Individual Contour Extraction for Robust Wide Area Target Tracking in Visual Sensor Networks

Xiaoling Wu, Hoon Heo, Riaz A. Shaikh, Jinsung Cho, Oksam Chae, and Sungyoung Lee[*]
Department of Computer Engineering, Kyung Hee University, Korea
{xiaoling, riaz, sylee}@oslab.khu.ac.kr, {hhoon, chojs, oschae}@khu.ac.kr

Abstract

In this paper, we propose an approach to collaboratively track motion of a moving target in a wide area utilizing camera-equipped visual sensor networks, which are expected to play an essential role in a variety of applications such as surveillance and monitoring. A genetic fitting method for efficient contour extraction is used as inter-scene approach to detect and track the target. We also considered the existence of faulty sensors in the network which deteriorate the difficulty of target tracking problem, and proposed a robust sensor collaboration method. The experimental results have shown that the proposed target tracking approach produces very successful target tracking compared with the existing method especially in case that the target is adjacent to neighboring objects of background.

1. Introduction

All Intelligent surveillance or monitoring is an emerging area for sensor networks application. An important task to be performed by a network of visual sensors (e.g. camera) or multimode sensors (e.g. camera with microphone and thermometer) distributed in a geographic area is to track target in a local area in order to monitor unusual activities. We use visual sensors and multimode sensors other than one single type of sensor because with the increase of computing power of hardware it becomes more feasible to utilize richer representations of features other than that used in current Decentralized Data Fusion algorithms. A lot of research works have been done based on common wireless sensor networks (WSNs) equipped with only acoustic sensors or temperature sensors, and they work well in ideal environment without much interference or noise. In case of real environment, however, the noise or uncertainty often has bad effect on target tracking results.

Before tracking, the sensors close to the predicted path of the target need to be alerted. The target can be a moving vehicle or can be a phenomenon such as an approaching fire. It is assumed that each individual sensor node is equipped with multimode sensory devices in order to detect the target based on the sensed data. The sensors that are triggered by the target collaborate to localize the target in the physical space [1]. One of the central issues for collaborative signal and information processing to be addressed is energy constrained dynamic sensor collaboration: how to dynamically determine who should sense, what needs to be sensed, and who the information must be passed on to.

Target tracking in visual sensor networks (VSNs) is a challenging problem that requires acquiring and processing data from multiple camera views (a single camera often cannot see every object). Besides, any practical sensing device has limitations on its sensing capabilities (eg. resolution, bandwidth, efficiency, etc.). The descriptions or physical models built on the data sensed by a device are, unavoidably, only approximations of the real nature. These uncertainties, coupled with the reality of occasional sensor failure greatly compromise reliability and reduce confidence in sensor measurements. In addition, the spatial or physical limitations of sensor devices often mean that only partial information can be provided by a single sensor. Thus, target tracking process should be robust because the existence of faulty sensors deteriorates the difficulty of target tracking problem. False alarms waste network resource as well as cause other negative effect. In WSN environment, the traditional double or triple redundancies are not adequate solutions due to their power consumption, space, and cost. We attempt to consider qualitatively fault tolerance in sensor collaboration using a simple but efficient method.

---

[*] Corresponding author.

IEEE
COMPUTER
SOCIETY

In case of existence of multiple objects including both target and uninterested background objects, which also complicates tracking problem, genetic fitting is proposed to distinguish them especially when the target is closely located with other object (s).

All in all, in this paper, our contribution is that we considered fault tolerant sensor collaboration in target tracking process by a low computation cost method in wireless VSN environment and a B-spline contour fitting approach based on genetic algorithm (GA) for efficient contour extraction is proposed as inter-scene vision method to detect the target. It produces accurate detection and tracking especially when the interesting target is closely located beside other objects. Compared with the existing classic contour extraction method B-spline, and existing graph matching method also based on VSNs, our method shows significant improvement in terms of success rate of target tracking.

This paper is organized as follows: related work and VSNs environment are sketched in section 2 and section 3, respectively. The method to extract individual contour is proposed in section 4, followed by experimental results in section 5. Some concluding remarks and future work are provided in section 6.

