OBJECT COLOR PROPAGATION IN AN UNREGISTERED DISTRIBUTED VIDEO SENSOR NETWORK

Vijay Venkataraman, Sabeshan Srinivasan, Anthony Stefanidis

Department of Spatial Information Science and Engineering and
National Center for Geographic Information & Analysis
University of Maine, 348 Boardman Hall, Orono, ME 04469-5711, USA
{vijay, sabeshan, tony}@spatial.maine.edu

ABSTRACT

In this paper we address the use of object color characteristics as a predominant linking parameter between disjoint video scenes in unregistered distributed video sensor networks. In our approach we proceed by tracking objects in high-resolution video scenes using accumulated frame differencing and morphological techniques. We continue by minimizing shadow effects in object tracking through background detection, in order to obtain more accurate object color signatures. These signatures are then used to link trajectory fragments across disjoint video feeds in a sensor network. We present experimental results to demonstrate the performance of our approach, especially addressing the effect of resolution in tracking, and the role of color signatures as a key linking element when video scene registration is not possible.

1. INTRODUCTION

Object detection and tracking techniques from motion imagery have progressed significantly and are currently employed in a variety of applications ranging from classic surveillance and human motion analysis [6] to complex traffic monitoring [1] and human-machine interfaces [4]. In addition to software solutions, hardware-based object tracking systems have also been developed to support the real-time processing of captured video data.

Parallel to these computer vision developments, we can identify a notable trend in geospatial applications towards the use of sensor networks for data collection [8]. While sensor technology is rapidly evolving towards the development of highly specialized micro sensors (e.g. collecting temperature or humidity readings), optical sensors still remain the dominant source of geospatial information (objects and their locations in space). Sensor networks of optical sensors may range from a fleet of UAVs monitoring an area of interest to a network of building-level surveillance cameras. Processing information collected from optical sensor networks requires efficient methods to track objects across disjoint video feeds, and this is the problem we address in this paper.

Common approaches to object tracking in a multi-sensor video network typically assume video registration, i.e., positions and orientation of the video cameras are known, and object trajectories in captured video streams can be brought to a common reference frame [2,5]. Within such a framework, Markov models of motion paths and probabilistic descriptions of a sensor network topology (and transitions within it) have been used to support the tracking of objects in non-overlapping video feeds [3].

While assuming a known location for our cameras is beneficial and desirable, sometimes this information may not be readily available. This is especially the case when dealing with mobile camera networks. In this paper we address the use of color to link tracked objects across disjoint camera feeds. Our motivation stems from the well-known robustness of color properties under rotation, scale and resolution changes [4] across disjoint scenes.

Our objective is to track cars in individual video feeds and determine use their color signatures to link the same object in two different feeds. We proceed by tracking objects in high-resolution video scenes using accumulated frame differencing and morphological techniques. We continue by minimizing shadow effects in object tracking through background detection, in order to obtain more accurate object color signatures. These signatures are then used to link trajectory fragments across disjoint video feeds in a sensor network. We present experimental results obtained using non-overlapping sensors monitoring traffic in a campus to demonstrate the performance of our approach. We are especially interested in the effect of resolution in our tracking results, and the role of color signatures as a key linking element when video scene registration is not possible.

The paper is organized as follows. In Section 2 we present a general framework for object detection and tracking, and some details behind key processes. In Section 3 we present our techniques to generate color signatures and minimize shadows. Experimental results are presented in Section 4, followed by conclusions in Section 5.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

2. OBJECT TRACKING IN A SENSOR NETWORK

Our approach to object motion detection and tracking is outlined in Figure 1. We use unregistered video scenes that are devoid of positional and orientation information. Scene, camera and object characteristics like scale, illumination, elevation, view azimuth, shape profile, velocity are assumed to vary across the sensor network. High-resolution color video captured by the sensors is pre-processed before objects are tracked.

