Real-Time Video Segmentation based on Motion Detection

Sharannya Sudarshan, Farnoush Banaei-Kashani, Cyrus Shahabi
Integrated Media System Center, USC
iCampus/MediaQ iWatch ✓ CT

Introduction

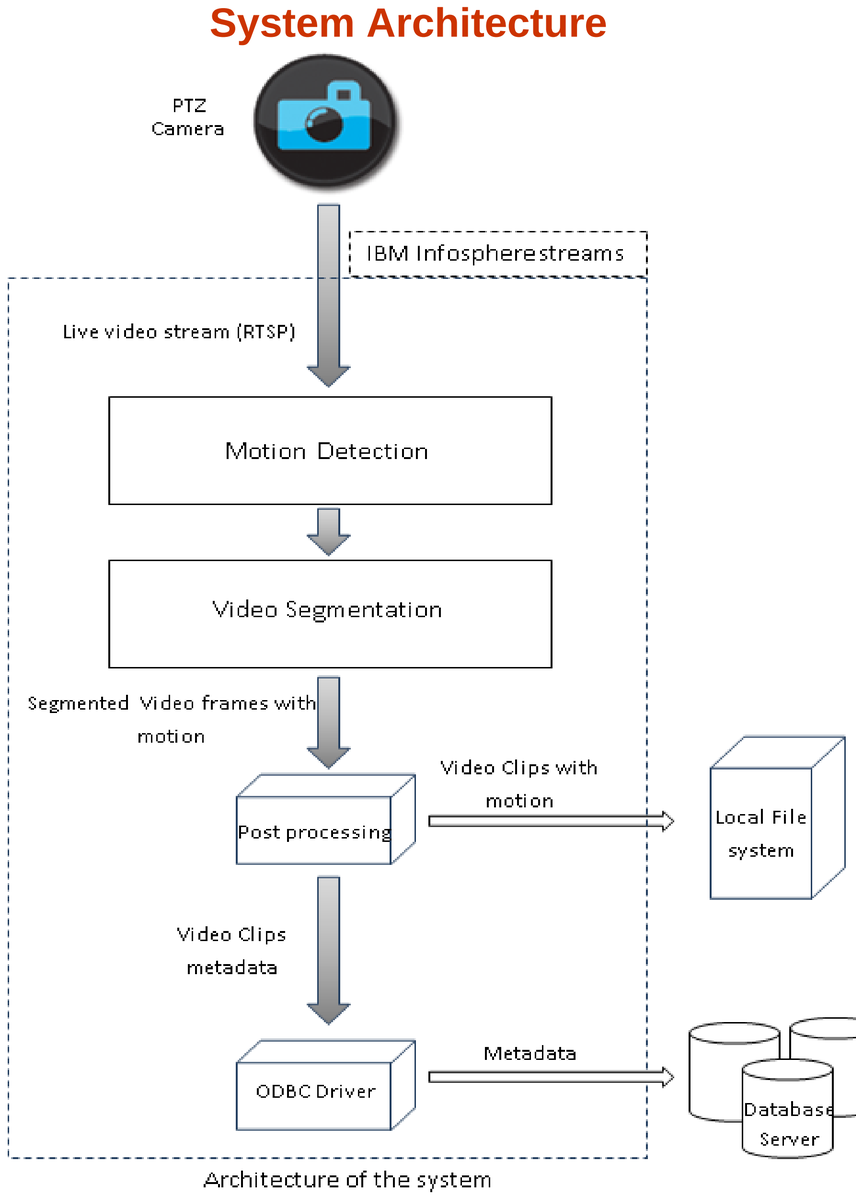
Identifying moving objects from a video sequence is a fundamental and critical task in many computer vision applications. Video segmentation is one such application which has been studied for several decades, and still remains a difficult problem for computer to automatically and accurately segment moving objects from video sequence with various backgrounds and global motions in real time. In this study, an IBM Infosphere Stream based approach for real time video segmentation has been implemented. Shi and Tomasi algorithm for feature detection and Pyramid Lucas-Kanade algorithm for motion vector detection are used to accurately segment video frames with moving objects from the live video sequence.

Motivation

Video segmentation on live data stream has been extensively used in video surveillance systems. Video surveillance systems monitor the surroundings and collect a large volume of audio/video information requiring great computational analysis. The system is expected to detect the object and monitor the respective motions and subsequently analyze the object behavior in order to prevent any untoward incidents. The first step in analyzing the behavior of an object in real time is object detection and tracking. Tracking involves detection of regions of interest in a frame and then finding frame-to-frame correspondence of each region's location and shape. However, current motion segmentation methods are mainly based on background subtraction, temporal differencing, optical flow or statistical methods.

The key problem in video segmentation is how to fully exploit spatial and temporal information provided by video sequence, especially temporal information over multiple frames. Some of the recent methods segment moving objects precisely using temporal information over multiple frames, but unfortunately they have high computation cost. For that reason in this study we use pyramidal implementation of the Lucas-Kanade method which has proved to be a very fast and a reliable optical flow estimator algorithm.