2. Related work

The problems of target detection and tracking have been explored in [2] on an individual node basis. There is, however, little research on distributed detection and tracking within WSNs. Topics of target tracking have been studied and developed extensively but primarily in the domain of active and passive radar. Graphical modeling techniques such as Kalman filtering and Hidden Markov Models [3, 4] have been employed successfully in this domain. Complex multiple hypothesis testing techniques are incorporated into their frameworks that rigorously evaluate every possible origin of the measurements received. However, they assume that all the measurements are available for processing at a centralized node. In [5], the authors proposed a distributed energy-aware collaborative target-tracking algorithm using Kalman filtering technique. However, their algorithm was only suitable for one single target case and it can move solely in a straight line. It is not practical in most cases as multiple targets may exist and they often move in at least a 2-D space. Another drawback is that they choose initial active sensor randomly for sensing information. It cannot be energy efficient and fault tolerant when the active sensor is located farthest away from the target. In [2], the authors used a classification algorithm to disambiguate closely located targets.

However signals received from targets are correlated and we cannot recover the uncorrelated signals always. Since we do not know in advance the number of targets around each sensor, the problem is ill-posed and very challenging even for a highend computer.

There are also some researches for target tracking based on VSN environment. In [13], the authors proposed a real remote monitoring system for all day outdoor observation using wireless communication, but the difference with ours is that this system was used for small scope area target tracking because it has only one camera equipped. If equipping multiple cameras for wide area tracking, the collaboration among sensors and cameras should be reconsidered. G. Kogut, etc [6] established wide-area camera network with non-overlapping fields of view in which a wide-area tracking algorithm is tested to track moving objects. This algorithm employs a variety of vehicle features in a feature vector: color, shape, and velocity. The color features are calculated by using partial implementation of the AutoColor Matching System to compensate for differences between illumination at cameras sites and between cameras. The tracking algorithm uses the color model and blob centroids from the segmentation module to help solve the data association problem. However their simulation results show that the graph matching method is not accurate enough for target tracking.

For fault tolerant target detection or tracking, Clouqueur, etc [7] seek algorithms to collaboratively detect a target region. Each sensor obtains the target energy (or local decision) from other sensors, drops extreme values if faulty sensors exist, computes the average, and then compares it with a pre-determined threshold for final decisions. For these algorithms, the challenge is the determination of the number of extreme values. This is unavoidable when using "mean" for data aggregation. In [8], the authors explored the utilization of "median" to effectively filter out extreme values for target region detection and claimed their algorithm more computationally inexpensive than [7]. However only readings of neighbors are included for computing the median value, and a sensor S_i is determined as an event sensor if the estimated value is larger than a predefined threshold. The drawback of this algorithm is that they omit the case when S_i is just the faulty sensing node.

In this paper, we attempt to solve the target-tracking problem on VSN platform using fault tolerant sensor collaboration, and genetic fitting method as in-network data fusion, especially when the target is closely located with other object (s).
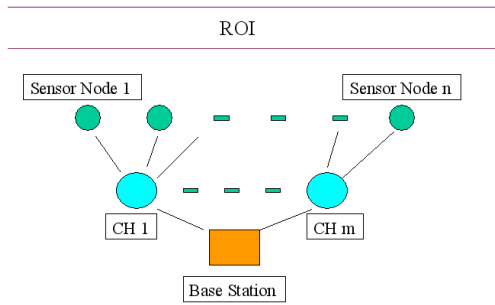
3. Visual sensor networks: architecture

Figure 1. Visual sensor networks architecture for target detection

We envision a system that covers a remote area with wireless sensors, cluster heads (CHs) [9] and base station as shown in Fig. 1, which has nearly uniform sensor distribution with optimal coverage and optimal CHs positions to reduce transmission energy consumption proposed by our previous work [10]. We assume that each sensor has a camera and other simple sensing devices equipped. Sensor nodes communicate with their CHs, and CHs transmit compressed image data to the base station, which is located far away. CH usually works based on an event driven method, i.e., it works according to the wireless sensor's signal indicating the existence of target. Only sensors positioned in the interested field of view and among them only those who have robust data, can be selected to send images to CH. In this way, large amount of energy can be saved. Fusion of information from different sensors would allow for tracking the interesting target. In addition to triggering appropriate responses, results from such an analysis would be stored in a database of base station [11]. This would allow statistical analysis of past events. Having multimode sensors provides the ability to reduce the probability of detecting some types of fault information, and to provide more information than with a single camera.

