Track-Based Video Compression

M. J. Carlotto[1], J. G. Ackenhusen, B. R. Suresh
General Dynamics Advanced Information Systems

An alternative to conventional techniques for compressing video data of moving objects is described. The method, known as track-based compression (TBC), detects, associates, and tracks moving objects between frames, sending only a small chip or ID around the moving object once the track has been established. The compression ratio achievable depends on scene content, sensor geometry, the degree to which the background can be stabilized, and other factors. Preliminary results range from 1,500:1 for oblique sensing geometries with significant parallax to more than 10,000:1 for near-nadir overhead and fixed ground-based surveillance video.

Key words: Video compression, tracking, background estimation, change detection, track-based filtering

Introduction

Given a limited bandwidth channel, as the space-time bandwidth of a video sensor increases the data stream must be compressed to a greater degree in order to prevent loss of information. For example, a gigapixel video sensor operating at 10 Hz requires ~ 50,000:1 compression ratio to be sent over a 10 Mb/sec downlink. The current state of the art in low-loss video compression (H.264), which can compress HDTV 1080i video (~ 2 Mpix/frame) to a rate of 5.5Mp/sec (2 x 24 x 30/5.5 = 261:1), is more than two orders of magnitude away, and is usually not realizable in real time.

Track-based compression (TBC) offers an alternative to conventional techniques for sending large volumes of video-derived information over limited bandwidth channels. Commercial video codecs attempt to optimize performance globally, i.e., over the full frame. An alternative approach is to concentrate only on the part of the video stream that is of interest to the user. TBC achieves extremely high compression ratios (> 10,000:1) in two ways: First, it only sends information over regions in the video frame that change. This in itself, depending on the density of moving objects and other changes, can achieve compression ratios between 100:1 and 1,000:1. Additional gain is achieved by associating and tracking changes between frames, sending only a small chip or ID around the moving object once a track has been established.

The processing chain is summarized in Figure 1. Registration includes both frame to ground and frame to frame (frame stabilization) processing. Moving object detection involves background modeling, change detection, and feature extraction. Detected changes are associated and tracked across frames. Track-based filtering eliminates many of the false alarms caused by glint, parallax, and sensor artifacts. The output data stream consists of a file generated at the frame rate containing the location and ID of all moving objects in track at that time. Chips are generated

[1] Point of contact: mark.carlotto@gd-ais.com

once a track has been confirmed and are sent lossless. After an initial training period, a background image is computed, which can be downlinked at a lower bit rate.

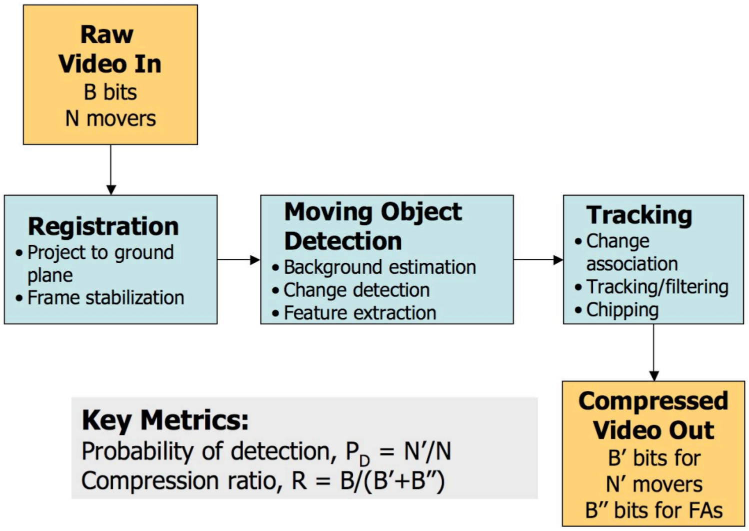


Figure 1 Track-based video compression processing

Registration Considerations

Achieving high compression ratios using TBC requires good tracking performance; i.e., long tracks with few false alarms. Key change detection performance metrics include the probability of detection (P_D) and false alarm rate (FAR). Our goal is to maximize the compression ratio while maintaining a high probability of detection, $P_D > 95\%$. Accurate registration and stabilization of the background is key to reducing false alarms and achieving high compression ratios, and can be particularly challenging in overhead images taken from a moving platform. For first-order Markov textured surfaces, the processing gain (SNR) of change detection over single image object detection (thresholding), is

$$\gamma = 2 \big/ 1 - E\left[a^{|d|}\right]$$

where a is the Markov coefficient and d is the local displacement resulting from parallax, both random variables. At a given P_D, the FAR depends on the processing gain. To achieve a 1000:1 compression ratio at the detection level (i.e., before track processing), the FAR must be less than 10^{-3}. Assuming Gaussian statistics, an SNR of 13.5 db is required to achieve a 95% P_D at that FAR. If we assume the target to clutter ratio is 3 db, the CD processing gain must be about 10 db. Using an analytic model for predicting CD performance (Carlotto 2007), the standard deviation of the displacement (misregistration) error must be less than about 0.65 pixels rms. The ability to accurately register the background is thus a critical driving requirement, followed by the need to effectively mitigate false alarms caused by 3-D parallax changes.

