International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol. 2, Issue. 5, Sept.-Oct. 2012 pp-3013-3016 ISSN: 2249-6645

Vehicle's Tracking and Recognition Using a Distributed Surveillance System for Urban Traffic Management

Peyman Babaei
Dep. of Computer, West Tehran Branch, Islamic Azad University, Tehran, Iran

Abstract: *This paper proposes an unsupervised vehicle's tracking and recognition methods for urban Traffic surveillance in a distributed cooperative manner. Vehicle's matching in a multi-camera surveillance system is a fundamental issue for increasing the accuracy of recognition. In intelligent transportation systems (ITS), especially in field of urban traffic management, intersections monitoring is one of the critical and challenging tasks. In multi-camera traffic surveillance system, videos have different characteristics such as pose, scale and illumination. Therefore it is necessary to use a hybrid scheme of scale invariant feature transform (SIFT) to detection and recognition vehicle's behavior in multi view more accurately and conveniently. The main focus of this paper is to analyze activities at intersection by a distributed cooperative system for tracking and recognition vehicles to extract traffic flows which assists in regulating traffic lights for using in smart cameras. Extracting the trajectories help to detect abnormal behavior which may be occluded in single- camera surveillance. Distributed cooperative's fundamental purpose is to efficiently reduce the transmission rate and also analyze an intersection scene and report statics and information of interest.*

Keywords: *Distributed system, Intersection monitoring, Multi-camera surveillance, Vehicle's behaviour learning, urban traffic management.*

I. INTRODUCTION

Video surveillance is widely employed in commercial applications and public transportation for purposes of statistics gathering, processing and traffic flow monitoring. The number of cameras and complexity of surveillance systems have been continuously increasing to have better coverage and accuracy. Tracking and behavior recognition are two fundamental tasks in this regard. Multi-camera systems become increasingly attractive in machine vision. Applications include multi view object tracking, event detection, occlusion handling and etc. For many applications, there may be constraints of transmission bandwidth and complexity in analyzing a huge amount of data centrally. In intelligent transportation systems (ITS), the convenient conditions are aroused from autonomous agents making decisions in a decentralized manner. In this paper, we develop method for tracking and recognition by a traffic video surveillance system of two distributed cameras with a partially overlapping field of view. We show how to develop methods for tracking and recognition in a system where processing and decision is distributed across the cameras.

This paper is organized as follows: an overview of the past works in section2. Our proposed architecture and algorithm is presented in section3. Results of subjective evaluations and objective performance measurements with respect to Ground-truth are presented in section4. Section5 contains the conclusion.

II. PAST WORKS ON MULTI-CAMERA SURVEILLANCE

Features' matching between multiple images of a scene is an important component of many computer vision tasks. In the last few years, a lot of works in detecting, describing and matching feature points has deployed. Although the correspondences can be hand selected, such a procedure is hardly conceivable as the number of cameras increases or when the camera configuration changes frequently, as in a network of pan-tilt-zoom cameras [1]. Other methods for finding correspondences across cameras [2] have been developed through a feature detection method such as the Harris corner detection method [3] or scale invariant feature transform (SIFT) [4]. In [5] shown that corners were efficient for tracking and estimating structure from motion. A corner detector is robust to changes in rotation and intensity but is very sensitive to changes in scale. The Harris detector finds points where the local image geometry has high curvature in the direction of both maximal and minimal curvature, as provided by the eigen-values of the Hessian matrix. They develop an efficient method for determining the relative magnitude of the eigen-values without explicitly computing them. Such color-based matching methods have also been used to track moving objects across cameras [6, 7]. Scale invariant features matching were first proposed in [8] and attracted the attention of the computer vision systems for invariant to scale, rotation, and view-point variations. Also uses a scale-invariant detector in the difference of Gaussian (DOG) scale space. In [4] fits a quadratic to the local scale-space neighborhood to improve accuracy. He then creates a Scale Invariant Feature Transform (SIFT) descriptor to match key-points using a Euclidean distance metric in an efficient best-bin first algorithm where a match is rejected if the ratio of the best and second best matches is greater than a threshold.