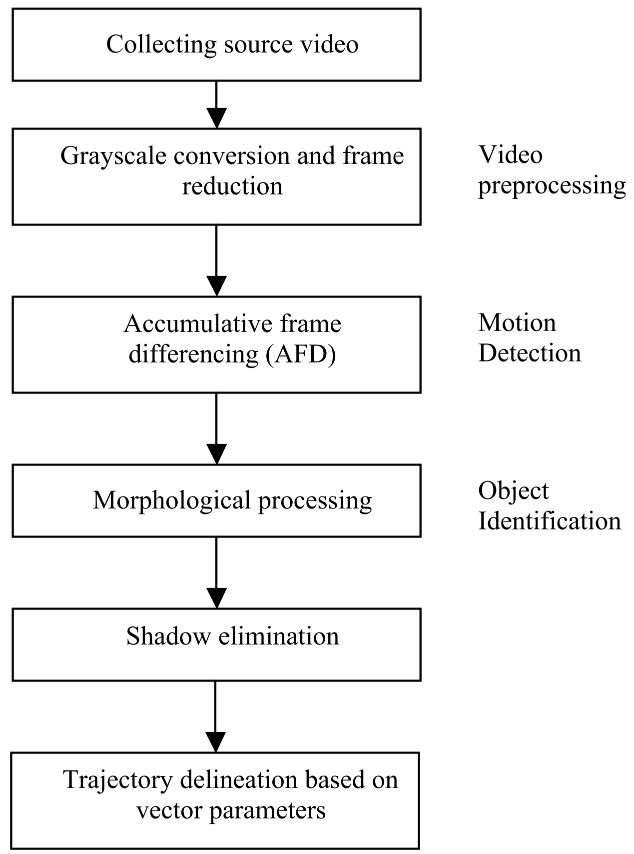


Figure 1: Object detection and trajectory delineation.

The general process we follow is outlined in Fig. 1. We start with accumulated frame differencing (AFD) to detect motion in our video feeds. This allows us to detect obtain object blobs and corresponding color pixels from the original videos. The blobs detected with AFD are then subject to morphological processing, in order to eliminate wholes in them. As shadows move together with the objects they correspond to, blobs detected through this process comprise both objects and their shadows. This would obviously affect the radiometric properties of the blobs, and to avoid that we make use of a shadow elimination process. Once this is accomplished, object trajectories are refined and linked across sensor feeds using the radiometric signatures of the tracked objects. In the next subsections we present details on key issues behind our approach.

2.1. Accumulative frame differencing parameters

Accumulative frame differencing determines motion on the basis of change in pixel color by subtracting or 'differencing' a sequence of frames. The Differencing Mode (DM) can be positive, negative, or absolute, thus tracking the back, front, or both sides of a moving object. The three modes differ in the manner in which the pixel intensity difference (PID) is compared with a pre-defined gray level threshold (*GLT)*. PID is the difference in color values for the same pixel (*x,y*) in the scene grid over time. *GLT* is the minimum gray level change that must occur for an object to be detected as moving across two frames. PID is computed between the reference and all other frames in the video stream as follows:

$$PID = R(x,y) - f(x,y,t_k) \quad (1)$$

where *R(x,y)* is the reference image and $f(x,y,t_k)$ is the frame at time t_k. Usually, the reference is the first frame. A new value for each pixel in the currently processed frame is obtained depending on the presence/absence of motion in the neighborhood of each pixel. The new value is binary and is allocated as follows:

$$I_k(x,y) = \begin{cases} I_{k-1}(x,y)+1 \\ I_{k-1}(x,y) \end{cases} \quad (2)$$

where $I_k(x,y)$ is the gray value at pixel *(x,y)* of the current frame and its value depends on its value in the previous frame $I_{k-1}(x,y)$. Pixel values are either incremented by one or left unchanged from the value in the previous frame depending on the differencing mode. In addition to the above, two other parameters influence AFD. Frame accumulation rate (FAR) is determined iteratively and is the number of frames over which the difference from the reference frame will be accumulated. Accumulation threshold (AT) is the minimum number of frames across which an object must exhibit change in order to be detected as moving. This helps us eliminate periodic noise. Its value may be selected such that noise is removed while still preserving slow moving objects.

2.2. Morphological parameters

Morphological parameters isolate objects that satisfy a certain shape criterion. The structuring element is a bounding box used to establish connection between neighborhood pixels with an object. The minimum (mOA) and maximum (MOA) object areas (pixels) constrain the allowable object sizes. Object compactness (C) is a shape metric defining the elliptical shape of the object, and ranges between 0 (straight line) and 1 (circle). Object compactness can be more reliably applied when dealing with objects with jagged edges or irregular shapes.