2-D Background Estimation and Change Detection

A 2-D background modeling approach is used for detecting changes across a registered sequence of video frames. The simplest technique (temporal anomaly detection) models the background brightness at each pixel location by its mean and variance. Moving objects are detected if the brightness exceeds a constant false alarm rate (CFAR) threshold

$$\left|x(t)-\mu(t)\right|/\sigma(t) > T$$

where

$$\mu(t) = (1-\alpha)\mu(t-1) + \alpha x(t-1)$$

$$\sigma^2(t) = (1-\alpha)\sigma^2(t-1) + \alpha\left[x(t-1)-\mu(t-1)\right]^2$$

and α is the learning rate. A Gaussian mixture model (Stauffer and Grimson 2000) provides better performance (at a higher computational complexity) when the background statistics change in a simple fashion, e.g., due to a blinking light. However, mixture models are not much better than simple anomaly detection in dealing with most kinds of moving clutter. Kernel density models (Elgammal et al 2002) are somewhat better, provided background changes occur less frequently than changes caused by moving objects.

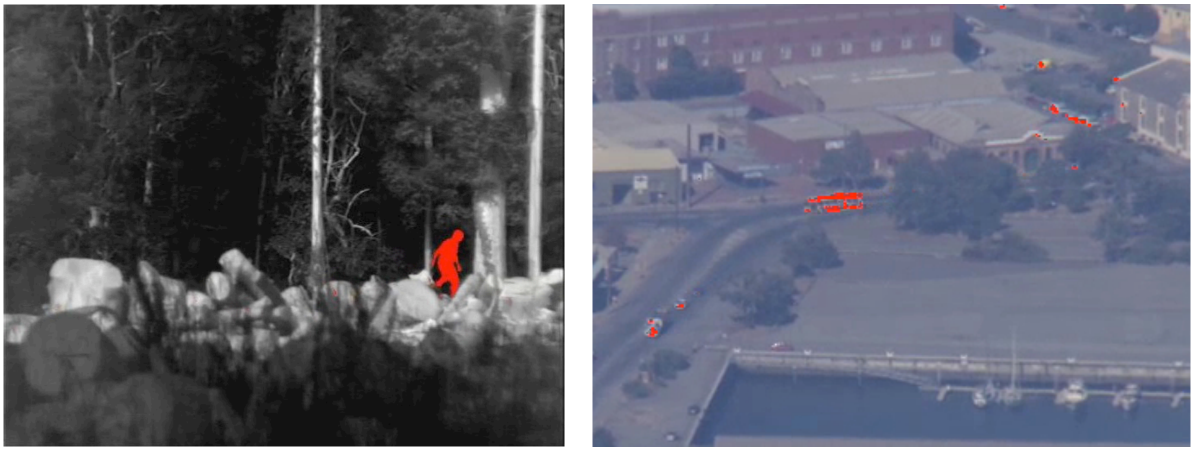


Figure 2 Moving object detection. Ground-based thermal IR video (left) and aerial video (right). (IR video courtesy NVL.)

Examples of moving object detection in two different types of imagery are shown in Figure 2. In the thermal IR video (640 x 480) at 10 frame/sec about 0.3% of a frame changes on average. Sending only chips and reports for moving objects at least 100 pixels in area (a person) results in a compression ratio (CR) of 355:1. By sending only the report location and size, the CR increases to more than 150,000:1. For a high definition (1280 x 720) aerial camera at a frame rate of 1 frame/sec, sending chips and reports for moving objects greater than 25 pixels in area results in a CR of 1470:1. The CR for sending only the report location and size is 4,400:1. Sending reports and chips, the CR is higher for the aerial video because the changes (chips) are smaller; sending only reports, the CR for the ground video is higher because there are fewer moving objects.

Following change detection, 2-D feature extraction estimates the physical length, width, and pose of the object, as well as other features which can be used as features for tracking and to determine the class of the object by its size (e.g., dismount, car, truck, etc.).