A comparative study of many local image descriptors [9] shows the superiority of SIFT with respect to other feature descriptors for the case of several local transformations. In [10] develop a scale-invariant Harris detector that keeps key points at each scale only if it's a maximum in the Laplacian scale-space [11]. More recently, in [12] integrate edge-based features with local feature-based recognition using a structure similar to shape contexts [13] for general object-class recognition. In [14] propose a matching technique based on the Harris corner detector and a description based on the Fourier transform to achieve invariance to rotation. Harris corners are also used in [15], where rotation invariance is obtained by a hierarchal

sampling that starts from the direction of the gradient. In [16] introduce the concept of maximally stable external region to be used for robust matching. These regions are connected components of pixels which are brighter or darker than pixels on the region's contour; they are invariant to affine and perspective transform, and to monotonic transformation of image intensities. Among the many recent works populating the literature on key-point detection, it is worth mentioning the scale and affine invariant interesting points recently proposed in [17], as they appear to be among the most promising key-point detectors to date. The detection algorithm can be sketched as follows: first Harris corners are detected at multiple scales, and then points at which a local measure of variation is maximal over scale are selected. This provides a set of distinctive points at the appropriate scale. Finally, an iterative algorithm modifies location, scale, and neighborhood of each point and converges to affine invariant points. In [18] describe a matching procedure wherein motion trajectories of objects tracked in different cameras are matched so that the overall ground plane can be aligned across cameras following a homograph transformation. A similar approach has been proposed in [19-21] which again motion tracks are matched together. However, although use scene dynamics to find matches, unlike our method, these methods first need to solve the problems of single camera tracking and data association across cameras, which is difficult in highly cluttered scenes or when moving objects occlude each other.

III. PROPOSED ARCHITECTURE AND ALGORITHM

First, we review the function of a typical single-camera and multi-camera surveillance system as presented in our previous work. At that work, as mentioned below the system was centralized. Next, the architecture and algorithm of distributed cooperative system is presented.

A. Single-camera and multi-camera surveillance functionality

As presented in our previous work [22], the function of a typical single-camera surveillance system is illustrated in Fig.1. The first part of the processing flowchart is very general, which is marked "Detecting & Matching Features Extraction Pipeline". This pipeline may produce all target information (pose, scale, illumination, color, shape, etc.), and potentially the description of the scene. The end of the processing pipeline, the vehicle tracking and classification is done.

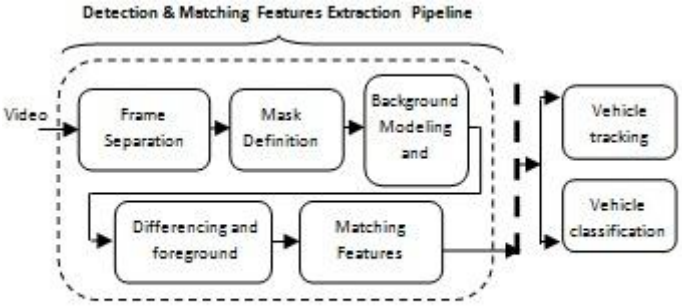


Figure1. Single-camera video surveillance Flowchart

Only the matching features have to be stored, instead of high quality video suitable for automated processing. This method enables the multi-camera surveillance system. The video surveillance system, as described in the above, cannot provide an adequate solution for many applications, Such as urban traffic management with all its associated limitations [23-27]. A multi-camera surveillance system tracking targets from one camera to the next can overcome all these limitations. A typical multi-camera surveillance system is illustrated in Fig.2. Fusing at the matching features level requires merging all the features from the cameras on to a full representation of the environment. This approach distributes the most time consuming processing between the different cameras, and minimizes communication, since only the extracted features needs to be transmitted, no video or image. Given these advantages, system communicates only the matching features for fusion.