The use of multimode-sensor architectures guarantees a proper coverage of all possible conditions of operation, thus satisfying the desired requirements in terms of system's global performance.

A. Image acquisition

In the image acquisition stage, the input to each single-sensor node is a stream of raw images. There are no requirements of the properties of the images. However the quality of the acquired data may affect the overall performance and quality of the processing steps. The rate of image acquisition may also influence processing results, depending on the velocity of the moving objects being tracked, and the perspective by which the objects are being viewed. Sensor location tends to be a contexts-dependent problem, determined largely by the specific application requirements.

B. Sensor Collaboration

Deploying VSNs aims to cover a wider area than possible with a single camera. Viewing the same object from different positions has two main advantages [12]. Firstly, event detection can be made more robust by cross-validating information. For example, if two camera-equipped sensors in a surveillance system see at the same time a moving object approximately located in the same position in space, then the confidence of this event detection increases. Secondly, range information can be estimated by triangulation, allowing for a very powerful geometric scene description.

For all kinds of weather and darkness, the camera can acquire images using the functions of auto focus, auto iris, lens zooming, and night vision. The camera can automatically change into an infrared one, depending on the lightness [13].

To achieve robust sensor collaboration, we proposed four steps fault tolerant method:

(1) Discover neighbors within cluster range.

(2) At every node in the cluster, compute the "median" value among the readings of node i and all its neighbors for eliminating the extreme values. This is different from average method [7] which firstly drops extreme values and then computes the average.

(3) Choose those whose readings are above a particular threshold (e.g., "temperature is higher than A", "acoustic signal intensity is higher than B") as the event nodes

(4) The chosen sensor nodes indicate the presence of a target in the vicinity. Then, binary messages are sent by the chosen nodes to CH. When such a message is received by CH, it demands the image data to be sent from the originator.

If a faulty sensor constantly reports wrong information (obviously different from other sensors), it should finally be excluded from the sensor network. This proposed fault tolerant idea has low computation and communication overheads compared with related fault tolerant methods, by analysis in a qualitative way.

C. Inter-scene tracking

Inter-scene tracking is the fusion of tracking data received from multiple individual sensors to allow target to be tracked as they move along nodes with or

without overlapping fields of view. This type of tracking is impossible with data from single nodes working on independently. The input of the inter-scene tracking data is the pre-processed output from the intra-scene tracking process [14]. Suitable communication protocol is assumed in the VSN, so the detailed design of novel protocol is out of the scope of this paper.

4. Object tracking by extracting individual contour

We propose the individual contour extraction approach based on GA in the inter-scene tracking stage. GA is a probabilistic technique for searching an optimal solution. The optimal solution is described by the value of vector X, which is called as the "chromosome" in GA, and it can be obtained by minimizing an objective function $f(X)$. Hence, the definition of the objective function significantly affects the solution state X. In the proposed algorithm, a chromosome consists of n control points of B-spline, which is similar to the chromosome design of MacEachern [15]. Since the chromosome represents a complete contour and a gene uses the actual location of a control point, the search algorithm has neither ambiguity on the contour location nor potential bias to particular shapes. To reduce the size of a gene, we use the index value as a gene, instead of two coordinate values. The chromosome representation using the indices enables us to use the search area with any shape and produce shorter chromosomes, which generally result in faster convergence rate.
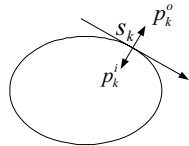

Figure 2. Definition of inner and outer regions

The fitting process to extract a contour starts with the generation of the initial population. The initial population consists of a set of chromosomes selected randomly from the search windows for control points. A new generation is evolved through evolutionary operations. The end of the evolutionary operation is determined by checking the fitness values, which represent the goodness of each chromosome in the population [16]. In this paper, we introduce a fitting function that rejects the contours of the nearby object specially designed for the contour extraction in interesting target image sequences.

