROM DCT BASED COMPRESSED VIDEO

A. Benazza-Benyahia[(1)], *N. Hamouda*[(1)], *F. Tlili*[(2)], *S. Ouerghi*[(2)]

(1) Communications, Signal and Image Lab. (COSIM),
(2) Green, Smart Communication Systems Lab. (GRESCOM)
SUP'COM, Carthage University, Tunisia
{benazza.amel,fethi.tlili}@supcom.rnu.tn, {noura.hamouda.12.01,s.ouerghi.safa}@gmail.com

ABSTRACT

Smart functions begin to be integrated into some camera acquisition systems for CCTV applications. In this regard, smoke detection based on the compressed video is a highly desirable functionality for the monitoring of forest that present a high risk of fire. In this paper, we propose a fast and early smoke detection method that measures the local fractal feature of smoke areas based on the Discrete Cosine Transform (DCT) coefficients. The latter are generally computed at the compression step which is concomitant in its acquisition. The two coding video standards MJPEG and MPEG2 are considered as they are widely available in the operating cameras. The novelty of our approach consists in resorting to a recursive DCT in order to improve the detection performance.

Index Terms— smoke detection; compression video standard; Hurst exponent; recursive DCT.

1. INTRODUCTION

Forests and woodlands of the Mediterranean present high risk of fire especially during periods of summer drought. About 50,000 fires destroy each year 600 000 ha [1]. The early detection of fire in the forest areas especially the Mediterranean ones is therefore essential to preserve their related ecosystems and the cultural heritage eventually located in such sites. In this respect, it is used to deploy smoke, temperature and humidity sensors [2]. However, the effectiveness of their detection strongly depends on both their distance to the fire source and the dimensions of the monitored area. In addition, these sensors cannot provide informations about the location of the fire neither about its extent. To alleviate such shortcomings, an easy and non expensive solution is to resort to computer vision techniques [3, 4]. For instance, the project FIRESENSE in which we are partners is committed to develop a CCTV system of fire detection in outdoor archaeological sites [5]. In this context, several methods of automatic smoke detection in the acquired videos have been developed [6, 7]. Generally, they exploit several features of the smoke such as its color, motion and disordered shape [7]. Regarding color, the space representation mainly retained is the RGB space in which the pixels of smoke areas have equal values in the three components [3, 6]. Tests based on the temporal drop of the chrominance value are also employed [6]. Dynamic parameters reflecting the fluctuations of the contours of the smoke can be extracted using the wavelet transform domain [6, 8]. The temporal evolution of smoke disordered areas (size variations, contour irregularity, ...) is also useful for the smoke detection [4, 6]. Recently, a new approach has been proposed based on the fact that smoke smoothes the edges of objects in the scene [9]. Another strategy is to characterize the texture of the smoke areas through the co-occurrence matrix [10] or the Hurst exponent by measuring the roughness of the region of interest [11]. In a recent work [12], we have exploited the fractal nature of the smoke zones highlighted recently in the work of Maruta *et al.* [11]. More precisely, we were interested in estimating the Hurst exponent which is known to reflect the fractal nature of the target areas. The specificity of our approach lies in the fact that the proposed detector only exploits the coding information of the moving sequence during its acquisition. Indeed, we aim at integrating an early detector to the acquisition/compression system of the camera in order to quickly alert the monitoring center. It is worth to note that our strategy does not compete with the conventional pixel domain detection techniques. Indeed, the latter are mainly devoted to run in a control center involving a video server with analytics functionality whereas our approach aims at anticipating the detection at an early stage in order to facilitate the action of the control centre. Generally, the two video compression standards MJPEG and MPEG2 are commonly available in most cameras. They both involve a blockwise Discrete Cosine Transform (DCT). In a recent work [12], we have proposed a novel design of a smart camera functionality which consists in estimating the Hurst exponent just from the DCT coefficients. The objective of this paper is to refine this esti-

This work was supported by FIRESENSE FP7 project http://www.firesense.eu.

mation by increasing the DCT block size while maintaining an acceptable computational load. The rest of this paper is organized as follows. In Section 2, we give an overview of the proposed detector. In Section 3, we present our contribution. In Section 4, we provide some experimental results and, some conclusions are drawn in Section 5.