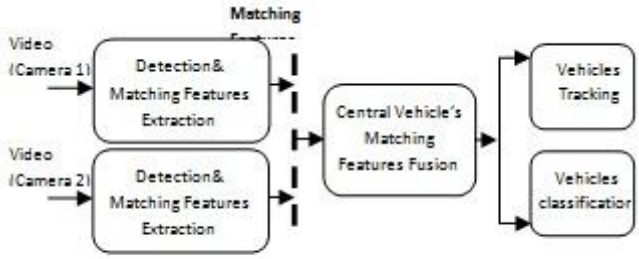


Figure2. Multi-camera video surveillance Flowchart

B. Proposed architecture and algorithm for distributed cooperative system

The problem of multi view activity recognition has been addressed in many papers, but almost the information of multiple views is fused centrally. Our proposed framework is decentralized. The pose of cameras at intersection is shown in Fig.3. In Fig.4, the structure of distributing levels is illustrated.

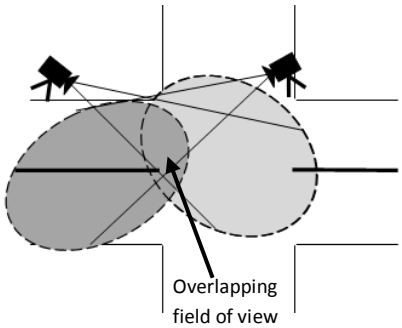


Figure3. Camera setup of cooperative system

Each of the cameras has processing cores in four levels which is described in the flowchart in Fig.4. The input stream is fed to detection level. At the decision level, control commands are issued to classify the detected vehicles based on extracted description features. Processing cores in three upper levels exchange the requisite information to track and recognition more accurately.

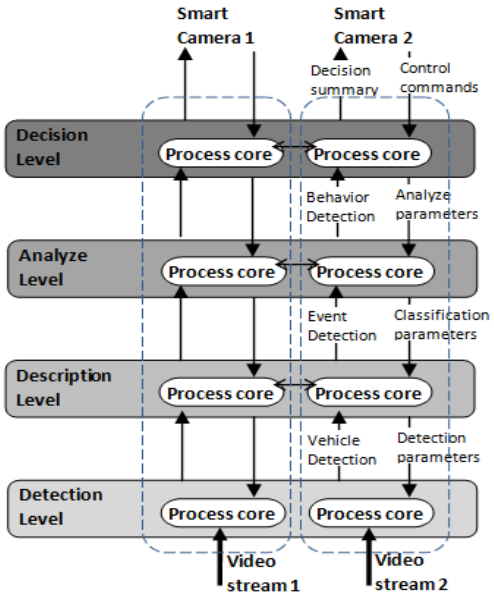


Figure4. Cooperative Levels in proposed distributed system

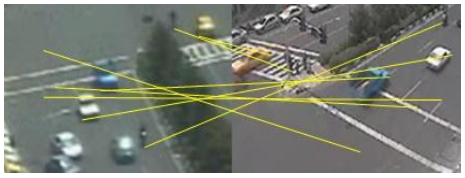
The principle features of our scheme are summarized in the following:

Communication Efficiency: distributed cooperative system is particularly well-suited for low bandwidth; therefore the requirement processing is done locally.

Unsupervised: The method does not require the pre-calibration into the scene and, hence, can be used in traffic scenes where the system administrator may not have control over the activities taking place.

The SIFT (Scale Invariant Feature Transform) [4] has been shown to perform better than other local descriptors [9]. Given a feature point, the SIFT descriptor computes the gradient vector for each pixel in the feature point's neighborhood and builds a normalized histogram of gradient directions. The SIFT descriptor creates a 16×16 neighborhood that is partitioned into 16 sub-regions of 4×4 pixels each. For each pixel within a sub-region, SIFT adds the pixel's gradient vector to a histogram of gradient directions by quantizing each orientation to one of 8 directions and weighting the contribution of each vector by its magnitude. Each gradient direction is further weighted by a Gaussian of scale $\sigma = n/2$ where n is the neighborhood size and the values are distributed to neighboring bins using interpolation to reduce boundary effects as samples move between positions and orientations. Fig.5 shows the matching results using SIFT created for a corresponding pair of points in two intersection scenes.

