Change Detection for Video

Alexia Briassouli, Ioannis Kompatsiaris
Informatics and Telematics Institute, Centre for Research and Technology Hellas
Thermi-Thessaloniki, 57001, Greece.
Email: abria@iti.gr,ikom@iti.gr

Abstract

The increased presence of digital multimedia in numerous applications, such as security, surveillance, the semantic web, has rendered the automated characterization of video content necessary. The localization of different activities/events in video content is of particular interest, however it is quite challenging to achieve in a principled manner and with no prior knowledge or training. This work presents an original, principled solution to the problem of detecting changes of activity in video, based on sequential change detection techniques. Initially, a binary mask of the active pixels, the Activity Area, is extracted in a pre-processing step by estimating the kurtosis values of inter-frame illumination variations. Sequential change detection is then applied to the illumination changes of the active pixels over time, leading to the separation of the video sequence into segments corresponding to different activities, which can be further processed for classification and recognition. Experiments with various indoors and outdoors videos demonstrate the system's good performance.

I. Introduction

Automated systems for event detection in video are constantly being developed and refined, in an effort to increase the usability of the video data. Event detection requires the initial segmentation of the video in sequences of frames that contain one kind of event or activity. Currently most work on the temporal processing of video separates sequences in shots, i.e. groups of frames filmed from the same camera or viewpoint. However, the resulting video segments do not necessarily contain different types of activities, since no motion, and hence activity information, is taken into account. Events and activities are often detected using Hidden Markov Models (HMMs) [1], but these approaches require significant amounts of training, and their architecture needs to be selected empirically in order to achieve accurate results. This fact, in combination with their high computational cost, renders them highly specialized, limiting their flexibility for use in a wide range of applications.

Motion is a significant indicator of changes in events and actions in video, so this work proposes a novel, motion-based non-parametric approach to the problem of activity detection. The times and locations of changes in activity are detected via statistical processing of the data. An important advantage of the methods used is that they are not tailored to specific data sets, nor do they depend on training a particular model, so the proposed system can be applied to a wide range of videos. Additionally, the processing can take place in real time, apart from a pre-processing step that can be implemented a priori.

II. Motion Analysis: Activity Area

The first step of the proposed system, which can take place offline, processes the entire video sequence at once, in order to extract a binary mask of the active pixels, called the Activity Area. An underlying assumption is that the camera and background are static, or that possible camera motion can be compensated for. It should be noted that the derivation of the Activity Area is robust to small background motions, as also shown in the experimental results. At each frame k, the inter-frame illumination variations at pixel $\bar{r}$ can be induced by displacement $u_k(\bar{r})$ in active pixels, or only measurement noise $z_k(\bar{r})$ in static ones. This can be expressed as the following two hypotheses:

$$\begin{aligned} H_0 : v_k^0(\bar{r}) &= z_k(\bar{r}) \\ H_1 : v_k^1(\bar{r}) &= u_k(\bar{r}) + z_k(\bar{r}), \end{aligned} \tag{1}$$

where $v_k^i(\bar{r}),\ i = 0, 1$ represents the measured illumination variations at frame k, pixel $\bar{r}$. Hypothesis H_0 corresponds to the case where only noise is present, represented by $z_k(\bar{r})$, and H_1 corresponds to the case where there is actual displacement, given by $u_k(\bar{r})$. The illumination variations between successive frames can be estimated from simple inter-frame differences or from optical flow estimates, for videos with noisier data. The distribution of the measurement noise $z_k(\bar{r})$ is not known, but can be approximated by a Gaussian probability density function (pdf), as is often the case in the literature [2]. This approximation can be further supported by the fact that the measurement noise originates from a large number of sources and is additive, so it can be approximated by a Gaussian pdf based on the Central Limit Theorem [3]. Then, active and static pixels can be separated by finding those whose illumination variations do not follow a Gaussian distribution. A classic measure of Gaussianity is the kurtosis $kurt(y) = E[y^4] - 3(E[y^2])^2$, as the kurtosis of Gaussian data is equal to zero [4]. The kurtosis is actually sensitive to outliers, so it can detect which $v_k^i(\bar{r})$ originate from pixel activity, even when the measurement noise is not strictly Gaussian. In practice, the illumination variations of each frame pixel over time are extracted, and their kurtosis is estimated. Its absolute value is then compared with a threshold that is equal to 10% of the mean absolute kurtosis value. This threshold is found to provide a robust separation of the high and low kurtosis values, and consequently the active and static pixels. Extensive experiments show that for a wide range of both indoors and outdoors videos, the kurtosis indeed leads to accurate Activity Areas at a low computational cost.