To compute the fitting value for a possible solution, we first generate the contour points of B-spline and trace the contour as shown in Figure 2. At the k'th contour point $\mathbf{r}(s_k)$, compute a normal vector $\mathbf{n}(s_k)$, and identify the inner region and outer region pixel location $\mathbf{p}_k^i$ and $\mathbf{p}_k^o$ from the curve by using equation (1) and (2).

$$\mathbf{p}_k^o = \mathbf{r}(s_k) + \mathbf{n}(s_k) \qquad (1)$$

$$\mathbf{p}_k^i = \mathbf{r}(s_k) - \mathbf{n}(s_k) \qquad (2)$$

The fitting value can be determined based on gradient magnitude information at each contour point by using the equation (3) and (4).

$$f(\mathbf{x}) = \left| \sum_{k=0}^{M-1} grad_k \right| \qquad (3)$$

$$grad_k = \begin{cases} \left| \nabla I(\mathbf{r}(s_k)) \right| & I(\mathbf{p}_k^i) - I(\mathbf{p}_k^o) > 0 \\ -\left| \nabla I(\mathbf{r}(s_k)) \right| & I(\mathbf{p}_k^i) - I(\mathbf{p}_k^o) \le 0 \end{cases} \qquad (4)$$

where $I(\mathbf{p}_k^i)$ and $I(\mathbf{p}_k^o)$ are grayscale values of the inside and outside of the k'th contour point. Hence, by judging the sign of the intensity difference, the proposed fitting function can subtract the gradient magnitude values of the contour pixel that lies on the neighboring object boundary from the accumulated value, to avoid detections of other objects from background.

5. Individual contour extraction: experimental Results

Using the VSN architecture aforementioned in section 3, we tested the proposed contour segmentation method with two sequences of images in each of which the moving target of interest passes by another adjacent static object. The sensors switch working by turn. Only event sensors selected by fault tolerance method send images to CHs. After receiving several image sequences from sensors, CH will take the responsibility to execute the proposed algorithm to process the image sequence. The test data are prepared to reveal the capability of the proposed algorithm in finding an accurate boundary among similar objects nearby if there are any existing. To generate the results, we construct a B-spline contour with 8 control points and select 20 initial solutions for each 30x30 window.

Fig. 3 (a) and (b) are tracking results with contour marked as red for 2 image frame sequences of a moving car in the Region of Interest (ROI). Each 3 images of (a) and (b) are taken by 3 event sensors selected by proposed fault tolerant sensor collaboration method respectively. As shown in Fig. 3, the proposed algorithm can successfully track the moving target of interest under unfavorable conditions, for example, with some other background objects nearby. Although we tested this algorithm only in case of single target, it is also applicable to multiple targets case.

Table 1 lists parts of numerical results of the segmentation algorithm for each image of the test set. FPE (False Positive Error) is the percent of area reported as a target by the algorithm, but not by manual segmentation. FNE (False Negative Error) is the percent of area reported by manual segmentation, but not by the algorithm. Similarity and dissimilarity indices [16], which show the amounts of agreement and disagreement between the area of the algorithm and manual segmentation, are computed by:

$$S_{agr} = 2\frac{A_{man} \cap A_{alg}}{A_{man} + A_{alg}} \ , \ S_{dis} = 2\frac{A_{man} \cup A_{alg} - A_{man} \cap A_{alg}}{A_{man} + A_{alg}} \tag{5}$$

In this expression, A_{man} and A_{alg} are the sets of pixels classified as measurement area of manual segmentation and the algorithm, respectively. These indices are calculated for validation on every image along the movement of target. Values computed are shown in Table 1 and we conclude that proposed method for segmentation isolates individual region of target successfully.