2.3. Vector building parameters

Vector building parameters control the creation of motion vectors of the filtered objects. The *Maximum link distance* specifies the maximum distance in pixels that the centroid can undergo change between successive motion frames. The *Maximum spectral distance* specifies maximum allowable variation in any color metric between successive motion frames. In a predominantly color based test, the link distance is set to a large value and the spectral distance becomes the key linking parameter.

3. COLOR SIGNATURES AND SHADOWS

The apparent surface color of an object is determined by source illumination, viewing geometry and camera parameters. Processing color information can be expensive and is typically restricted to the pixels that have been obtained from the morphological operations on the object shape profile. Color information can be operated upon in either the RGB or the HSI (Hue; Saturation; Intensity) domains. In the RGB domain, isolating color information is not easy since all the three values (R,G,B) change significantly with illumination. In the HSI domain, the Hue value plays a vital part in object determination since it does not vary with intensity changes. This is particularly true when the object is subject to varying illumination levels as under shadows and under bright sunlight. Hence we attempt to track objects based on proximity in Hue values after shadows are eliminated. In order to get a Hue signature, the median hue value of pixels in an object blob is used.

Figure 2: Object before and after shadow filtration.

In order to eliminate effects of shadows we define a rectangular bounding box containing the detected object. Pixel values outside the object outline but inside the bounding box belong to the surface on which the object traveled, in this case, the road. The frequency distribution of the hue values of these pixels was studied and a range of hue values which represented the road surface was chosen. The difference between object shadows and road surface is simply in illumination, as the actual material is the same. Hence, the hue values of object shadows are a very close match to the hue values of the road pixels. The basis for this assumption is the premise that hue value of a surface does not vary with incident illumination variations when surface material remains the same.

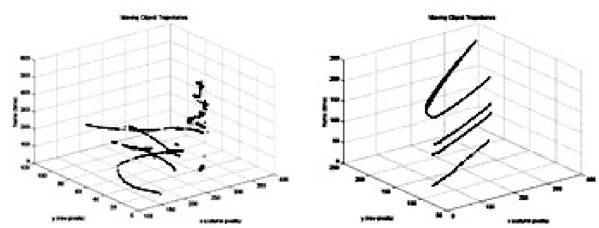
Once the range of hue values corresponding to shadows has been identified, those pixels in the vicinity of an object with hue values falling in the 'shadow range' are eliminated from the corresponding object blob (Fig. 2). By removing shadows from the tracked blobs, the corresponding color signatures are more robust for object linking across feeds.

4. EXPERIMENTS

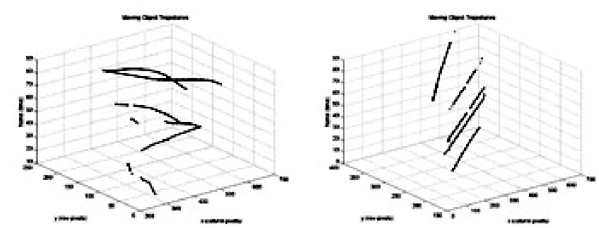
In order to evaluate the performance of our approach we used video camcorders monitoring traffic in our Campus. We captured color video at 720*480 pixels and a frame rate of 30 fps. This was reduced to 320*240 at 8 fps for testing low-resolution analysis and 720*480 at 3 fps for high-resolution analysis. Fig. 3 shows samples of our feeds, with camcorder S1 monitoring a parking lot, while camcorder S2 monitors a perpendicular exit road, offering a view disjoint from S1. The data from both the video sensors contain realistic scenarios of variations in illumination within a single sensor and across sensors. S1 has a distinct building shadow that introduces non-uniformity in illumination within individual objects.

Figure 3**:** Sample video feeds from sensors S1 (left) and S2 (right).

The algorithms accept preset values for parameters described in Section 2 that are derived iteratively. Parameter values are applied consistently on both video scenes and act as global constraints, establishing a knowledge base for that distributed video sensor network.



(a) S1: 320*240; 8fps (b) S2: 320*240; 8fps



(c) S1: 720*480; 3fps (d) S2: 720*480; 3fps

Figure 4: Motion vectors: Low- (a, b) and High- (c, d) resolutions

The color characteristics of two autos, a Red Truck (RT) and a Black SUV (BS) that appear in both video feeds were analyzed, and trajectories were generated based on hue proximity. The trajectories were then compared across scenes based on the color signatures from both objects to see if they matched uniquely. The motion trajectories for the two objects in Scenes 1 and 2, as they were captured on different resolution feeds are shown in Fig. 4. The trajectories are plotted on the X-Y plane with time along the vertical axis. Shadow filtering was performed only on the high resolution analysis due to high loss of object surface pixels when performing the filtration. Parameters used to delineate these trajectories are shown in Table 1.

