International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol. 3, Issue. 3, May - June 2013 pp-1843-1846 ISSN: 2249-6645

Human Detection, Tracking and Trajectory Extraction in a Surveillance Camera network

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

ABSTRACT: *This paper proposes human tracking and recognition method in a camera network. Human matching in a multi-camera surveillance system is a fundamental issue for increasing the accuracy of recognition in multiple views of cameras. In camera network, videos have different characteristics such as pose, scale and illumination. Therefore it is necessary to use a hybrid scheme of scale invariant feature transform to detection and recognition human's behaviors. The main focus of this paper is to analyze activities for tracking and recognition humans to extract trajectories. Extracting the trajectories help to detect abnormal behavior which may be occluded in single- camera surveillance.*

KEYWORDS: *Camera network, Multi-camera surveillance, Human's behavior, Trajectories extraction.*

I. INTRODUCTION

Tracking and behavior recognition are two fundamental tasks in video surveillance systems which are widely employed in commercial applications for purposes of statistics gathering and processing. The number of cameras and complexity of surveillance systems have been continuously increasing to have better coverage and accuracy. Multi-camera systems become increasingly attractive in machine vision. Applications include multi view object tracking, event detection, occlusion handling and etc. In this paper, we develop method for tracking and recognition by a traffic video surveillance system of two cameras with a partially overlapping field of view.

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

In the last few years, a lot of works in detecting, describing and matching feature points has deployed. In a camera network features' matching between multiple images of a scene is an important component of many computer vision tasks. 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 [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. Then creates a Scale Invariant Feature Transform 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 this method 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 [19-21].

III. PROPOSED ARCHITECTURE

First, we review the function of a typical single-camera and multi-camera surveillance system 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 human tracking and classification is done.

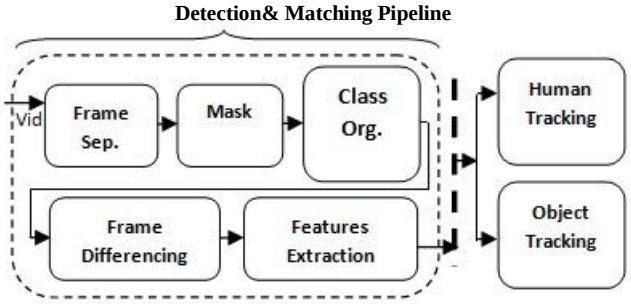

Fig. 1: Single Camera Processing

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 [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 network 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.

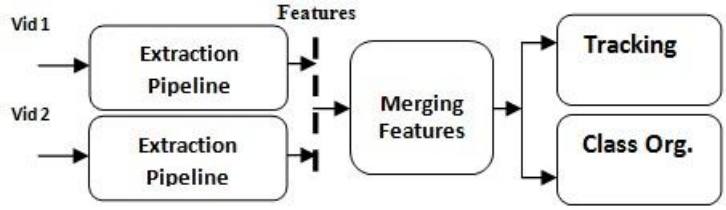

Fig. 2: Multi camera network Processing

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.

Fig 3: Camera setup in a network

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

Fig 4 Structure camera network

The Scale Invariant Feature Transform has been shown to perform better than other local descriptors [9]. Given a feature point, the descriptor computes the gradient vector for each pixel in the feature point's neighborhood and builds a normalized histogram of gradient directions. The descriptor creates a neighborhood that is partitioned into sub-regions of 4×4 pixels each. For each pixel within a sub-region and 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. Principle features of our scheme are summarized as Communication Efficiency: camera network is particularly well-suited for low bandwidth; and 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. Fig.5 shows the matching results using descriptor created for a corresponding pair of points.

Fig 5 Matching results using descriptor.

IV. EXPERIMENTAL RESULTS

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. We showed that suing SIFT point descriptors in a camera network can improves the performance with respect to the other calibration systems. Here it is shown that descriptor lead to excellent performances compared to other existing approaches. As explained, 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.

Table1: Number of matching by features descriptors.

True Positive & False Positive					
S1 Results (number of occlusion: 31)			**S2 Results (number of occlusion: 23)**		
t **seconds**	**True Positive**	**False Positive**	***t*** **seconds**	**True Positive**	**False Positive**
120s	8	0	120s	4	0
180s	10	0	180s	12	0
240s	17	0	240s	17	0
300s	22	1	300s	27	1
360s	33	1	360s	31	1
420s	45	2	420s	38	1
480s	52	3	480s	41	2

In table2 counting and classification results are presented. As shown, the overall accuracy is about 90% for using DOG detector in counting cars and about 94% for Bus and Trucks. This system can be as an input to calibration system in multi-camera surveillance system.