2. PROPOSED SMOKE DETECTOR

2.1. A bried background on video compression standards

The MJPEG standard performs the compression of a video sequence by encoding frames separately according to the JPEG standard. When applied to disjoint 8×8 blocks, the DCT only exploits the spatial redundancies. Indeed, it is worth recalling that the DCT almost decorrelates the coefficients and concentrates the energy of most natural frames on a small number of low-frequency coefficients. For compression purposes, these properties are very appealing since the high frequency coefficients can be coarsely quantized or even canceled without affecting the reconstruction quality. The MPEG2 video compression standard operates on the video sequence initially divided into Group Of Pictures (GOP) and the coding of a GOP is carried out independently of the remaining ones to achieve some robustness against transmission errors and a quick access to any frame of the sequence. Note that the composition is common to all the GOPs of a sequence and, it is chosen at the beginning of the encoding session. A GOP is composed of three types of images: the intra (I), prediction (P) and bidirectional (B) images. In general, each GOP begins by an I frame coded according to the JPEG encoder. The coding of P and B images is hybrid as it combines a temporal differential prediction of 16×16 macroblocks and the DCT of the error prediction subblocks. Linear quantization of the DCT coefficients followed by a binary encoding are the last operations of the chain compression. In [12], we have proposed a novel smoke detector applied to a video sequence coded either by MJPEG or MPEG2 standards. In what follows, we describe this novel compressed-domain smoke detector.

2.2. Principle of the proposed method

The objective is to discriminate between "smoke" and "non smoke" blocks in a given frame by accounting for the fractal nature (reflected by the Hurst exponent) [14]. Several tests have indicated that an acceptable early alert should be given in a delay not exceeding 5 s. Hence, depending on the temporal sampling rate, it is not required to activate the detector for each acquired image. For instance, for the MJPEG standard, the detection is performed periodically every k_0 image, the value of k_0 being set by the user. In the case of the MPEG2 standard, the detector is applied to the first I image of each GOP. In this case, k_0 is the GOP length. Besides, the smoke is generally dynamic and corresponds to moving areas. Hence, the first step is to discriminate between moving and stationary objects in the target image. As the considered coding standards operate by blocks, moving blocks are firstly extracted an will be considered as candidates containing smoke areas. Next, the smoke blocks are assumed to have a roughness texture characterized by a specific range of values of the Hurst exponent which is different of that of non smoke blocks. Therefore, the challenge consists in estimating accurately the Hurst exponent and, classifying the dynamic blocks according to the resulting estimated fractal parameter.

2.3. Fractal feature estimation

The first step aims at detecting moving blocks in the analyzed image $I(m, n, pk_0)$ that could be potential candidate smoke regions. To this purpose, we have chosen the basic image subtraction technique in order to maintain a low computational complexity. The difference image $J(m, n, pk_0) = I(m, n, pk_0) - I(m, n, pk_0 - k_1)$ is computed with respect to the past image of a delay k_1 whose value is adjusted according to the temporal sampling rate of the sequence. An $N \times N$ block is considered as moving if:

$$\sum_{m,n=1}^{N} \sum_{n=1}^{N} |J(m, n, pk_0)| > S \tag{1}$$

where S is a given threshold set thanks to an off-line learning step based on the image entropy [13]. More precisely, if we assume that an additive Gaussian noise of variance σ^2 corrupts $J(m, n, pk_0)$, S is considered as proportional to the entropic deviation ($S = 4\sigma$) in order to account for 100% of the normal distribution. Hence, it is enough to estimate σ^2 through an estimation of the entropy power.

The second step consists in assigning a binary label c such that $c = 1$ for any moving block b containing smoke. This binary classification is based on the fractal feature (the local Hurst parameter H_b of the underlying block). Let us recall that for a 1D auto-similar signal, the Power Spectral Density (PSD) could be expressed as $1/f^{2H_b+1}$ where f is its frequency and H_b the Hurst parameter [14]. Consequently, H_b could be approximated by the slope of the logarithm of any estimate of the PSD. Note that in order to maintain a very low complexity we have chosen the conventional least square criterion. Regarding our problem, we should first of all compute an estimate of the PSD. A basic one is the periodogram based on the Discrete Fourier Transform coefficients. We propose to calculate the periodogram from the DCT coefficients instead of the conventional discrete Fourier transform coefficients. We move next from the Cartesian coordinates (u, v) in which the DCT coefficients are initially computed to the polar ones (f, θ): $u = f\cos(\theta)$, $v = f\sin(\theta)$. For a given value of θ, the PSD of a self-affine fractal signal is approximated by $1/f^{2h_b(\theta)+1}$ where f is the radial frequency. Therefore, the parameter $h_b(\theta)$ could be estimated according to the least