Resolution	Size	FR	C	mOA	MOA
Low	320*240	8 fps	0	200	10000
High	720*480	3 fps	0	4000	10000

Table 1: Parameter values for scenes 1 and 2

Since gaps appear in the AFD output video due to the objects stalling to turn or reverse, gaps appear in the video sequence and fragment the trajectories of object motion as seen in the S1 trajectories (Figs. 4a,c). In S2 (Figs. 4b,d) we find more continuous and unique trajectories, as the corresponding movement is smooth and unobstructed.

Object	S1 Traj.	S2 Traj.	Match Reliability
RT	1-8	2	99.8%
BS	9-13	3	FAIL

(a) Low-resolution results (320*240) – Shadows intact

Object	S1 Traj.	S2 Traj.	Match Reliability
RT	1-7	2	99.5%
BS	8	3	83.8%

(b) High-resolution results (720*480) – Shadows removed

Table 2: Results of comparing color linked trajectories

In table 2 we show representative results when linking trajectories identified in these two feeds, in both low and high resolution. In order for a trajectory pairing to be considered as a match, we requested that trajectories should map uniquely from S1 to S2. Thus we aimed at unambiguous, robust matches. Considering low resolution feeds we can see that the red truck matched very well across the two feeds, while the black SUV in S1 failed to match exclusively to itself in S2. In high resolution analysis, followed by shadow removal, this problem was remedied and both vehicles in S1 were successfully mapped uniquely to their counterparts in S2, with a high degree of reliability.

5. CONCLUDING REMARKS

In this paper we presented our approach to object tracking in video feeds and linking across non-overlapping feeds using primarily object color signatures, and shadow elimination. Experimental results verify the importance of color as an invariant property to link trajectory segments across multiple feeds without using any sensor geopositioning information. By combining high resolution feed processing and shadow removal we can successfully link objects using solely their radiometric content. We are currently working on extending our work on object signatures to include geometric and spatiotemporal properties, and the development of a fuzzy control system to form a base for detecting different classes of objects (autos, bikes, people, etc.).

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation through grant ITR-0121269, and the National Geospatial-Intelligence Agency through NMA 401-02-1-2008.

REFERENCES

[1] Chachich A., A. Pau, A. Barber, K. Kennedy, E. Olejniczak, J. Hackney, Q. Sun, and E. Mireles, Traffic Sensor using a Color Vision Method, *SPIE Proceedings 2902*, pp. 156-165, 1996.

[2] Collins R.T., A. Lipton, H. Fujiyoshi, and T. Kanade, Algorithms for Cooperative Multisensor Surveillance, *Proceedings of the IEEE*, 89(10), pp 1456-1477, 2001.

[3] Jaynes C., Acquisition of a Predictive Markov Model using Object Tracking and Correspondence in Geospatial Video Surveillance Networks, *GeoSensor Networks,* (A. Stefanidis & S. Nittel eds.), CRC Press, pp. 149-166, 2004.

[4] Khan S., and M. Shah, Consistent Labeling of Tracked Objects in Multiple Cameras with Overlapping Fields of View, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 25(10), pp. 1355-1360, October 2003.

[5] Lee L., R. Romano, and G. Stein, Monitoring Activities from Multiple Video Streams: Establishing a Common Coordinate Frame, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 22(8), pp. 758-767, August 2000.

[6] Lim S.N., L. Davis, & A. Elgammal, A Scalable Image-based Multi-Camera Visual Surveillance System, *IEEE Conf. on Advanced Video & Signal Based Surveillance (AVSS)*, Miami, pp. 205-212, 2003.

[7] Raja Y., S. J. McKenna, and S. Gong, Colour Model Selection and Adaptation in Dynamic Scenes, *European Conf. on Computer Vision (ECCV),* Freiburg, pp. 460-474, 1998.

[8] Stefanidis A. and S. Nittel, *GeoSensor Networks*, CRC Press, Boca Raton, 2004.