International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 2 Feb 2021, pp: 879-883 www.ijaem.net ISSN: 2395-5252

Extraction of Semantic Dynamic Content from Videos

Dr. Khalid Ahmed Ibrahim[1], Prof.G.K.Viju[2]

1. Associate Professor, Faculty of CS & IT, Karary University, Omdurman, Sudan
2. Professor and Dean (E-Learning), University of Garden City, Khartoum, Sudan

Submitted: 15-02-2021 Revised: 02-03-2021 Accepted: 05-03-2021

ABSTRACT: The exploitation of video data requires to extract information at a rather semantic level, and then, methods able to infer "concepts" from low-level video features. This paper adopts a statistical approach and focus on motion information. Because of the diversity of dynamic video content, appropriate motion models are designed and learn them from videos. This paper defines original and parsimonious probabilistic motion models, both for the dominant image motion and the residual image motion. Motion measurements include affine motion models to capture the camera motion, and local motion features for scene motion. The two-step event detection scheme consists in pre-selecting the video segments of potential interest, and then in recognizing the specified events among the pre-selected segments, the recognition being stated as a classification problem.

I. INTRODUCTION:

A video retrieval system is a computer system for browsing, searching and retrieving video images from a large database of digital images. It can be based on a textual description of the images in the database.

Exploiting the tremendous amount of multimedia data, and specifically video data, requires to develop methods able to extract information at a rather semantic level. The characteristics of a semantic event have to be expressed in terms of video primitives (color, texture, motion, shape ...) sufficiently discriminant with respect to content. This remains an open problem at the source of active research activities.

In this paper, the problem of inferring concepts from low-level video features is tackled and a statistical approach involving modeling, (supervised) learning and classification issues is followed. This paper deals with concepts related to events in videos, more precisely, to dynamic content. Therefore, motion information was focused. To this end, a new probabilistic motion model was introduced. Such a probabilistic modelling allows us to derive a parsimonious motion representation while coping with errors in the motion measurements and with variability in motion appearence for a given type of event. In a distinct way the scene motion (i.e., the residual image motion) and the camera motion (i.e., the dominant image motion) is handled, since these two sources of motion bring important and complementary information. As for motion measurements, consider, on one hand, parametric motion models to capture the camera motion, and on the other hand, local motion features to account for the scene motion.

A two-step event detection method is designed to restrict the recognition issue to a limited and pertinent set of classes since probabilistic motion models have to be learnt for each class of event to be recognized. This allows us to simplify the learning stage, to save computation time and to make the overall detection more robust and efficient. The first step consists in selecting candidate segments of potential interest in the processed video. Typically, for sport videos, it involves to select the "play" segments. The second step handles the recognition of the relevant events (in terms of dynamic content) among the segments selected after the first step and is stated as a classification problem.

II. MATERIALS AND METHODOLOGY

Motion Measurements:

It is possible to characterize the image motion, by computing at each pixel a local weighted mean of the normal flow magnitude. However, the image motion is actually the sum of two motion sources: the dominant motion and the residual motion. More information can be recovered explicitly while considering these two motion components rather than the total motion only. Thus, first compute the camera motion

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 2 Feb 2021, pp: 879-883 www.ijaem.net ISSN: 2395-5252

between successive images of the sequence. Then, cancel the camera motion allows to compute local motion related measurements revealing the residual image motion only. The dominant image motion is represented by a deterministic 2D affine motion model

$$\mathbf{w}_\theta(p) = \begin{pmatrix} a_1 + a_2 x + a_3 y \\ a_4 + a_5 x + a_6 y \end{pmatrix}$$

Where θ= (a_i, i=1,…., 6) is the model parameter vector and p=(x, y) is an image point.

The residual motion measurement $v_{res}(p,t)$ is defined as the local mean of the magnitude of normal residual flows weighted by the square of the norm of the spatial intensity gradient. The normal residual flow magnitude is given by the absolute value of the Displaced Frame Difference DFDθt, evaluated with the estimated dominant motion, and divided by the norm of the image spatial gradient. The final answer is

$$v_{res}(p,t) = \frac{\sum_{q \in \mathcal{F}(p)} \|\nabla I(q,t)\| . |DFD_{\hat{\theta}_t}(q)|}{\max\left(\eta^2, \sum_{q \in \mathcal{F}(q)} \|\nabla I(q,t)\|^2\right)},$$

$DFD_{\theta t}(q)=I(q+w_{\theta t}(q),t+1)I(q,t).F(p)$ is a local spatial window centered in pixel p.ΔI(q,t) is the spatial intensity gradient of pixel q at time t.η^2 is a predetermined constant related to the noise level.