square criterion and this by considering the estimated PSD logarithm. The estimated Hurst parameter $\widehat{H}_b$ of the candidate block is obtained by averaging $h_b(\theta)$ values when θ describes the finite set Θ of admissible values [15]:

$$\widehat{H}_b = \frac{1}{\text{card}[\Theta]} \sum_{\theta \in \Theta} \hat{h}_b(\theta). \quad (2)$$

At this level, an off-line training procedure has indicated that the distribution of the fractal attributes of each trained class c could be modeled by a Gaussian law defined by the empirical mean $\bar{H}_c$ and the empirical variance σ_c^2. Indeed, we have manually selected blocks containg smoke to compute the histogram of the estimated values of the fractal parameter. The same procedure was also applied to nonsmoke objects. The Gaussianity of the two distributions was checked by displaying Q-Q plots. Finally, given $\widehat{H}_b$, the last step consists in classifying each candidate block b into a smoke class $c = 1$ if

$$\mid \frac{\widehat{H}_b - \bar{H}_1}{\sigma_1} \mid < \mid \frac{\widehat{H}_b - \bar{H}_0}{\sigma_0} \mid . \quad (3)$$

3. IMPROVING THE DETECTION PERFORMANCE

3.1. Motivation

A key issue of our detector is the estimation of the fractal feature. It is important to note that $\widehat{H}_b$ value depends on the employed set Θ [15]. The higher is the block size, the larger is Θ and the more accurate is the estimated $\widehat{H}_b$. To illustrate this point, a synthetic 256×256 fractal image is generated and its Hurst exponent is estimated on different blocks of size N with $N \in \{8, 16, 32, 64, 128, 256\}$. The final estimated Hurst exponent is obtained by averaging the estimations of all the considered blocks. Figure 1 shows the evolution of the quadratic estimation error versus the block size N. As expected, the accuracy increases with the size N. Our motivation is to refine the estimation of the fractal feature by considering larger blocks of coefficients *without* increasing the complexity (number of add and multiply operations). It is mandatory to satisfy the latter constraint to ensure a quick detection by applying a recursive DCT.

3.2. The recursive DCT

Regarding the two coding standards MJPEG et MPEG2, only DCT coefficients of blocks of size 8×8 are available. Our idea is to exploit the recursivity of the DCT according to the block size. More precisely, our contribution consists in considering 4 adjacent DCT 8×8 coefficients available at the encoder to compute DCT coefficients of the counterpart 16×16 macroblock. Indeed, it has been shown that a $N \times N$ DCT coefficient matrix could be easily computed from a $N/2 \times N/2$ one, N being a power of 2 [16]. More precisely, given a block of size $N \times N$, $\mathbf{x}_{N \times N} = [x(m,n)], 0 \leq m, n \leq N-1$, the DCT output coefficients $\mathbf{X}_{N \times N} = [X(u,v)]$, $0 \leq u, v \leq N-1$ is defined by:

$$\mathbf{X}_{N,N} \propto \mathbf{C}_N \mathbf{x}_{N,N} \mathbf{C}_N^T \quad (4)$$

where $\mathbf{C}_N$ is the DCT coefficient matrix of size $N \times N$ defined by its generic element $C_N(u,m)$:

$$\forall u, m = 0, \ldots, N-1, \quad C_N(u,m) = \cos\left(\frac{(2m+1)u\pi}{2N}\right). \quad (5)$$

It is worth recalling the following salient property of matrix $\mathbf{C}_N$ for every $m = 0, \ldots, N-1$:

$$C_N(u,m) = \begin{cases} C_N(u, N-1-m) & \text{if } u \text{ even} \\ -C_N(u, N-1-m) & \text{if } u \text{ odd.} \end{cases} \quad (6)$$

