Robust Tracking Algorithm of Multiple Objects Under Dynamic Environment

Sung-Wook Seol[1], Jee Hye Jang[1], Hyo-Sung Kim[1], Ki-Gon Nam[1], Chul-Hun Lee[2]
[1]Department of Electronics Engineering, Pusan National University, Pusan, Korea,
[2]Department of Electronics, Pusan Info-Tech College, Pusan, Korea
susul@vision3d.ee.pusan.ac.kr

Abstract

In this paper, we present the motion estimation approach based on double difference to detect and track the moving objects in the sequence of images under the moving camera. The double difference method operates on the two consecutive difference images with the logical AND operation. It accumulates the information on edge region and eliminates the information on un-textured region from the sequence images. The candidate regions for the moving objects are the blocks containing the accumulated information on edge region. The candidate block region has both objects and background. Therefore, we use gradient feature to remove the background region. The key points are initialized as the grid point over objects region. The initialized key point is matched with some frames and determined candidate blocks as objects. The final candidate blocks are grouped by the clustering algorithm like ISODATA. The region of moving object is detected by XY-projection in the marginal searching window. That is we match and track the detected object using auto-correlogram. The experimental results have shown that the proposed algorithm is reliable and can detect and track successfully the moving objects in the image sequences obtained by the moving camera.

Keywords: tracking, detection, moving camera, double difference, multiple vehicles

1 Introduction

The study on the tracking of the moving objects is one of the most major branch in computer vision and other several fields. Especially, the tracking of the moving object under the moving camera is extremely active field of research in the computer vision. The tracking algorithm of the moving objects is applied to object recognition, robotics, surveillance system, and automatic navigation system. The conventional approach to track of the moving object is the difference image [1], optical [2], geometric transform [3], and camera motion information [4]. The difference image method has a weak point to apply to the moving camera. The optical flow method has a large amount of the data and time to search the motion vector [5]. The geometric method and the camera motion information method have some difficult to extract the unified information from the background.

Moreover, the several systems mainly deal with the tracking process given the detected objects. Those systems need to pre-process that detects the position of the moving objects before tracking them. That is, the conventional systems consist of two parts; the detection and track of the moving objects.

However, our proposed system makes a contribution to remove this pre-processing and unify two processes; detection and track.

Therefore, we present the double difference based motion estimation algorithm to integrate the detection and tracking of the moving objects. Using proposed algorithm, we construct the tracking system of the moving objects under the environment of a moving camera. After initial transient states over a few frames, the proposed system transits the stable state that can detect and track the moving objects at the same time. The proposed system can deal with detection and tracking efficiently of the new object that appear in a tracking process. Experimental results have shown that the proposed algorithm is reliable and can successfully detect and tract the moving objects in the sequence of images obtained by the moving camera

2 The Proposed Tracking System

The block diagram of the system is shown in figure 1.

The proposed system converts the spatio-temporal information on the moving objects to the double difference image using the intensity information on a current frame **I** and a previous **I'**.

The key point is extracted in the region with maximal contour feature of the region that occurred motion in double difference image. The regions extracted key points are compared with the key points of the previous frame by block matching. The matched key points are growing the region of objects via clustering. Tracking is performed by association of the relation the object of previous frame with the tracking object of the current frame within assigned regions continuously.

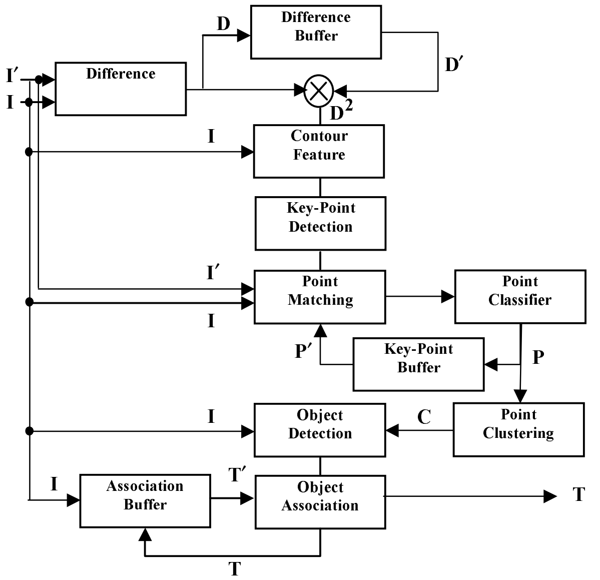


Figure 1: The proposed tracking system.