Probabilistic Modelling of Motion:

The proposed method for the detection of important dynamic events relies on the probabilistic modelling of the motion content in a video. Indeed, the large diversity of video contents leads to favor a probabilistic approach which moreover allows to formulate the problem of event recognition within a Bayesian framework. Due to the different, nature of the information brought by the residual motion (scene motion) and by the dominant motion (camera motion), two different probabilistic models are defined.

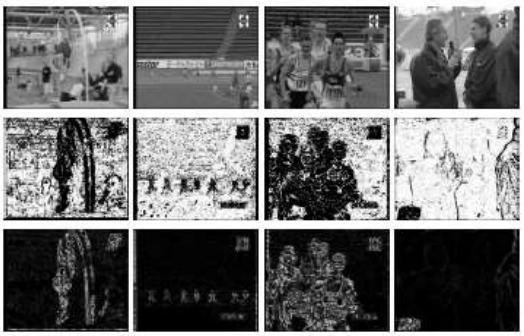


Fig. 1. Athletics video: First row: four images of the video. Second row: the corresponding maps of dominant image motion supports (inliers in white, outliers in black). Third row: local residual motion measurements v_{res} (zero-value in black).

Residual Motion:

The probabilistic model of scene motion derived from statistics on the local residual motion measurements. The histograms of these measurements computed over different video segments were found to be similar to a zero-mean Gaussian distribution except a usually prominent peak at zero. Therefore, the distribution of the local residual motion measurements within a video segment by a specific mixture model involving a truncated Gaussian distribution and a Dirac distribution is modeled. It can be written as:

$$F_{vres}(\gamma)=\beta\delta_0(\gamma)+(1-\beta)\phi_t(\gamma;0,\sigma^2)\, I_{\gamma\neq 0}(\gamma)$$

where β is the mixture weight,δ_0 denotes the Dirac function at 0 ($\delta_0(\gamma)=1$ if $\gamma = 0$ and $\delta_0(\gamma)= 0$ otherwise) and $\phi_t(\gamma;0,\sigma^2)$ denotes the truncated Gaussian density function with mean 0 and variance σ^2 . I denotes the indicator function ($I_{\gamma\neq 0}$ = 1 if $\gamma\neq 0$ and $I_{\gamma\neq 0} = 0$ otherwise). Parameters β and σ^2 are estimated using the Maximum Likelihood criterion.

The full probabilistic residual motion model is then defined as the product of these two models as follows: $P_{Mres}(v_{res},\Delta v_{res}) = P(v_{res}).P(\Delta v_{res})$ The probabilistic residual motion model is completely specified by four parameters only which are moreover easily computable.

Dominant Image Motion:

Design a probabilistic model of the camera motion to combine it with the probabilistic model of the residual motion in the recognition process. A first choice could be to characterize the camera motion by the motion parameter vector θ and to represent its distribution over the video segment by a probabilistic model. Mathematical representation of the estimated motion models was

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 2 Feb 2021, pp: 879-883 www.ijaem.net ISSN: 2395-5252

proposed, that is the camera-motion flow vectors and to consider the 2D histogram of these vectors. Finally, this histogram is represented by a mixture model of 2D Gaussian distributions. The number of components of the mixture is determined with the Integrated Completed Likelihood criterion (ICL) and the mixture model parameters are estimated using the Expectation-Maximisation (EM) algorithm.

Event Detection Algorithm:

Now exploit the designed probabilistic models of motion content for the task of event detection in video. Concepts of dynamic content to be involved in the event detection task is learned. The videos to be processed are segmented into homogeneous temporal units. To segment the video, a shot change detection technique or a motion based temporal segmentation method can be used. Let $\{s_i\}_{i=1,\cdots,N}$ be the partition of the processed video into homogeneous temporal segments.

Selecting Video Segments :

The first step of our event detection method permits to sort the video segments in two groups, the first group contains the segments likely to contain the relevant events, and the second one is formed by the video segments to be definitively discarded. Typically, consider sport videos, try to first distinguish between "play" and "no play" segments. This step is based only on the residual motion which accounts for the scene motion; therefore only single-variable probabilistic models are used, which saves computation. Denote $\{M^{1,n}{}_{res} 1 \leq n \leq N_{1,}\}$ as residual motion models learnt for the "play" group and $\{ M^{2,n}{}_{res} 1 \leq n \leq N_2\}$ as residual motion models learnt for the "no play" group. Then, the sorting consists in assigning the label ζ_i, whose value can be 1 for "play" or 2 for "no play", to each segment s_i of the processed video using the ML criterion defined as follows:

ζ_i, $=\arg\max_{k=1,2} (\max_{1=n=Nk} P^{k,n}{}_{Mres}(z_i))$

$z_i = \{(v_{res\ i}, \Delta v_{res\ i}\ \}$ denote the local residual motion measurements and their temporal contrasts for the video segment s_i.