Indeed, this property allows appropriate row and column permutations of $\mathbf{C}_N$ to express $\mathbf{X}_{N \times N}$ as follows:

$$\mathbf{X}_{N \times N} = \mathbf{S}_N \begin{pmatrix} \mathbf{X}_{\text{ee}} & \mathbf{X}_{\text{eo}} \\ \mathbf{X}_{\text{oe}} & \mathbf{X}_{\text{oo}} \end{pmatrix} \mathbf{S}_N^T \quad (7)$$

where $\mathbf{S}_N^T$ denotes the row permutation matrix that rearranges rows of an $N \times N$ matrix to even ones followed by odd ones, both in an ascending order $0, 2, 4, \ldots N-2, 1, 3, 5, \ldots N-3, N-1$ and where the sub-matrices $\mathbf{X}_{\text{ee}}, \mathbf{X}_{\text{eo}}, \mathbf{X}_{\text{oe}}$ and $\mathbf{X}_{\text{oo}}$ are expressed by means of $N/2 \times N/2$ DCT coefficient matrix. In our application, the 8×8 DCT coefficients are already available to enable the computation of the sub-matrices $\mathbf{X}_{\text{ee}}, \mathbf{X}_{\text{eo}}, \mathbf{X}_{\text{oe}}$ and $\mathbf{X}_{\text{oo}}$. Then, from (7), we can obtain the DCT coefficients of blocks 16×16. In [16] the complexity of the recursive DCT cost was evaluated for different sizes N. For $N = 16$, the computational load is of 704 multiplications and 2592 additions while it is of 36,864 multiplications and 19,968 additions for 4 DCTs of 8×8 blocks [16]. If the sequence is of size 720×576 (definition D1) at 30 fps and, if we assume that 10% of the 16×16 blocks are detected as motion ones, the recursive DCT consumes about 16 MIPS. For example, considering the D1-MPEG2 encoder developed by Analog Devices Inc for their Blackfin DSP, the recursive DCT extra computational load remains unsignificantit compared to the 880 MIPS required by the whole encoder [1]. Once the 16×16 DCT coefficients are recursively computed, it only remains to estimate the local fractal parameter from the resulting coefficients.

4. EXPERIMENTAL RESULTS

In the training phase, the 3 training sequences "Forest1"[2], "sWasteBasket" and, "Smoke Manavgat Raw"[3] were employed. It has been found that $(\bar{H}_1, \sigma_1) = (0.4251, 0.0891)$ and, $(\bar{H}_0, \sigma_0) = (0.1412, 0.127)$. Also, after this training

[1] http://www.analog.com/en/processors-dsp/blackfin/BF_MPG2E_00/processors/product.html

[2] www.pond5.com

[3] http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SampleClips.htm

procedure, the threshold value S was set to 2048. The test sequences are the sequences V1, V4, V5 and, V7 retained in the FIRESENSE project. Their main characteristics are listed in Table 1. Only the luminance images are considered. The considered detectors are the basic detector $\mathcal{D}0$ described in Section 2 and the detector $\mathcal{D}1$ when a recursive computation of DCT coefficient blocks of size 16×16 is performed. Note that the set Θ of admissible values has a cardinality of 37 for $\mathcal{D}0$ and, 145 for $\mathcal{D}1$. This clearly indicates a better richness in directionality. A variable number L of images were submitted to the detector. The length k_0 of the GOP concerning the MPEG2 encoder is taken as the corresponding frame rate. Consequently, every 1s, the detector is activated which is a sufficient duration to not miss a possible smoke appearance. The performances of the detector are measured by the rates of false alarm P_f, of missing detection P_m averaged on the considered sequence. For both $\mathcal{D}0$ and $\mathcal{D}1$, there is no shift between the frame where really smoke appears and the frame where the detector decides that a smoke has occured (except for V1, the shift is equal to 1). This clearly corroborates the detection earliness. Tables 2 and 3 provide the detection performances obtained with MJPEG and MPEG2: $\mathcal{D}1$ improves dramatically $\mathcal{D}0$ one. The next round of experiments consists in comparing $\mathcal{D}1$ with two fractal based smoke detectors denoted by $\mathcal{D}2$ and $\mathcal{D}3$. The only difference between $\mathcal{D}1$ ad $\mathcal{D}2$ is that the fractal parameter estimation is performed in the initial spatial domain as described in [17] whereas $\mathcal{D}3$ corresponds to the method od Maruta *et al.* with its own setup [11]. Tables 4 and 5 clearly indicate the benefit drawn by Hurst exponent estimation in the compressed domain (detector $\mathcal{D}0$) and by resorting to larger DCT blocks (detector $\mathcal{D}1$) w.r.t. to competitive fractal based detectors $\mathcal{D}2$ and $\mathcal{D}3$. Figure 2 shows an example of the detection results of the $\mathcal{D}1$ method for respectively V1 and V5 frames.