2.1 The Double Difference and Contour Features

The image obtained in the physical environment contains a lot of the un-textured background like the sky and a road. The double difference method operates in three consecutive image frames. This method is proposed and applied under fixed camera environment [6, 7]. However we prove that it is excellent under moving camera in our experiment. It can use to eliminate the un-textured region like background and detects the textured region of the moving objects. The difference image is to subtract the current frame image with previous frame by pixel unit. That is, the double difference value is very high in the region where take place some spatial and temporal motion. Even though there is an object, if it hasn't any motion, the object can't be detected. The difference image **D** is represented by

$$\mathbf{D} = \begin{cases} 1 & if \ |\mathbf{I} - \mathbf{I}'| > T \\ 0 & otherwise \end{cases}, \tag{1}$$

where T is the threshold value, eg. the mean value of the difference image. The double difference image is obtained by the logical AND operation with the two consecutive difference images. The double difference $\mathbf{D}^2$ is represented by

$$\mathbf{D}^2 = \mathbf{D} \cap \mathbf{D}', \tag{2}$$

where **D** and **D'** are difference images obtained by current and previous frame, respectively.

We extract the region of moving objects through double difference method. The extracted regions contain both objects and background region. Therefore, it has to eliminate the region with the edge of a background, and to select the region with the edge of objects. For this, the contour feature of the objects is calculated in the moving region where extracted from double difference image method. After masking the current image with the moving region obtained from double difference method, the contour feature is calculated within the region where the motion is occurred. The contour feature is calculated from image **F** that is masked current image with double difference image [8]. The contour feature is represented by

$$\mathbf{G} = \begin{bmatrix} \mathbf{F}_x^2 & \mathbf{F}_x\mathbf{F}_y \\ \mathbf{F}_x\mathbf{F}_y & \mathbf{F}_y^2 \end{bmatrix}, \tag{3}$$

using the gradient feature with intensity of masked image **F**. In every pixel which motion is detected via the double difference method in current frame, **G** value is calculated. That is, the pixels ordered by maximal contour feature are selected as contour feature points because the greater minimum eigenvalue are the more strong contour features. The minimum eigenvalue is

$$\lambda = \min\{\lambda \mid \det(\mathbf{G} - \lambda\mathbf{I}) = 0\}. \tag{4}$$

As in order of the contour features, i.e. in order of λ, larger λ is used as the contour features.

2.2 The Key points Detection and Classification

The key points are initiated to the grid point that has an fixed length in the image coordinate as shown in the figure 2. The searching window of W x W size is established on the initial key points. If the searching window has contour features more than threshold number, it is registered as a key point block. If not, the window is removed from key point.

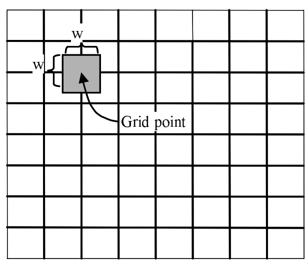


Figure 2: Set the initial key points.

The registered key point blocks determine the matching point in current frame using block matching with a registered blocks in previous frame [9]. As the index of a matching between two blocks, sum of the absolutely difference (SAD), sum of the squared difference (SSD) and normalized cross-correlation coefficient (NCC) are generally used. The NCC offers the most exact result among them, but it has large amount of computation. So, the SAD is used as the index of a matching because it has less mount of computation than others, and it is convenient to make up with hardware.

The matching process using SAD is searching that matching point that has minimum absolute error by comparing the key point in current frame with the matched point in previous frame. The matching point in current frame calculated as matching point with previous frame is a (x_i,y_i) that obtained by

$$\arg\min_{i} \sum_{(u,v)\in WxW} \left| I(x_i+u, y_i+v) - I'(x'+u, y'+v) \right|, \quad (5)$$

where WxW is a searching window, (x',y') is matched point in previous frame, (x_i,y_i) is the key point in current frame that is central in the searching region.