III. Sequential Change Detection

The Activity Area is extracted offline, as its estimation requires use of the entire video sequence. Once it is available, the video can be processed in real-time to find changes in the activity in it. The time instants where activities change are extracted by sequential change detection techniques [5]. The input is, as before, a sequence of the illumination variations from frame k_0 to k, i.e. $v_{k_0,k} = [\bar{v}_{k_0}, \bar{v}_{k_0+1}, ..., \bar{v}_k]$, which follow a distribution f_0 before a change occurs, and f_1 after the change, at an unknown frame k_{ch}. Here, $\bar{v}_{ki}$ contains the illumination variations of all pixels in the activity area at frame ki, so an activity area consisting of N_a pixels has $\bar{v}_{ki} = [v_{ki}(1), ..., v_{ki}(Na)]$. All active pixels from each frame are examined simultaneously to ensure that there will be enough data samples for the sequential change detection to provide reliable results. At frame k, $v_{k_0,k}$ is input into the log-likelihood ratio to detect a change:

$$T_k = LLRT_k(f_1||f_0) = \ln \frac{f_1(v_{k_0,k})}{f_0(v_{k_0,k})} = \ln \prod_{ki=k_0}^{k} \frac{f_1(\bar{v}_{ki})}{f_0(\bar{v}_{ki})} = \sum_{ki=k_0}^{k} \ln \frac{f_1(\bar{v}_{ki})}{f_0(\bar{v}_{ki})}, \tag{2}$$

where it has been assumed that the data samples $\bar{v}_{ki}$ are independent identically distributed (i.i.d.) under each hypothesis, so $f_H(v_{k_0,k}) = \prod_{ki=k_0}^{k} f_H(\bar{v}_{ki}),\ H = 0, 1$. The values $v_{ki}(1), ..., v_{ki}(Na)$ are also assumed to be i.i.d., so $f_H(\bar{v}_{ki}) = \prod_{n=1}^{Na} f_H(v_{ki}(n))$. The log-likelihood ratio is then equal to:

$$T_k = LLRT_k(f_1||f_0) = \sum_{ki=k_0}^{k} \sum_{n=1}^{N_a} \ln \frac{f_1(v_{ki}(n)}{f_0(v_{ki}(n))}, \tag{3}$$

The test statistic is expressed with the following iterative form [6], known as the CUSUM (Cumulative Sum) test:

$$T_k = \max\left(0, T_{k-1} + \ln \frac{f_1(\bar{v}_k)}{f_0(\bar{v}_k)}\right), \tag{4}$$

and a change is detected at a frame k_{ch} when the test statistic becomes higher than a pre-defined threshold. Unlike the threshold for sequential probability likelihood ratio testing [7], [8], the threshold for the CUSUM testing procedure cannot be determined in a closed form manner. It has been proven in [9] that the optimal threshold for the CUSUM test, for a pre-defined false alarm γ, is the threshold that leads to an average number of changes under H_0 equal to γ. In the general case examined here, the optimal threshold is estimated empirically from the data being analyzed [10]. After extensive experimentation, it is found that, for videos like the ones examined here, the optimal threshold is equal to $\eta_{opt} = \mu_T + \cdot\sigma_T$, where μ_T and σ_T are the mean and standard deviation of the test statistic T_k until frame k.