Table 1. Segmentation results for 8 images

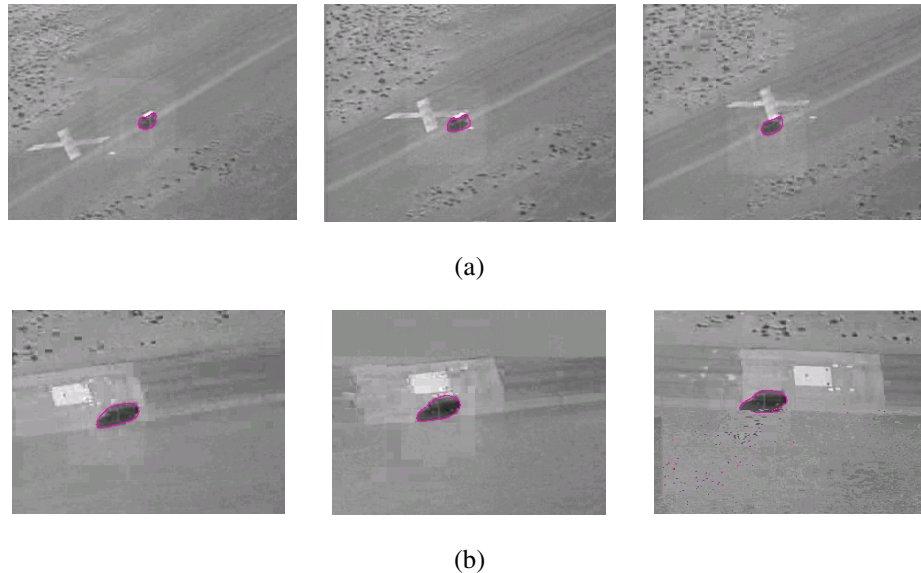| Image Number	FPE [%]	FNE [%]	S_{agr}	S_{dis}
1	3.87	3.45	0.965	0.161
2	7.76	4.49	0.947	0.122
3	6.46	8.47	0.896	0.218
4	2.81	8.29	0.909	0.198
5	2.47	7.89	0.933	0.124
6	4.21	6.79	0.935	0.089
7	4.67	1.42	0.975	0.057
8	3.73	6.57	0.955	0.122

(a)

(b)

Figure 3. Tracking results of the proposed method for a sequence of interesting target images (a) tracking result for the 1st original image frame sequence of a moving car in one area (b) tracking result for the 2nd original image frame sequence of the moving car in another area
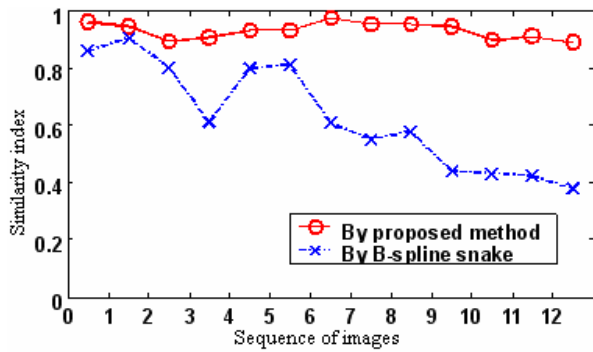
Figure 4. Comparison between proposed method and B-spline snake on the similarity index values at each picture

As the segmentation is performed piece by piece, in contrast with the result of proposed method, mal-fitting error contained in results of existing classic contour extraction method increases. Figure 4 shows the test result in terms of the similarity index, S_{agr} along with the sequence of pictures. The plots are made using the proposed segmentation algorithm, and B-spline snake respectively. It shows that B-spline snake algorithm fails when the number of pictures containing adjacent objects increases while the proposed algorithm detects the accurate contours.

We also compared with the experimental results of a wide area target tracking system in a VSN proposed in [6] as shown in Table 2, which has similar assumptions with ours. Their algorithm success rate was defined versus a manually tracked ground truth. The trial data consisted of several data sets: the deliberate running of an electric cart along various paths through the sensor network, the recording of full-sized vehicles traveling through the network, and the recording and tracking of pedestrians in the areas which adjacent vehicle and pedestrian traffic. The experiments involved simultaneous calculations on five video streams, and ran on three separate computers. Comparing our results shown in Table 1 with the results of [6] shown in Table 2, we can see clearly the significantly higher success rate of our algorithm which is represented by S_{agr}. We didn't compare our results with any other tracking approaches which are based on common WSN equipped with only acoustic sensors or temperature sensors, although many of their algorithms show successful tracking results. The reason is our totally different assumptions. Their approaches work well in ideal environment without much interference or noise. In case of real environment, however, the noise or uncertainty always has negative influences on target tracking results. With

visual sensors in addition to other types of sensors, we can get richer information from the target.