To track the moving object with a moving camera, the tracked object should have a small displacement and the background will move extremely in the image coordinates. That is, in matching process with SAD, the distance d_i is length over two points between (x_i,y_i) and (x',y'). When d_i is smaller then threshold value T_d, the block is considered as a object region. If not, the block can be considered as a background.

Therefore, the key points in current frame is classified as two categories; the region of an object and a background. If the region of an object is classified as a background in previous frame, it is necessary that this region is saved until next frame, because it can be classified as the region of an object in next frame. In case of classification as a background consecutively with two frames, it is removed perfectly among key points.

The operating form of the matched points and key points are classified with six cases, and it is shown in the figure 3.

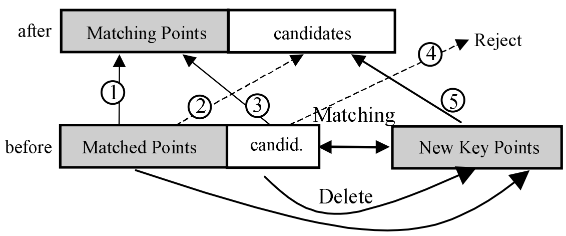


Figure 3: The matching and classifying operation.

When the tracked points in previous frame are mismatched, they are classified as a background region. In the next frame, the block matching is operated again. If block matching is continuously failed twice, that block is removed away because it is considered as a vanished object.

2.3 Clustering and Object Detection

The detected points as a moving object is classified some small groups via clustering algorithm. The typical models of the clustering algorithm are the simple cluster searching algorithm, the min-max distance algorithm, the k-means algorithm, the ISODATA algorithm and the LVQ algorithm etc. In this paper, we use the ISODATA algorithm to cluster images.

The region of the moving object is detected using each matched points that grouped by clustering algorithm because those points are in the moving objects. To detect the region of the moving objects, the searching window is set with some margin in the rectangular region surrounding the last shell of the clustered points. To extract exactly the region of the moving objects in searching window, the XY-projection is operated what horizontal and vertical edges is accumulated to each direction in the searching window [10]. The edge profile projected horizontal and vertical direction is checked from each border. If any point is larger than threshold value, there is determined an edge of a moving objects. Thus, the two points obtained from association of a XY-projection make up with the smallest rectangular region contained moving objects. The information of a moving object is extracted in that rectangular region from input image.

2.4 The Moving Objects Association

It is necessary to inspect that the extracted object in current frame is the same as tracked objects in previous frame. If the extracted object in current frame is associated with tracked any object in previous frame, the object tracking is performed continuously. However the tracked object in previous is not associated with anything in present frame, the tracked object should be removed. And a new object is appeared that is not associated with any moving

Table 1 : Operating explanation of the image buffer in the Fig. 3.

no.	Operation	Explanation
①	Matching	The points of the object in previous frame are classified as a objects continuously
②	Miss-matching	The matched points in previous frame are classified as a background region in current frame
③	Matching	The region of the objects was classified as background and next time it is classified as a region of objects
④	Miss-matching	It is removed because of failure of a block matching continuously
⑤	Matching	The new key points in current frame are classified as the region of an object

objects in previous, it should be registered and tracked continuously.

The index of a similarity that is associated with objects require some conditions. First of all, it must recognize the object robustly regardless of rotation, translation, and variable scale change of the objects. Next, it is robust for the partial occlusion of the moving objects. The correlogram is known that satisfy all of these conditions. The correlogram is used as an index that represent the similarities of tracked object in previous frame with detected object in current frame.