A. Data Modeling

The CUSUM test of Eq. (4) requires knowledge about the family of distributions before and after the moment of change in order to be implemented, although the time of change is unknown. The illumination variations of active pixels over time contain outliers introduced by their change in motion, as these pixels are usually not active in all frames. Data which contains outliers follows more heavy-tailed distributions than the Gaussian, such as the Laplacian or generalized Gaussian [11]. In this work we focus on the Laplacian, whose parameters can be estimated at a lower computational cost than the generalized Gaussian. The Laplacian distribution is given by:

$$f(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right), \tag{5}$$

where μ is the data mean and $b = \sigma/\sqrt{2}$ is its scale, for variance σ^2. The exponent of this distribution contains an absolute difference instead of the difference squared, so its tails decay more slowly than those of the Gaussian, indicating that the data contains more outliers than Gaussian data. The test statistic of Eq. (3) is written as:

$$T_k = \sum_{ki=k_0}^{k} \sum_{n=1}^{N_a} \ln\left[\frac{b_0}{b_1} \exp\left(-\frac{|v_{ki}(n)-\mu_1|}{b_1} + \frac{|v_{ki}(n)-\mu_0|}{b_0}\right)\right] = NN_a \ln \frac{b_0}{b_1} + \sum_{ki=k_0}^{k} \sum_{n=1}^{N_a} \left(-\frac{|v_{ki}(n)-\mu_1|}{b_1} + \frac{|v_{ki}(n)-\mu_0|}{b_0}\right) \tag{6}$$

The data used in the experiments, namely the illumination variations, are modeled with the Gaussian and Laplacian distributions in order to determine which one is more appropriate. The Root Mean Square error (RMS) between the actual empirical data distribution and the corresponding Gaussian and Laplacian model is estimated for all videos. For reasons of space, only its mean is presented, which is found to be 0.0915 for the Gaussian and 0.0270 for the Laplacian model, justifying the choice of the latter as a better fit for our data. In practice, the time of change is not known, so the data to be used for the estimation of the model parameters needs to be selected carefully. In the experiments, the first $w_0 = 10$ frames are considered as "baseline data", which is used to derive the parameters of f_0. At each frame k, the data in that frame and the previous $w_1 = 10$ frames is used to approximate f_1. This entails the necessary assumptions that no changes occur in the first $10 - 20$ frames that contain the baseline data and the data used for the first approximations of f_1. These windows are found to give good change detection results for a wide range of videos, both indoors and outdoors, from different domains.

IV. Experiments

The proposed approach is applied to a wide range of videos, both indoors and outdoors, from different domains, in order to determine its performance in practice. The videos can be found on http://mklab.iti.gr/content/videos.

A. Basketball hoop

Initially a video of a kid throwing a ball through a basketball hoop is examined (Fig. 1(a), (b)). The corresponding activity area, shown in Fig. 1(c), shows the pixels where the kid moves and the trajectory of the ball. The application of the CUSUM algorithm leads to the detection of changes at frames $20, 35$. Indeed, at frame 20 the kid starts to throw the ball, which moves up until frame 35, after which it falls through the hoop.