Table2. Counting and classification results

Number of object matching by algorithm						
Algorithm	**Objects**			**Human**		
	Count	**Video**	**Acc.**	**Count**	**Video**	**Acc.**
DoG	61	73	83%	53	56	94%
HCD	68	73	93%	55	56	98%
CSD	67	73	91%	54	56	96%

V. CONCLUSION

In this paper we considered the problem of features matching in a camera network with overlapping fields of view. We showed that suing SIFT point descriptors in a camera network can improves the performance with respect to the other calibration systems. In particular it returned good results for scale changes, zoom and image plane rotations, and large view-

point variations. These conclusions are supported by an extensive experimental evaluation, on different scenes. 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]. Z.Fu, W.Hu and T.Tan, "Similarity Based Vehicle trajectory clustering and Anomaly Detection", In Proc. Intl. Conf. on Image Processing (ICIP'5), Vol 2, pp.602-605, 2005.
[2]. T.V.Mathew and K.V.Krishna Rao, "Introduction to Transportation Engineering." Chapter 39,'Traffic Intersections', NPTEL, 2007.
[3]. P.Babaei and M.Fathy,"Multi-Camera Systems Evaluation in Urban Traffic Surveillance versus Traditional Single-Camera Systems for Vehicles Tracking," 3rd International Conference on Computer and Electrical Engineering (ICCEE 2010), vol.4, pp.438-442, 2010.
[4]. P.Babaei and M.Fathy, "Vehicles tracking and classification using traffic zones in a hybrid scheme for intersection traffic management by smart cameras", In IEEE Conf. on Signal and Image Processing, (ICSIP 2010), pp. 49–53, 2010.
[5]. T.Bouwmans, F.E.Baf and B.Vachon, "Background modeling using mixture of Gaussians for foreground detection: a survey", In proc of Patents on Computer Science 1, pp.219–237, 2008.
[6]. M.A.Najjar, S.Ghosh and M.Bayoumi, " A Hybrid adaptive Scheme based on selective gaussian modeling for real time object detection", In IEEE Symp. Circuit and systems, pp.936-939, 2009.
[7]. S.S.Cheung and C.Kamath, "Robust techniques for background subtraction in urban traffic video," EURASIP J.Applied Signal Processing, vol. 5, pp. 2330-2340, 2005.
[8]. E.B.Ermis, P.Clarot, P.Jodoin, "Activity Based Matching in Distributed Camera Networks," IEEE Transaction on Image Processing, vol. 19, no. 10, OCT. pp.2595–2613, 2010.
[9]. D.Devarajan, Z.Cheng, and R Radke, "Calibrating distributed camera networks," Proc. IEEE, vol. 96, no. 10, pp. 1625–1639, OCT. 2008.
[10]. C.Harris and M.Stephens, "A combined corner and edge detector," in Proc. of 4th Alvey Vision Conf., pp. 147–151, 1988.
[11]. D.Lowe,"Distinctive image features from scale-invariant keypoints,"In Int. J. Comput. Vis., vol.60, no.2, pp. 91–110, 2004.
[12]. B.Song and A.R.Chowdhury,"Stochastic adaptive tracking in a camera network," in Proc.IEEE Int.Conf.Computer Vision, pp.1–8, 2007.
[13]. 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 , pp. 505–519, Mar. 2009.
[14]. D.G.Lowe, "Object recognition from local scale-invariant features," in Proc. of ICCV, pp. 1150–1157, 1999.
[15]. K.Mikolajczyk, C.Schmid,"A performance evaluation of local descriptors," in Proc. Of CVPR, 2003, pp. 257–263.
[16]. K.Mikolajczyk and C.Schmid, "Indexing based on scale invariant interest points," in Proc. Of ICCV, pp. 525-531, 2001.
[17]. T.Lindeberg, "Feature detection with automatic scale selection," in Proc. Of IJCV, vol. 30, no.2, pp.79-116, 1998.
[18]. K.Mikolajczyk, A.Zisserman,"Shape recognition with edge-based features," in Proc. of the British conf. Machine Vision 2003.
[19]. S.Belongie, J.Malik and J.Puzicha, "Shape context: A new descriptor for shape matching and object recognition," in Proc. of NIPS, pp. 831-837, 2000.
[20]. A.Baumberg,"Reliable feature matching across widely separated views," in Proc. of CVPR, pp.774– 781, 2000.
[21]. N.Allezard, M.Dhome, F.Jurie,"Recognition of 3D textured objects by mixing view-based and model based representations," in Proc. of ICPR, 2000.
[22]. J.Matas, O.Chum, M.Urban, T.Pajdla,"Robust wide baseline stereo from maximally stable external regions," in Proc. of BMVC, pp. 384–393, 2002.
[23]. K.Mikolajczyk, C.Schmid,"Scale and affine invariant interest point detectors,"In Int. Com. Vis. Vol. 60, no.1, pp. 63–86, 2004.
[24]. 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, pp.758–767, Aug. 2000.
[25]. 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, pp.56–71, Jan. 2010.
[26]. D.Makris, T.Ellis, and J.Black, "Bridging the gap between cameras," in Proc. of CVPR, vol. 2, pp. 205–210, 2004.
[27]. T.Chang and S.Gong, "Tracking multiple people with a multi-camera system," in Proc. Of IEEE Multi-Object tracking, pp. 19–26, 2001.