International Journal of Engineering Research ISSN:2319-6890)(online),2347-5013(print)
Volume No.3, Issue No.4, pp : 258-260 01 April 2014

Motion Based Summarization and Grouping of Events for Video Surveillance System

L. Arun Raj, C. Thangapandian

Department of Computer Science, BS Abdur Rahman University, Chennai, India
arunraj@bsauniv.ac.in, mailmeat89@gmail.com

Abstract— Motion Based Summarization and Grouping of Events for Video Surveillance is one of the approach for detecting dynamic and complex scenes in computer vision. It aims to automatically recognize and track people and objects from image sequences in order to understand and describe dynamics and interactions among them. Furthermore we propose a grouping of events people running together, fighting, etc. This method can handle both symmetric and asymmetric group activities. Video based summarization have the potential to assist in maintaining public safety and security.

Keywords **: Video Surveillance, Video Summarization, Grouping of Events, Motion based summarization.**

INTRODUCTION

Video Summarization and Grouping of Events for Video Surveillance of dynamic and complex scenes is one of the most active research topics in computer vision. It aims to automatically detect, recognize and track people and objects from image sequences in order to understand and describe dynamics and interactions among them [1]. Event classification [2] and grouping event is one key task involved in it. Being automatically able to detect group activities of humans is very important for public safety. We use group activity detection algorithm which can handle both symmetric and asymmetric group activities. In early the most of the researchers used Hidden Markov Model (HMM) [11] for group event detection. However it detects group activities with fixed number of group members and they cannot handle flexible and varying number of group members, where the input feature vector length is fixed. So the input feature vector length should vary with respect to group of activities. Video based surveillance has the potential to assist in maintaining public safety and security.

Virtually all public spaces and critical infrastructures in the world have a multiple sensor surveillance system installed, many of which demands to have automatic surveillance features. Typical application domains for video surveillance include public areas (city streets, school campuses, and museums), transport (airports, train stations, underground, motorways) and retail (theft prevention, understanding hopper behavior).

Figure. 1. Security issues that can be managed by video surveillance system.

Figure. 1. shows the list of security issues that can be managed by video surveillance system. Traditional video surveillance system has two drawbacks. They are finding available human resources to observe the output and manual system are ineffective when the number of cameras exceeds the ability of human operators to keep track of the evolving scene. This drawback can be overcome in automatic video surveillance system which can work 24 hours a day, 7 days a week allowing for accurate event detection and their cost is lower than maintaining a group of operators. The motion based summarization and grouping of events for video surveillance system made of six modules.

ARCHITECTURE

The motion based summarization and grouping of events in video surveillance involves serious of steps. They are Video Segmentation, Background Subtraction, Object Extraction, Object Detection, Event Classification and Grouping of events. The architecture of motion based summarization is shown in Figure. 2.

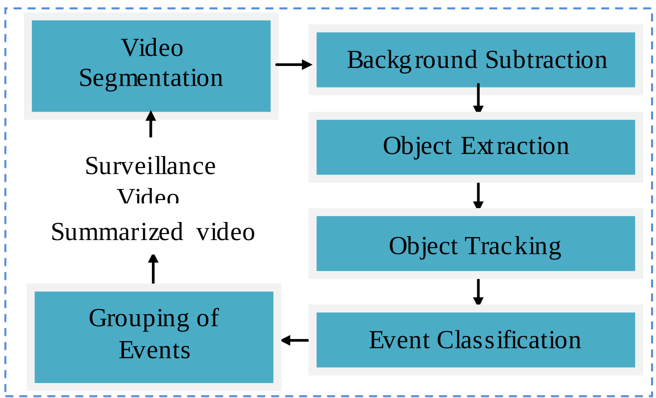

Figure. 1. Architecture of Motion based summarization and grouping of events for video surveillance system.

The components of motion based summarization are explained below.

1. VIDEO SEGMENTATION

Its goal is to divide the video stream into a set of meaningful and manageable segments (shots) [3] [6] that are used as basic elements for indexing. Each shot is then represented by selecting key frames and indexed by extracting spatial and temporal features. The retrieval is based on the similarity between the feature vector of the query and already stored video features.