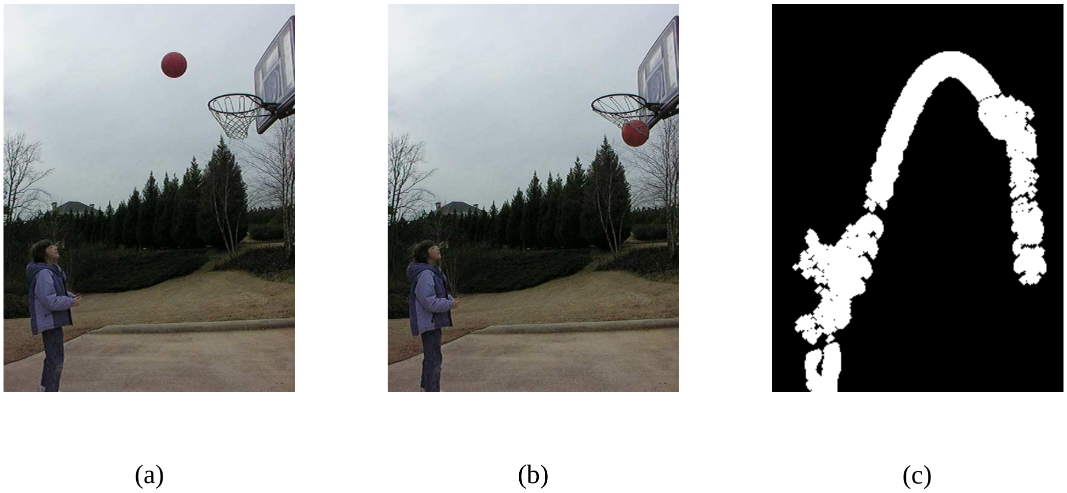
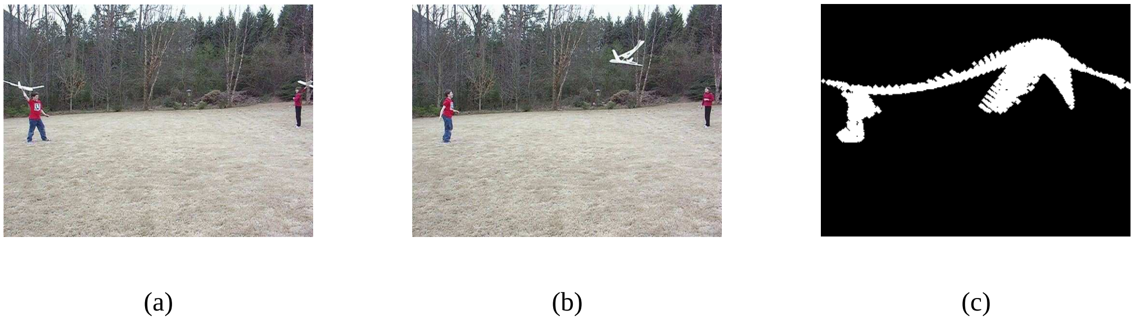

(a) (b) (c)

Fig. 1. Basketball hoop. (a) Frame 20. (b) Frame 35. (c) Activity Area.

B. *Kids flying planes*

In this experiment, a video of two kids flying planes is examined (Fig. 2(a), (b)). The corresponding activity area, shown in Fig. 2(c), shows the pixels where the kids moved and the plane trajectories. Although the background illumination in the trees area is not entirely constant, the activity area is extracted with good accuracy because of the sensitivity of the kurtosis to outliers, which correspond to the actual childrens' motions and not small background motions and illumination variations. The application of the CUSUM algorithm leads to the detection of changes at frames $25, 60$. Indeed, at frame 25 the kids start to move their hands to throw the planes, and the airplanes' motion does change at frame 60 when they crash into each other and fall, verifying the results of the CUSUM method.


(a)

(b)

(c)

Fig. 2. Kids flying planes (a) Frame 15. (b) Frame 60. (c) Activity Area.

C. *Fight, run away*

An indoors surveillance video is examined, where two people walk towards each other, start fighting, and then run away from each other (Fig. 3(a), (b)). The resulting activity area contains two disjoint regions, one corresponding to the people walking and fighting, and on corresponding to minor activity near the receptionist's desk (Fig. 3(c)). The small activity area contains too few pixels to reliably detect changes, so only the data in the large activity area is processed. Changes are detected at frames $45, 57, 90, 117, 141, 172, 225$. The change in frame 45 is an error, as the people approach each other during frames $1 - 57$. At frame 57 one person opens his arms, at frame 90 they meet and start to fight, so those are true change points. During the fight they move towards the right at frame 117 and then to the left at frame 141. At frame 172 they stop fighting and run away from each other, disappearing from the scene at frame 225. Thus, the change detection results agree with the ground truth for this more complex activity. The two false alarms can be eliminated by estimating the mean displacement in the subsequences before and after those instants: if it is the same, those change points are removed.