5. CONCLUSION

In this paper, we have proposed a new method of smoke detection from DCT based compressed video which exploits the coding information of the moving sequence during its acquisition. Our main contribution consists not only in the estimation of the Hurst exponent just from the DCT coefficients but also in the refinement of this estimation by accounting for the block size while maintaining an acceptable computational load. This was performed by use of a recursive DCT. Experimental results on real sequences are satisfactory and encourages to integrate this early detection in the acquisition/compression system of the camera. These satisfactory experimental results on real sequences encourage to integrate this early detector in the acquisition/compression system of the camera. Further works should also be investigated to exploit the motion and the chrominance information.

6. REFERENCES

[1] H. N. Le Houérou, "Vegetation wildfires in the mediterranean basin: evolution and trends," *Ecologia Mediterranea*, vol. 13, pp. 13-24, 1987.

[2] W. Jones, "An algorithm for fast and reliable fire detection," *8th Fire Suppression and Detection Research Application Symp.*, January 2004, Orlando.

[3] T. H. Chen, P. H. Wu and Y. C. Chiou, "An early fire-detection method based on image processing," *IEEE Internat. Conf. on Image Processing*, Singapore, pp. 1707-1710, Oct 2004.

[4] Z. Xiong, R. Caballero, H. Wang, A. Finn, M. A. Lelic and P. Peng, "Video-based smoke detection possibilities, techniques, and challenges," *Suppression and Detection Technical Working Conf.*, Orlando, FL, USA, March 2007.

[5] N. Grammaladis *et al.*, "A multisensor network for the protection of culturel heritage," *European Signal and Image Processing Conf.* Barcelona, Spain, 2011.

[6] B. U. Toreyin, Y. Dedeoglu, U. Gudukbay, A. E. Çetin, "Computer vision based method for real time fire and flame detection," *Pattern Recognition Letter*, no. 27, pp. 49-58, 2006.

[7] S. Verstockt, P. Lambert, R. Van de Walle, B. Merci and B. Sette, "State of the art in vision-based fire and smoke dectection," *Internat. Conf. on Automatic Fire Detection*, 2009.

[8] A. Rafiee, R. Tavakoli, R. Dianat, S. Abbaspour and M. Jamshidi, "Fire and smoke detection using wavelet analysis and disorder characteristics,, *Internat. Conf. on Computer Research and Development*, Shanghai, China, pp. 262-265, March 2011.

[9] R. Gonzalez-Gonzalez *et al.*," Wavelet-based smoke detection in outdoor video sequences," *IEEE Midwest Symp. on Circuits and Systems*, Seattle, Washington, USA, pp. 383-387, August 2010.

[10] H. Maruta, A. Nakamura and F. Kurokawa, "A novel smoke detection method using support vector machine," *IEEE TENCON*, pp. 210-215, Nov. 2010.

[11] H. Maruta, A. Nakamura, T. Yamamichi and F. Kurokawa, "Image based smoke detection with local Hurst exponent," *IEEE TENCON*, pp.4653-4656, Nov.2010.

[12] N. Hamouda, A. Benazza-Benyahia, F. Tlili, "Détection de fumée dans des zones de forêt à partir de séquences vidéo codées, *GRETSI'11*, Bordeaux, France, Sept. 2011 (in French).

[13] F. Luthon, M. Liévin, F. Faux, "On the use of entropy power for threshold selection," *Signal Processing*, vol. 84, pp.1789-1804, 2004.

[14] B. Mandelbrot, "Selfaffine fractals and fractal dimension," *Physica Script*, vol. 32, pp. 257260, 1985.

[15] M. Petrou, P. Garcia-Sevilla. *Dealing with texture*, Wiley, Chichester, West Sussex, England, 2006.

[16] P. Z. Lee and F.-Y. Huang, Restructured recursive DCT and DST algorithms, *IEEE Transactions on Signal Processing*, vol. 42, no. 7, pp. 16001609, July 1994.

[17] S. Hoefer, F. Heil, M. Pandit, R. Kumaresan, "Segmentation of textures with different roughness using the model of isotropic two-dimensional fractional Brownian motion," *IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, pp.53-56,1993.