System Architecture

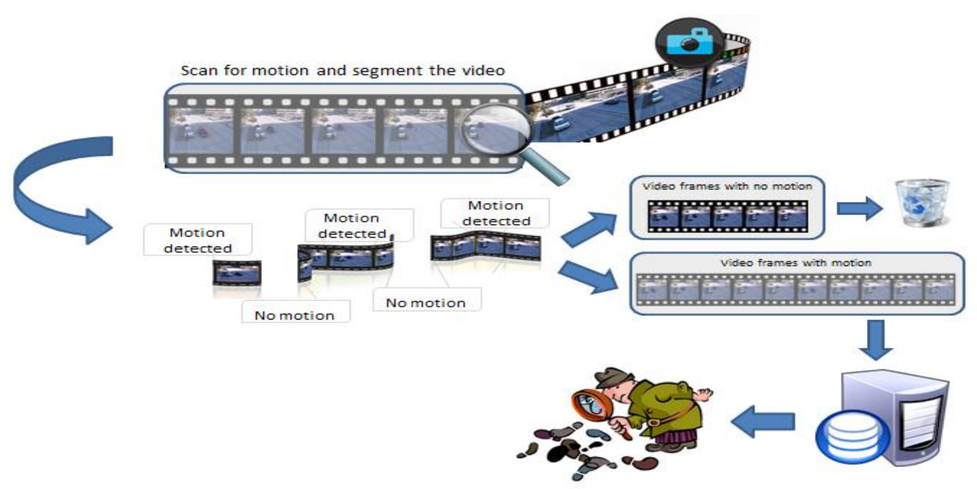


Architecture of the system

The main component of the system is the IBM InfoSphere Streams processing unit. The live video stream is captured from the PTZ cameras using RTSP and then is processed using the user-defined functions which apply Computer Vision algorithms to identify segments of the video which contain motion. These frames are grouped together to form time-stamped video files which maybe be examined later to identify any untoward incidents.

Capability

1. Segmentation is based on motion information from multiple frames unlike the previous methods where the segmentation is based on successive frames only.
2. Real time Segmentation
   - -Real-time monitoring;
   - -Real-time motion detection and moving object segmentation.
3. Intelligent frame processing
   - -Fully exploits temporal information over multiple frames;
   - -Handles low frame rate of input data;
   - -Differentiates between motion of significant features and insignificant.
4. IBM streams for stream processing
   - -Computationally efficient;
   - -Motion Detection and Segmentation handled by OpenCV toolkit provided for IBM InfoSphere Streams. Database handled via IBM InfoSphere Streams Database toolkit;
   - -With IBM Streams's local memory, it allows the processing algorithm to look back in time for processing of older frames. Current implementation looks back in history and also looks ahead to identify the valid candidates of video frames containing motion components.



Simplified view of the system

Related Research

Many techniques and algorithms have been proposed for moving object segmentation and these are generally divided into two categories. The first category of techniques usually segment moving objects on every key frame and use them as initialization values, and then track these in the following frames [1, 2, 3]. The main drawback in this category is that it is difficult to automatically achieve the initial segmentation due to the lack of temporal information. Another category of techniques utilize motion information between two successive frames, and then segment each frame of video sequence the same way [4, 5]. The main drawback is that the performance may be unstable when objects either move too fast or pause between two successive frames.

Conclusion and Future Work

The system efficiently handles motion detection and segmentation of live video stream. The segmented videos are stored in the database server and undergo further processing to extract significant information which along with other supplementary information can be used to identify any untoward incidents.

The current implementation handles input live video stream from only one PTZ camera at a time. Development work is currently in progress towards building a system which handles multiple inputs from a series of PTZ cameras simultaneously and processes them in real-time. The resultant set of segmented videos will be stored in the DB server as was done in the previous implementation, and further forensic analysis can be performed on these stored videos.

References

[1] D. Wang, "Unsupervised video segmentation based on watersheds and temporal tracking", IEEE Trans. Circuits Syst. Video Technol., vol. 8, pp. 539-546, Sept. 1998.

[2] T. Meier, K. N. Ngan, "Video segmentation for content-based coding", IEEE Trans. Circuits Syst. Video Technol., vol. 9, pp. 1190-1203, Dec. 1999.

[3] L. Wu, J. B. Pineau, Ph. Delagnes, and D. Barba, "Spatial-temporal segmentation of image sequences for object-oriented low bit-rate image coding", Signal Processing: Image Commun., vol. 8, pp. 513-543, 1996.

[4] J.G. Choi, S-W. Lee, and S-D, Kim, "Video segmentation based on spatial and temporal information", IEEE Int. Conf. Acoust., Speech, Signal Processing, ICASSP'97, pp. 2661-2664, Munich, Germany, April 1997.

[5] R. Mech, M. Wollborn, "A noise robust method for segmentation of moving objects in video sequence", IEEE Int. Conf. Acoust., Speech, Signal Processing, ICASSP'97, pp. 2657-2660, Munich, Germany, April 1997.