V. Conclusions

A novel method for the analysis of video based on statistical sequential techniques is presented. Binary masks indicating the regions of activity in a video are extracted from the higher order statistics of the illumination variations,

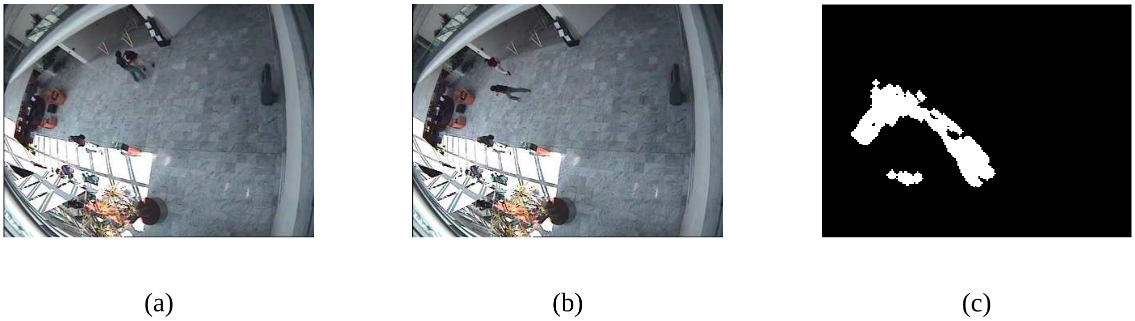


Fig. 3. Fight, run away. (a) Frame 100. (b) Frame 160. (c) Activity Area.

so that only the active pixels inside these masks are processed in the sequel. Changes in the activity taking place are detected via the application of sequential change detection techniques, namely the CUSUM algorithm. The proposed system does not require training or model selection in order to be implemented, therefore it can be applied to a variety of video data. Experiments with various both indoors and outdoors videos verify that the method gives good results. Future work involves the extension of the proposed approaches for the removal of false alarms in various contexts.

ACKNOWLEDGMENTS:

The research leading to these results has received funding from the European Communitys Seventh Framework Programme FP7/2007-2013 under grant agreement FP7-214306 - JUMAS, from FP6 under contract number 045547-VidiVideo.

REFERENCES

[1] J. Huang, Z. Liu, and Y. Wang, "Joint scene classification and segmentation based on hidden markov model," *IEEE Transactions on Multimedia*, vol. 7, no. 3, pp. 538 – 550, 2005.

[2] T. Aach, A. Kaup, and R. Mester, "Statistical model-based change detection in moving video," *Signal Process.*, vol. 31, no. 2, pp. 165–180, 1993.

[3] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 2nd edition, 1987.

[4] G.B. Giannakis and M. K. Tsatsanis, "Time-domain tests for Gaussianity and time-reversibility," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3460–3472, Dec. 1994.

[5] I.V. Nikiforov, "A generalized change detection problem," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 171 – 187, Jan. 1995.

[6] E. S. Page, "Continuous inspection scheme," *Biometrika*, vol. 41, no. 1, pp. 100–115, June 1954.

[7] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 2nd edition, 1994.

[8] A. Wald, *Sequential Analysis*, Dover Publications, 2004.

[9] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, 1986.

[10] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1993.

[11] L. Alparone B. Aiazzi and S. Baronti, "Estimation based on entropy matching for generalized gaussian pdf modeling," *IEEE Transactions on Signal Processing*, vol. 6, no. 6, pp. 138–140, 1999.