(a)



(b)

Figure5. Three different intersection scenes, (matching results using SIFT)

IV. EXPERIMENTAL RESULTS

Here it is shown that SIFT lead to excellent performances compared to other existing approaches. As explained, SIFT description is computed as follows: once a key-point is located and its scale has been estimated, one or more orientations are assigned to it based on local image gradient direction around the key-point. Then, image gradient magnitude and orientation are sampled around the key-point, using the scale of the key-point to select the level of Gaussian blur. The gradient orientations obtained are rotated with respect to the key-point orientation previously computed. Finally, the area around the key-point is divided in sub-regions, each of which is associated an orientations histogram weighted with the magnitude. We have experimented with various feature detectors including the Harris corner detector (HCD), curvilinear structure detector (CSD), and difference of Gaussian (DoG) scale space. In Fig.6, the experimental result contain the comparison of these methods is shown.

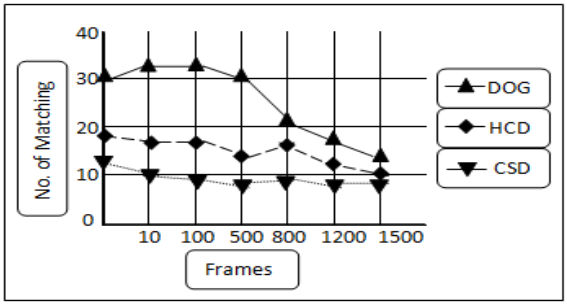


Figure6. Efficiency comparison in intersection traffic scenes

In table1 counting and classification results are presented. As shown, the overall accuracy is about 91% for using DOG detector in counting cars and about 90% for Bus and Trucks. This system can be as an input to calibration system in multi-camera surveillance system.

Table1. Counting and classification results

No. of Vehicle's matching in overlapping field of view						
Vehicle Type	Bus & Truck			Car		
	DOG	HCD	CSD	DOG	HCD	CSD
Count by Detector	27	21	18	116	108	101
Ground Truth	30	30	30	127	127	127
Precision	90%	70%	60%	91%	85%	79%

V. CONCLUSION

In this paper we considered the problem of features matching in a distributed cooperative system with overlapping fields of view. We showed that suing SIFT point descriptors in a distributed cooperative surveillance system can improves the performance with respect to the other calibration systems. In particular it returned good results for scale changes, severe zoom and image plane rotations, and large view-point variations. These conclusions are supported by an extensive experimental evaluation, on different traffic scenes in urban traffic. Therefore, tracking and recognition using SIFT becomes feasible. This should result in highly robust trackers.

ACKNOWLEDGEMENTS

This work was supported by Islamic Azad University, West Tehran Branch.