Table 1. Characteristics of training and test sequences.

Sequence	Frame rate (fps)	Image Size
Smoke Manavgat Raw	25 images/s	288×352
forest1	29 images/s	256×480
sWasteBasket	10 images/s	576×720
V1	7 images/s	720×576
V4	25 images/s	352×288
V5	9 images/s	352×288
V7	5 images/s	352×288

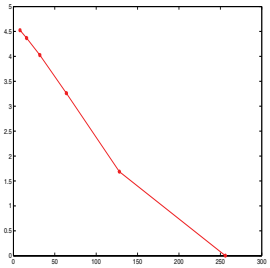


Fig. 1. Quadratic estimation error *versus* the block size for a synthetic fractal image.

Table 2. Performances of the $\mathcal{D}0$ and $\mathcal{D}1$ detectors operating with k_1 = 30 over L frames coded with the MJPEG standard.

Video	P_f		P_m	
	$\mathcal{D}0$	$\mathcal{D}1$	$\mathcal{D}0$	$\mathcal{D}1$
V1, L=350	0.4042	**0.2501**	0.0678	**0.0305**
V4, L=1,000	0.4604	**0.1536**	0.0042	**0.0018**
V5, L=950	0.1345	**0.0314**	0.1385	**0.1335**
V7, L=2,470	0.2607	**0.1187**	0.1450	**0.0630**
Average	0.2720	**0.1103**	0.0986	**0.0595**

Table 3. Performances of the $\mathcal{D}0$ and $\mathcal{D}1$ detectors operating with k_1 = 30 over L frames coded with the MPEG2 standard.

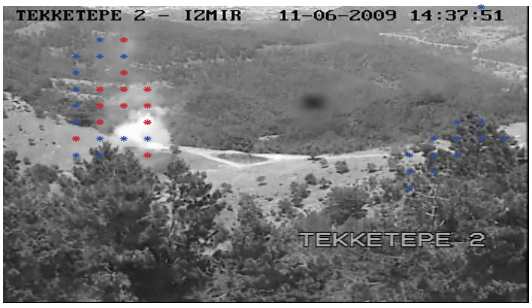
Video	P_f		P_m	
	$\mathcal{D}0$	$\mathcal{D}1$	$\mathcal{D}0$	$\mathcal{D}1$
V1, L=50	0.02	**0.018**	0.1459	**0.0563**
V4, L=40	0.3657	**0.1615**	0.05	**0.018**
V5, L=106	0.1358	**0.0311**	0.1569	**0.1335**
V7, L=494	0.2519	**0.0756**	0.1485	**0.1777**
Average	0.1881	**0.0675**	0.1224	**0.0809**

Table 4. Performances of detector $\mathcal{D}2$ operating with k_1 = 30 over L frames coded with the MJPEG (a) and MPEG2 (b) standards. The employed values of L are the same as in Table 2.

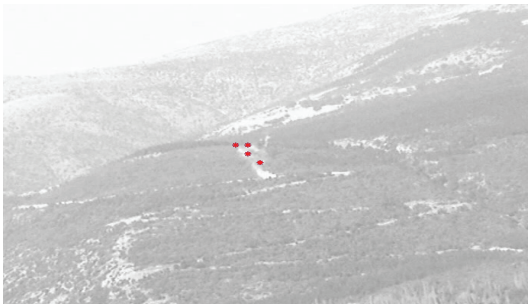
Video	P_f		P_m	
	(a)	(b)	(a)	(b)
V1	0.5577	0.021	0.0392	0.0986
V4	0.2086	0.2038	0.018	0.019
V5	0.0448	0.0456	0.1935	0.1923
V7	0.1202	0.1155	0.2318	0.259
Average	0.1775	0.0937	0.1039	0.1180

Table 5. Performances of detector $\mathcal{D}3$ operating with k_1 = 30 over L frames.

Video	P_f	P_m
V1, L=350	0.0219	0.0077
V4, L=1,000	0.0178	0.0193
V5, L=950	0.0054	0.3372
V7, L=2,470	0.1453	0.0652
Average	0.0474	0.1073



(a) Frame # 128 of V1



(b) Frame # 222 of V5

Fig. 2. MJPEG coded frame blocks classified on $c = 0$ (in blue) or $c = 1$ (in red) with $(k_0, k_1) = (1, 30)$.