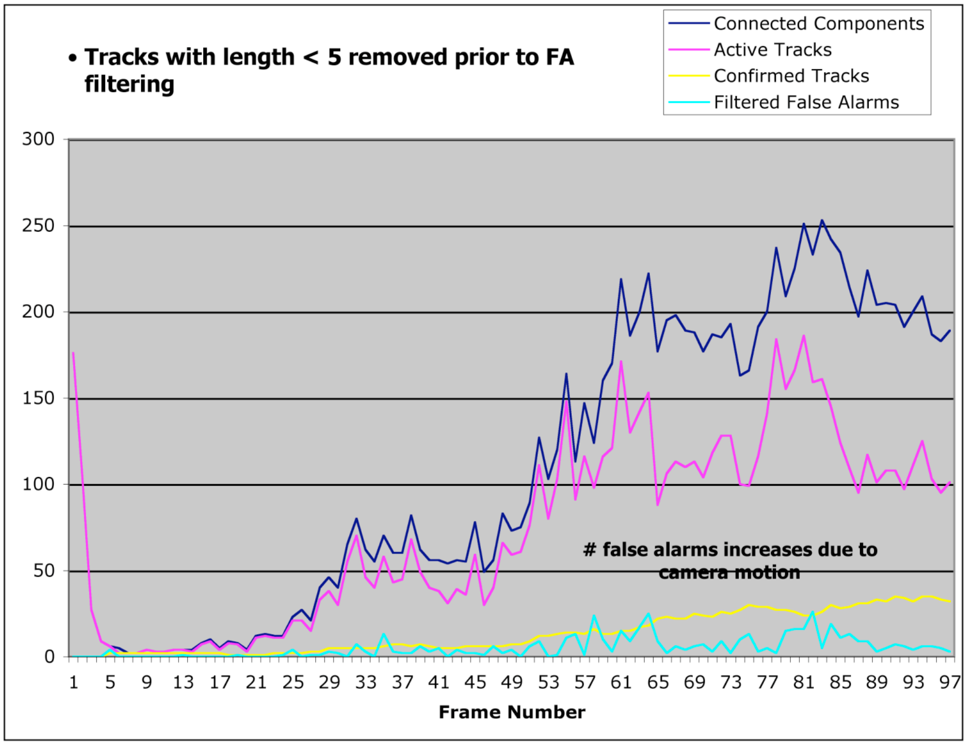


Figure 3 Track-based filtering characteristics of aerial video

Track-Based Filtering

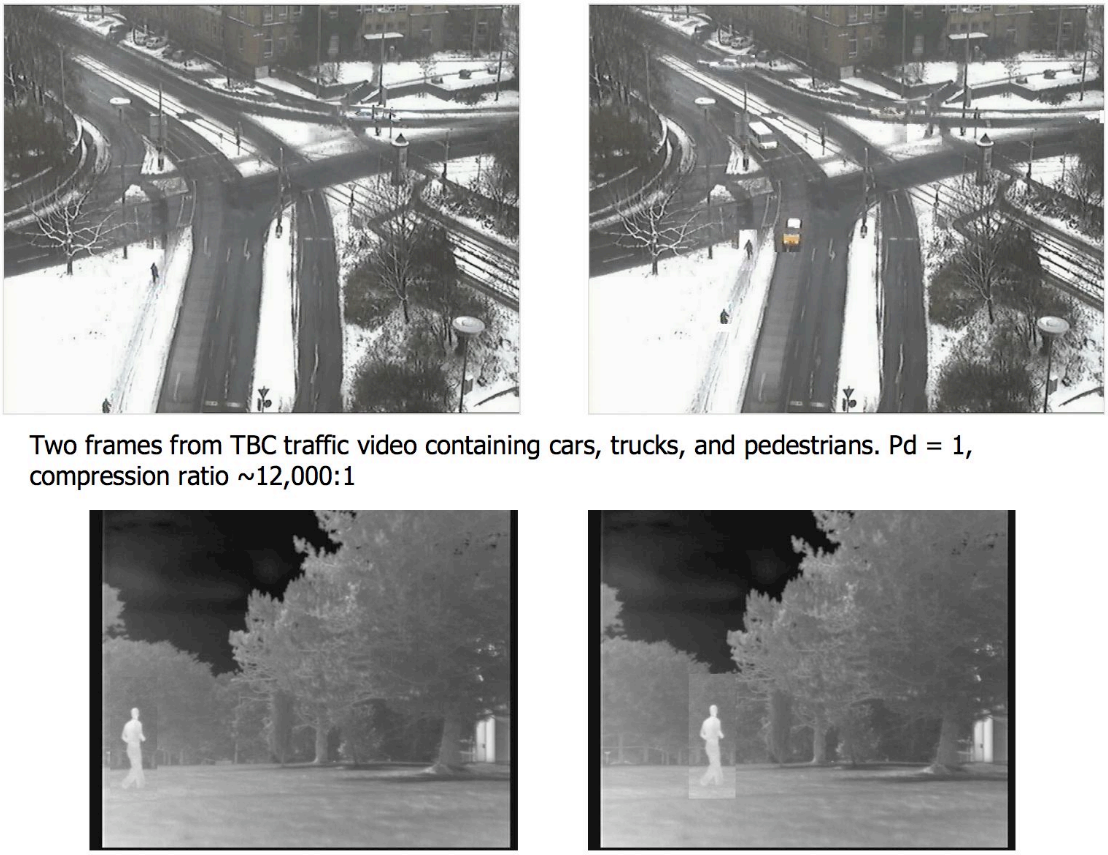
In addition to standard kinematic trackers, we have developed a pixel-level tracker that can maintain track on thousands of objects in wide field of view (WFoV) video. Our approach is to track all changes, filtering out those that do not meet certain criteria such as persistence (minimum number of reports), displacement (distance and direction), and consistency (size and brightness). Pixel-level associations instantiate track hypotheses in an *M x L* element track table where *M* is the maximum number of active tracks that can be maintained at a time. Tracks that pass filtering are output after a delay of *L* frames.