2. BACKGROUND SUBTRACTION

Background subtraction (BS) [7] is a widely used segmentation technique able to achieve real-time performance. BS aims to segment moving regions in image sequences comparing current frame to a model of the scene background. A pixel is classified as being from a moving object if the difference between the current frame and the background model is above a given threshold. Background subtraction methods can be organized in:

1) per pixel,
2) per region and
3) per frame.

A per-pixel approach is formed by methods that consider each pixel signal as an independent process. Region-based algorithms usually divide the frames into blocks and calculate block-specific features in order to obtain the foreground.

Frame-level class is formed by methods that look for global changes in the scene. Usually, they are used jointly with other pixel or region background subtraction approaches.

3. OBJECT EXTRACTION

It extracting foreground objects from color images and videos with very little user interaction. Object extraction [8] is a critical task in video summarization. It extracting foreground objects from color images and videos with very little user interaction. This task is usually accomplished by chroma keying, where principal subjects are first captured against a background consisting of a single color and then the object will be extracted. The major two steps in object extraction are

- Background subtraction
- Foreground extraction

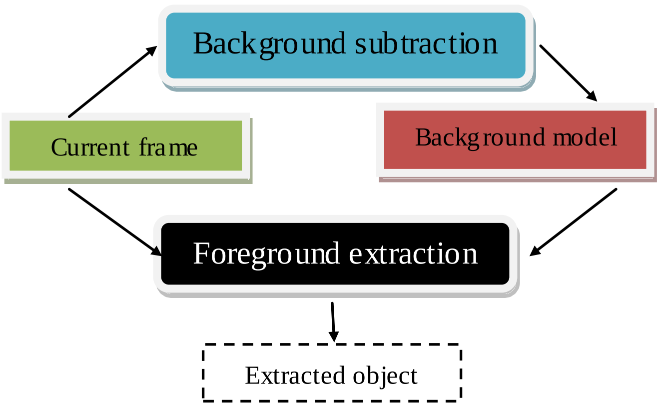

Figure. 3. Object Extraction

4. OBJECT TRACKING

The aim of an object tracker is to generate the trajectory of an object over time by locating its position in every frame of the video [8]. Object tracker may also provide the complete region in the image that is occupied by the object at every time instant. The tasks of detecting the object and establishing correspondence between the object instances across frames can either be performed separately or jointly. Two major components can be distinguished in object tracking.

- Target representation and localization is mostly a bottom-up process which has also to cope with the changes in the appearance of the target.

- Filtering and data association is mostly a top-down process dealing with the dynamics of the tracked object, learning of scene priors, and evaluation of different hypotheses.

The way the two components are combined and weighted is application dependent and plays a decisive role in the robustness and efficiency of the tracker. The object tracking deals with track initialization, track update (including prediction and data association), track deletion.

5. EVENT CLASSIFICATION

Video event classification has video classification has the intent of classifying an entire video, some authors have focused on classifying segments of video such as identifying violent or any unwanted action took place or distinguishing between different news segments within an entire surveillance video.

5. GROUPING EVENT

Video event classification has video classification has the intent of classifying an entire video, some authors have focused on classifying segments of video such as identifying violent or any unwanted action took place or distinguishing between different news segments within an entire surveillance video. In this paper, we address the following issues for group event detection.

i. Group Event Detection with a Varying Number of Group Members.

ii. Group Event Detection with a Hierarchical Activity Structure.

Table I : List of Group Activities

Activity	**Definition**
In Group	The people are in group.
Walk Together	People walking together.
Fight	Two or more group fighting.
Run Together	The group running.
Ignore	Ignoring each other.
Approach	Two people or Group with one approaching the other.
Chase	One group chasing another.

Some example of group activities are shown below in Figure. 3. Group events address the problem of detecting events with varying number of people.

(a)

(b)

(c)

Figure. 3. Some example frames of group activities.
(a) People in Groups.
(b) People walk together.
(c) People Fighting.