References

[1] E.B.Ermis, P.Clarot, P.Jodoin, "Activity Based Matching in Distributed Camera Networks," IEEE Transaction on Image Processing, vol. 19, no. 10, OCT. 2010, pp.2595–2613.
[2] D.Devarajan, Z.Cheng, and R Radke, "Calibrating distributed camera networks," Proc. IEEE, vol. 96, no. 10, OCT. 2008, pp. 1625–1639.
[3] C.Harris and M.Stephens, "A combined corner and edge detector," in Proc of 4th Alvey Vision Conf., 1988, pp. 147–151.
[4] D.Lowe,"Distinctive image features from scale-invariant keypoints," in Int. J. Comput. Vis., vol. 60, no. 2, 2004, pp. 91–110.
[5] C. Harris, Geometry from visual motion. in: A. Blake, A.Yuille (Eds), Active Vision, MIT Press, 1992.
[6] B.Song and A.R.Chowdhury,"Stochastic adaptive tracking in a camera network," in Proc.IEEE Int.Conf.Computer Vision, 2007, pp.1–8.
[7] S.Khan and M.Shah,"Tracking multiple occluding people by localizing on multiple scene planes,"IEEE Trans. Pattern Anal. Mach. Intell.,vol. 31, no. 3 , Mar. 2009, pp. 505–519.
[8] D.G.Lowe, "Object recognition from local scale-invariant features," in Proc. Of ICCV, 1999, pp. 1150–1157.
[9] K.Mikolajczyk, C.Schmid,"A performance evaluation of local descriptors," in Proc. Of CVPR, 2003, pp. 257–263.
[10] K.Mikolajczyk and C.Schmid, "Indexing based on scale invariant interest points," in Proc. Of ICCV ,2001, pp. 525-531.
[11] T.Lindeberg, "Feature detection with automatic scale selection," in Proc. Of IJCV, vol. 30, no.2, 1998, pp.79-116.
[12] K.Mikolajczyk, A.Zisserman,"Shape recognition with edge-based features," in Proc. of the British conf. Machine Vision 2003.
[13] S.Belongie, J.Malik and J.Puzicha, "Shape context: A new descriptor for shape matching and object recognition," in Proc. Of NIPS, 2000 pp. 831-837.
[14] A.Baumberg,"Reliable feature matching across widely separated views," in Proc. of CVPR, 2000, pp.774– 781.
[15] N.Allezard, M.Dhome, F.Jurie,"Recognition of 3D textured objects by mixing view-based and model based representations," in Proc. of ICPR, 2000.
[16] J.Matas, O.Chum, M.Urban, T.Pajdla,"Robust wide baseline stereo from maximally stable extremal regions," in Proc. of BMVC, 2002, pp. 384–393.
[17] K.Mikolajczyk, C.Schmid,"Scale and affine invariant interest point detectors," in Int. J. Com. Vis. Vol. 60, no.1, 2004, pp. 63–86.
[18] L.Lee, R.Romano, and G.Stein,"Monitoring activities from multiple video streams: Establishing a common coordinate frame," IEEE Trans.Pattern Anal.Mach Intell., vol.22, no.8, Aug.2000. pp.758–767.
[19] S.Khan and M.Shah,"Consistent labeling of tracked objects in multiple cameras with overlapping fields of view," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 10, Oct. 2003, pp.1355–1360.
[20] X.Wang, K.Tieu, and E.Grimson,"Correspondence-free activity analysis and scene modeling in multiple camera views," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 1 , Jan. 2010, pp. 56–71.
[21] D.Makris, T.Ellis, and J.Black, "Bridging the gap between cameras," in Proc. of CVPR, vol. 2, 2004, pp. 205–210.
[22] P.Babaei and M.Fathy,"Multi-Camera Systems Evaluation in Urban Traffic Surveillance versus Traditional Single-Camera Systems for Vehicles Tracking," in Proc. of ICCEE 2010, vol.4, Nov.2010, pp.438-442.
[23] H. Veeraraghavan and N. Papanikolopoulos, "Combining multiple tracking modalities for vehicle tracking at traffic intersections," in IEEE Conf. on Robotics and Automation, 2004.
[24] S.Khan and M.Shan, "Consistent labeling of tracking objects in multiple cameras with overlapping fields of view", In IEEE Trans., on PAMI, vol.25, no.10, Oct.2003 , pp.1355-1360.
[25] Q.Zhou, J.Park and J.K.Aggarwal, "Quaternion-Based Tracking of Multiple Objects in synchronized video", In Proc. of Intl. Sym. on Computer and Information Sciences , 2003, pp.430-438.
[26] S.L.Dockstsder and A.M.Tekalp, "Multiple camera tracking of interacting and occluded human motion", In Proc. of IEEE. vol.89, no.10, Oct.2001, pp.1441-1455.
[27] L.Lee, R.Romano and G.Stein, "Monitoring activities from multiple video streams: establishing a common coordinate frame", In IEEE Trans. on PAMI, Vol.22, no.8, Aug.2000, pp.758-767.