Detecting Relevant Events :

The method deals with the detection of the events of interest within the previously selected segments .The detection is performed in two sub-steps.

Video segment labeling
Event label validation
Video segment labeling :

For each video segment s_i, $z_i = \{v_{res\ i}, .v_{res\ i}\}$ are the residual motion measurements and their temporal contrasts, and wi represent the motion vectors corresponding to the 2D affine motion models estimated between successive images over the video segment si.

Event label validation:

However, consider that there might be "no play" segments wrongly labeled as "play" after the first selection step. These segments are called as "intruders" These segments are forced to be assigned one of the event classes using relation

$L_i = \arg\max_{j=1,\ldots.J} P_{M}{}^{j}{}_{res}(zi) \times P_{M}{}^{j}{}_{cam}(w_i)$

which creates false detection.

III. HOW IT WORKS?

The described method was applied on sports videos which involve complex contents while being easily specified. Moreover, events or highlights can be naturally related to motion information in that context.The results obtained on athletics and tennis videos are reported here.

Experimental Comparison: First, an experimental comparison between our statistical approach and a histogram-based technique were carried out. In order to evaluate the probabilistic framework which have designed, consider the same motion measurements for the histogram technique. Thus, the latter involves three histograms: the histogram of residual motion measurements v_{res} (2), the histogram of their temporal contrasts Δv_{res}, and the 2D histogram of the camera-motion flow vectors (subsection 3.2). Each event j is then represented by three histograms:

$H^j{}_{vres}$, $H^j{}_{\Delta\ vres}$ and $H^j{}_{cam}$,

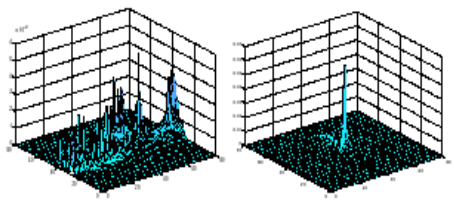
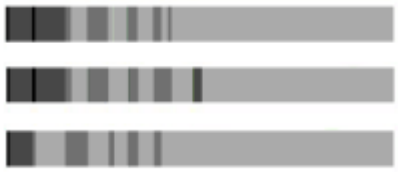


Fig. 2. Athletics video: 2D histograms of the camera-motion flow vectors. Left: for a pole vault shot, right: for a long-shot of track race.



International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 2 Feb 2021, pp: 879-883 www.ijaem.net ISSN: 2395-5252

Fig. 3. Athletics video: Detection of relevant events: Top row: ground-truth, middle row: results obtained with the probabilistic motion models, bottom row: results obtained with the histogram-based technique.

To compare two histograms, consider the Euclidian distance, denoted by d_1 for 1D histograms and by d_2 for 2D histograms. However, the computed motion measurements are all real values and have a huge number of available computed values. Thus consider a very fine quantization and the resulting histograms are very close to the real continuous distributions. Moreover, the histogram distance is only used to rank the classes. The Euclidean distance is then a reasonable choice while easy to compute. These histograms are computed (and stored) for each event j from the training set of video samples. Then, consider the test set
and compute the three histograms:
H^{si}_{vres} , $H^{si}_{\Delta vres}$ and H^{si}_{cam} , for each video segment s_i of the test set. The classification step is now formulated as assigning the label l_i of the event which minimizes the sum of the distances between histograms:

$$l_i = \arg\min_{j=1,\ldots,J} \left(d_1(H^{si}_{vres}, H^{j}_{vres}) + d_1(H^{si}_{\Delta vres}, H^{j}_{\Delta vres}) + d_2(H^{si}_{cam}, H^{j}_{cam}) \right)$$

In order to focus on the classification performance of the two methods, the test set only involves "play" segments. A part of an athletics TV program which includes jump events and track race shots are processed. The training set is formed by 12500 images and the test set comprises 7800 images. Four events are recognized: Pole vault, Replay of pole vault, Long-shots of track race and Close-up of track race.

Event Detection Method :

Event detection method is applied to a tennis TV program. The first 42 minutes (63000 images) of the video are used as the training set (22 minutes for the learning of the motion models involved in the two first steps and 20 minutes for the learning of intruder models), and the last 15 minutes (18000 images) form the test set.