The correlogram is a two dimensional probability of an intensity distribution among pixel that have fixed distance in the region of an object. The intensity in the n x n sized region on the object is quantized to m levels. In given any pixel which has intensity c_i, $\Gamma^{(k)}_{c_i,c_j}$ are a pixel numbers that intensity is c_j apart from k distance from given pixel. The auto-correlogram value of (k, c_i) is

$$\gamma^{(k)}_{c_i} = \frac{\Gamma^{(k)}_{c_i,c_i}}{8k \times h_{c_i}}, \qquad (6)$$

where h_{ci} is the histogram of the c_i intensity. The total auto-correlogram is composed of two-dimensional table with k x m. The region of the objects obtained in the current frame and in the previous frame is converted to auto-correlogram. The object that the smallest SAD of the auto-correlogram about the region of the objects in the previous and current frame is associated each other. That is, the object which has the smallest value of

$$D = \arg\min_{1 \le n \le N} \{\sum_{j=1}^{k} \sum_{i=0}^{m} [\mathbf{A'}(i,j) - \mathbf{A}_n(i,j)]^2\}, \qquad (7)$$

is associated each other, where **A'** is an auto-correlogram of any object in previous frame, and $\mathbf{A}_n$ is an auto-correlogram of the n-*th* object in current frame.

3 The Experimental results

The image used in the experiment is obtained by fixed camera that is setting up the 2.3m height of the vehicle and the size of image is 256 x 256 pixels, the intensity levels of image are 8bit grey. The vehicle mounted our system is moving forward in the general road and the sequence of images is acquired by CCD camera. The acquired images are sampled by Matrox Meteor-Ⅰ/MC Frame Grabber of Matrox Company.

For the real-time tracking system, the required time within algorithm is less than time interval between two continuous frames. The average operating time of the algorithm within this system is 16ms. So, the real-time operation system can be realized.

The figure 4 shows detection and tracking processes of the moving object picture obtained from other researches. The images show a current image, double difference image, contour feature image, key point detected image, clustered image and detected and tracked image, respectively.

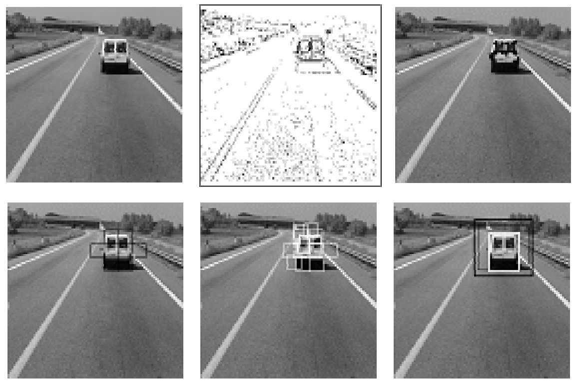
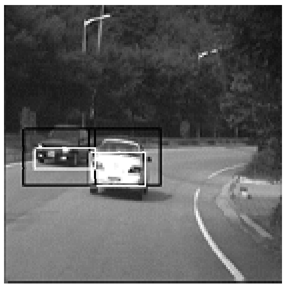
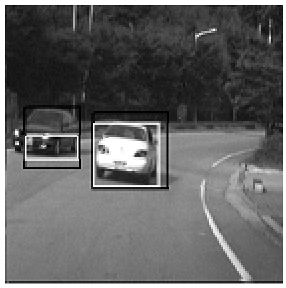


Figure 4: The detection and tracking process.

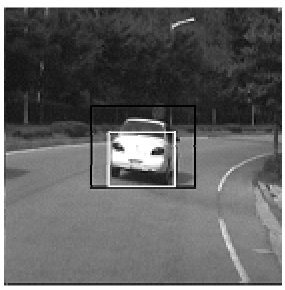
The experimental result of the moving object that is entering into the ramp of the highway is shown in the figure 5. It is shown that the algorithm is reliable rotational and scale change of the moving objects.