Table 2. Experiment results of graph matching in visual sensor network platform [6]

Data Set	#Nodes Entered	Mean Platoon Size	#Successful Reident.	Success Rate%
Electric Cart	18	1.0	11	61%
Full-size Vehicles	27	1.3	14	52%
Pedestrians	33	2.8	19	58%

6. Conclusions and future work

In VSN environment, fault tolerant sensor collaboration in target tracking process by a low computation cost method has been considered. A B-spline contour fitting approach based on GA for efficient contour extraction is used as inter-scene vision method to detect the target. Our proposed method produces accurate detection and tracking especially when the interesting target is closely located beside other objects. Compared with the existing classic contour extraction method B-spline and graph matching based on VSNs, our method shows significant improvement in terms of success rate of target detection and tracking. Although we have used target tracking as an example to carry out the discussions, the method of individual contour extraction is quite general. Our future work includes accurate target position identification, image data compression and more flexible architectural design.

Acknowledgement

References

[1] Rahul Gupta and Samir R. Da, "Tracking Moving Targets in a Smart Sensor Network", *Proc. of Vehicular Technology Conference*, Vol.5, 2003, pp. 3035-3039.

[2] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, Classification and Tracking of Targets", *IEEE Signal Processing Magazine*, Vol. 19, 2002, pp. 17-29

[3] F. Maainerie, "Data Fusion and Tracking Using HMMs in a Distributed Sensor Network", *IEEE Trans on Aerospace and Electronics System*. Vol. 33, No. 1, 1997, pp. 11 - 28.

[4] D. Reid, "An Algorithm for Tracking Multiple Targets", *IEEE Transactions on Automatic Control*, Vol, 24, No. 1 ,1979, pp. 843- 854.

[5] S. Balasubramanian, I. Elangovan, S. K. Jayaweera and K. R. Namuduri, "Distributed and Collaborative Tracking for Energy-Constrained Ad-hoc Wireless Sensor Networks", *WCNC/ IEEE Communications Society,* 2004, pp. 1732-1737.

[6] G. Kogut and M. Trivedi, "A Wide Area Tracking System for Vision Sensor Networks", *9th World Congress on Intelligent Transport Systems*, Chicalgo, Illinois, 2002.

[7] T. Clouqueur, K.K. Saluja, and P. Ramanathan, "Fault Tolerance in Collaborative Sensor Networks for Target Detection", *IEEE Transactions on Computers*, Vol. 53, No. 3, 2004, pp. 320-333.

[8] Min Ding, Dechang Chen, Andrew Thaeler and Xiuzhen Cheng, "Fault-Tolerant Target Detection in Sensor Networks", *Wireless Communications and Networking Conference*, IEEE Vol. 4, 2005, pp. 2362-2368.

[9] Wendi B. Heinzelman, Anantha P. Chandrakasan, and Hari Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks", *IEEE Transactions on Wireless Communications*, Vol. 1, No. 4, 2002, pp. 660-670.

[10] Wu Xiaoling, Shu Lei, Yang Jie, Xu Hui, Jinsung Cho and Sungyoung Lee, "Swarm Based Sensor Deployment Optimization in Ad hoc Sensor Networks", *ICESS/ LNCS (SCIE)*, Springer, 2005, pp. 533-541.

[11] Mohan M. Trivedi, Ivana Mikie, and Greg Kogut, "Distributed Video Networks for Incident Detection and Management", *Proceedings of IEEE ITSC*, 2000, pp. 155-160.

[12] K. Obraczka, R. Manduchi, and J. J. Garcia-Luna-Aveces, "Managing the information flow in visual sensor networks", *Int. Symp. Wireless Personal Multimedia Communications*. Vol. 3, 2002, pp. 1177-1181.

[13] Shinichi Masuda and Tetsuo Hattori, "Flexibly Configurable Multivision Remote Monitoring System", *The 47th IEEE International Midwest Symposium on Circuits and Systems*, 2004, pp. 301-304.

[14] Greg T. Kogut, Mohan M. Trivedi, "Real-time wide area tracking: Hardware and software infrastructure", *Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems*, 2002, pp. 587-593.

[15] L. A. MacEachern, T. Manku, "Genetic algorithms for active contour optimization", *IEEE Int. Sym. for Circuits and Systems*, 1998, pp. 229-232.

[16] Hoon Heo, M. Julius Hossain, JeongHeon Lee, OkSam Chae, "Visualization of Tooth for 3-D Simulation", *Asia Simulation Conference, Proceedings of LNCS (SCIE)*. Vol.3398, 2004, pp. 645-684.