Selecting video segments: Distinguish between "play" segments involving the two tennis players in action and the "no play" segments including views of the audience, referee shots or shots of the players resting, as illustrated in Figure 4. Exploit only the first subset of the training set to learn the residual motion models that need for the selection step. 205 video segments of the training set represent "play" segments and 95 are "no play" segments. 31 residual motion clusters and their associated models are supplied by the AHC algorithm for the "play" group, and 9 for the "no play" group. The high number of clusters obtained reveals the diversity of dynamic contents in the two groups of the processed video. Quite satisfactory results are obtained, since the precision rate for the play group is 0.88 and the recall rate is 0.89.

Fig. 4. Tennis video: Three image samples extracted from the group of "play" segments and three image samples extracted from the group of "no play" segments.

	Rally	Serve	Change of side
P	0.92	0.63	0.85
R	0.89	0.77	0.74

Table 1. Tennis video: Results of the event detection method based on probabilistic motion models (P: precision, R: recall).

Detecting relevant events:

The goal is now to detect the relevant events of the tennis video among the segments selected as "play" segments. For this second step, introduce the probabilistic camera motion model. The three events we try to detect are the following: Rally, Serve and Change of side. In practice, consider two sub-classes for the Serve class, which are wide-shot of serve and close-up of serve. Two sub-classes are considered too for the Changeof-side class. As a consequence, five residual motion models and five camera motion models have to be learnt. Also determine the residual motion models accounting for the intruder segments for each class. The obtained results are reported in Table 1. Satisfactory results are obtained specially for the rally class. The precision of the serve class is lower than the others. In fact, for the serve class, errors come from the selection step (i.e., some serve segments are wrongly put in the "no play" group, and then, are lost). It appears that a few serve

segments are difficult to distinguish from some "no play" segments when using only motion information.

IV. COMPARISON WITH EXISTING SYSTEMS :

In general retrieval system have the following limitation There are various methods to extract texture features from temporal slices for motion retrieval. Since tensor histogram features offer the best performance, further combine tensor histograms and color histograms for clustering and retrieving of video shots. The validity of the proposed approaches have been confirmed by extensive and rigorous experimentations in sport video domain. To apply the proposed methods to other video sources such as movie and documentary films, the current works need to be further pursued in two directions: motion segmentation and the integration of various video features. Therefore, the decomposition of camera and object motion is a preferable step prior to feature extraction.

In Extraction of Semantic Dynamic Content from Videos the proposed motion and color features can be easily incorporated with other features such as audio and textual information for a more sophisticated video retrieval system.

V. CONCLUSION:

In this paper it have been addressed the issue of determining dynamic content concepts from lowlevel video features with the view to detecting meaningful events in video. Mainly focused on motion information and designed an original and efficient statistical method. New probabilistic motion models representing the scene motion and the camera motion was introduced. They can be easily computed from the image sequence and can handle a large variety of dynamic video contents. The considered statistical framework outperforms a histogram-based technique was demonstrated. Moreover, it is flexible enough to properly introduce prior on the classes if available, or to incorporate other kinds of video primitives (such as color or audio). The proposed two-step method for event detection is general and does not exploit very specific knowledge (e.g. related to the type of sport) and dedicated solutions.

REFERENCES:

[1]. A. Ekin, A.M. Tekalp, and R. Mehrotra. Automatic soccer video analysis and summarization. IEEE Int. Trans. on Image Processing, 12(7):796–807, July 2013.

[2]. J. Li and J.Z. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Trans. on PAMI, 25(9):1075–1088, Sept. 2013.

[3]. R. Fablet,P. Bouthemy, and P. Perez. Non-parametric motion characterization using causal probabilistic models for video indexing and retrieval. IEEE Trans. On Image Processing, 11(4):393–407, 2012.

[4]. C-W. Ngo, T-C. Pong, and H-J. Zhang. On clustering and retrieval of video shots through temporal slices analysis. IEEE Trans. Multimedia, 4(4):446–458, Dec.2012.

[5]. L. Zelnik-Manor and M. Irani. Event-based video analysis. IEEE Int. Conf. on Computer Vision and Pattern Recognition, Kauai, Hawaii, Dec. 2001.

[6]. Y. Rui and P. Anandan. Segmenting visual actions based on spatio-temporal motion patterns. CVPR'2000, Hilton Head, SC, 2000.

[7]. N. Vasconcelos and A. Lippman. Statistical models of video structure for content analysis and characterization. IEEE Trans. on IP, 9(1):3–19, Jan. 2000.

[8]. C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the Integrated Completed Likelihood. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(3):719–725, 2000.

International Journal of Advances in Engineering and Management
ISSN: 2395-5252

IJAEM

Volume: 03 Issue: 02 DOI: 10.35629/5252

www.ijaem.net
Email id: ijaem.paper@gmail.com