Figure 3 illustrates the benefit of track filtering. (The vertical axis represents the number of connected components, active or confirmed tracks, or false alarms that have been filtered, depending on the color of the curve.) Over the period of the video, two abrupt shifts in camera position introduce a large number of false changes (two spikes in the figure). Some of the detected object regions (connected components) can be filtered based on their size and shape. Most do not persist long enough to become confirmed tracks. Those that do persist but do not displace by an expected amount are further filtered. Even though the number of detections varies over a 10:1 range, the number of confirmed tracks remains relatively constant.

Table 1 Sample track-based compression performance results

Data set	CR	PD*	Description
UGS (thermal IR)	13,000:1	1	Fixed platform. Tracks people moving in/out of scene with moving background clutter (trees and shadows).
Traffic videos	12,000:1	0.95	Fixed platform. Tracks cars, trucks, and pedestrians.
Low altitude aerial surveillance	1470:1	0.75	Moving platform. Oblique video (elevation angle ~ 30°) with severe parallax. PD reduced due to obscuration. Uncorrected parallax errors increases FAR and reduces R.
WFoV	9130:1	0.95	Moving platform. Tracks large well-spaced targets including trucks, airplanes, and ships.

*For objects within a given size range.



Two frames from TBC traffic video containing cars, trucks, and pedestrians. Pd = 1, compression ratio ~12,000:1

Two frames from TBC IR UGS video containing persons walking and moving background clutter. Pd = 1, compression ratio ~13,000:1

Figure 4 Track-based compression (TBC) exhibits. (Videos courtesy U. Karlsruhe and NVL.)

Track-based Compression Performance

Table 1 lists TBC performance for several different scenarios. These numbers include sending the reports and chips. Even greater compression ratios (> 100,000:1) are possible by sending only the track reports, e.g., inserting icons onto maps that have already been stored at the client side. Figure 4 shows a basic TBC reconstruction where the chip is inserted into the background image at the centroid of the target track. More sophisticated renderings are also possible.

The compression ratio is affected by the FAR and P_D. Reduced P_D causes track fragmentation, which increases the number of tracks, and thus the number of chips that must be sent. Best performance is achieved in overhead imagery near nadir, and in ground surveillance with minimal obscuration and low background motion clutter.

Summary

A preliminary version of a track-based compression algorithm has been implemented and tested on several different kinds of video data with promising results. In addition to compression, TBC also serves as a data conditioner for downstream processing of video information. Chip data can be fed downstream to automatic target recognition algorithms that provide IDs on targets in track. This information, together with the track reports can then be fused with other intelligence data to develop a comprehensive picture of the battlespace. Other uses of the compressed video stream are also possible.

Acknowledgement

The authors thank Army Night Vision Lab for supplying the UGS video.

References

Mark Carlotto, “Detecting change in images with parallax,” *SPIE’s 2007 Signal Processing, Sensor Fusion, and Target Recognition XV*, Orlando FL.

Chris Stauffer and W. Eric L. Grimson, “Learning patterns of activity using real-time tracking,” *IEEE Trans. PAMI*, Vol. 22, No. 8, August 2000.

Ahmed Elgammal, Ramani Duraiswami, David Harwood and Larry Davis, “Background and foreground modeling using nonparametric kernel density estimation for visual surveillance,” *Proc. IEEE*, Vol. 90, No. 7, July 2002.