In this paper we use Asynchronous Hidden markov model (AHMM) [10] to model activity correlation between two people. AHMM was introduced to handle asynchronous feature. Since the feature streams of different people in the same group may not be perfectly synchronized (e.g., when two people walk together, one person may stretch the leg earlier than the other person), AHMM can help reduce the possible recognition errors from these action frames.

EXPERIMENTAL EVALUATION

we recorded 8 hours of training practice in two different days and with different light conditions. For testing the accuracy of our system we visually examined a set of randomly chosen frames taken from different moments in the day and compared for each frame how many people are actually in the camera field of view (FOV) and how many centroids are located by the segmentation algorithm.

The error $\mathbf{e_i}$ for each scene $\mathbf{i}$ is computed as

$$\mathbf{e_i} = \frac{|n' - n|}{n}$$

where n' is the number of detected people and n is the real number of people in the FOV. The accuracy a_i for each scene i is

$$\mathbf{a_i = 1 - e_i}$$

The average accuracy A is $\mathbf{A} = \frac{1}{n}\sum_{i=1}^{n} a_i$

The result is showed in Table, where different type of situations is considered depending on the number of people in the FOV. A comparison with other similar methods is not easy because those consider quite often up to 3 or 4 people in the scene, while we examined more crowded situations.

Table II : Summarization Accuracy

No. of people in the scene	No. of samples considered	Accuracy in %
0-4	25	99.00 %
5-9	25	97.00%
10-14	25	95.00%
15-19	25	92.00%

CONCLUSION

In this system, we have presented our approaches for the exploitation of motion based summarization and grouping of events in video surveillance. Our video summarization method effectively retrieves key frames from the perspective of human perception. And then classify the events into groups (walking together, fighting, people in groups etc).

REFERENCE

i. *Feng Wang, Chong-Wah Ngo, "Summarizing Rushes Videos by Motion, Object, and Event Understanding" Volume: 14 , Publication Year: 2012.*

ii. *Hanjalic. A, Yan. S, LIU. Q, Smeaton, A. F. , "Special Section on Object and Event Classification in Large-Scale Video Collections" Volume: 14, Publication Year: 2012.*

iii. *Marcin Detyniecki, Christophe Marsala. "Adaptive acceleration and shot stacking for video rushes summarization" , Volume: 7, Publication Year: 2008.*

iv. *Thanikachalam. V, Thyagharajan. K.K, "Human action recognition using accumulated motion and gradient of motion from video" in Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference, Publication Year: 2012.*

v. *Wenshuo Gao, Xiaoguang Zhang, Lei Yang, Huizhong Liu "An improved Sobel edge detection" in Computer Science and Information Technology (ICCSIT), Volume: 5; Publication Year: 2010.*

vi. *Abdelati Malek Amel, Ben Abdelali Abdessalem, Mtibaa Abdellatif, "Video shot boundary detection using motion activity descriptor", Volume 2, Publication Year: 2010.*

vii. *Yiran She, Wen Hu, Junbin Liu, Mingrui Yang, Bo Wei and Chun Tung Chou;. "Efficient Background Subtraction for Real-time Tracking in Embedded Camera Networks" Publication Year: 2012.*

viii. *Chih-Yuan Chung and Homer H. Chen, "Video Object Extraction via MRF-Based Contour Tracking" Volume: 20; Publication Year: 2010.*

ix. *Panagiotis Sidiropoulos, Vasileios Mezaris, Ioannis Kompatsiaris, Hugo Meinedo, Miguel Bugalho, and Isabel Trancoso, "Temporal Video Segmentation to Scenes Using High-Level Audiovisual Features" Volume: 21; Publication Year: 2011.*

x. *D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan, "Modeling individual and group actions in meetings with layered HMMs," IEEE Trans. Multimedia, vol. 8, no. 3, pp. 509–520, Jun. 2006.*

xi. *D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan, "Modeling individual and group actions in meetings with layered HMMs," IEEE Trans. Multimedia, vol. 8, no. 3, pp. 509–520, Jun. 2006.*