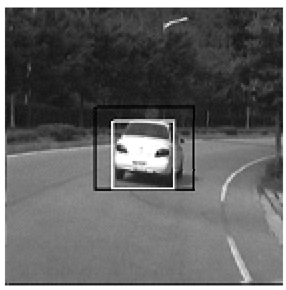
(a)



(b)



(c)

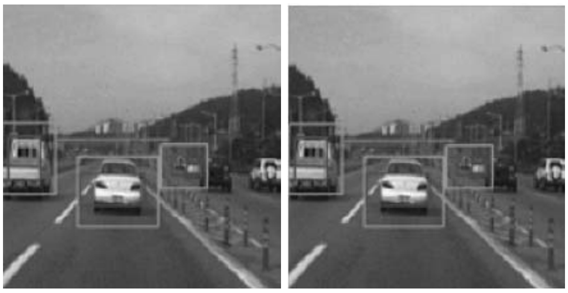


(d)

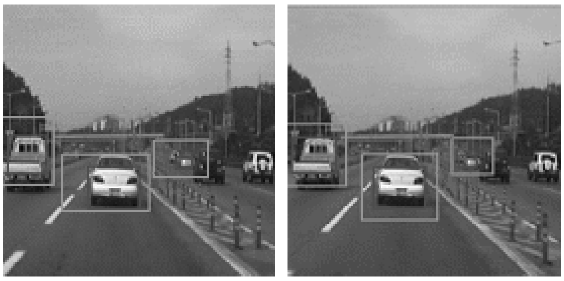
(d) (f)

Figure 5: The simulations for tracking.

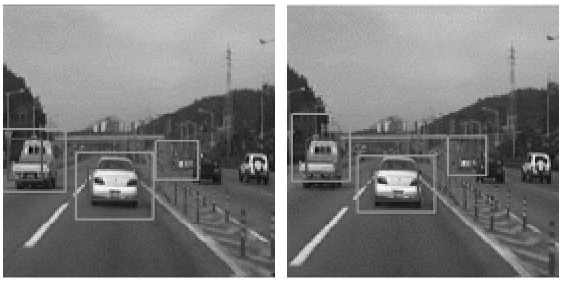
Also the experimental results show that the tracking of multiple objects in the figure 6.



(a) (b)



(c) (d)



(e) (f)

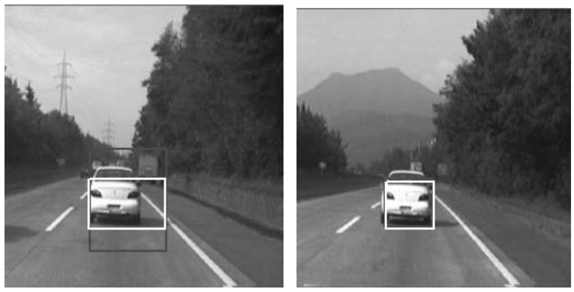
Figure 6: The simulations for tracking of multiple objects.

The table 1 shows the processing time of our system that is detecting and tracking the moving object in the sequence of images to be shown in the figure 7. The processing time is averaged by experiments of several times.

Table 2 : Processing Time during Detection of the Moving Object.

	Fig.7(a)	Fig.7(b)	Fig.7(c)	Fig.7(d)
Processing Time per Frame [*ms*]	16	15	15	16

(a) (b)



(c) (d)

Figure 7: Some variable images to check up processing time of the proposed algorithm.

The frame rates are 30. It means that the time interval between two images is about 33ms. It is proved that the system is reliable at real time tracking in an outdoor environment.

4 Conclusion

In this paper, we have proposed motion estimation approach based on the double difference to detect the moving objects in the sequence of images under the moving camera.

The experimental results have shown that the propose algorithm is reliable and can detect and track the moving objects successfully in the image sequences obtained by the moving camera.

5 References

[1] R. Jain, D. Militzer and H. Nagel, "Separating non-stationary from stationary scene components in a sequence of real world TV images," *Proc. Fifth Int. Joint Conf. Artif. Intell.*, pp. 425-428, 1977.

[2] T. Suzuki and T. Kanade, "Measurement of vehicle motion and orientation using optical flow," *IEEE Conf. On Intelligent Transportation systems*, Tokyo, Japan, Oct. 1999.

[3] S. Araki, T. Matsuoaka, N. Yokoya and H. Takemura, “Real-time tracking of multiple moving object contours in a moving camera image sequences,” *IEICE Trans. Inf. & syst.* vol. E83-D, no.7, Jul. 2001.

[4] D. Murray and A. Basu, “Motion tracking with an active camera,” *IEEE Trans. Pattern Anal. & Mach. Intell.,* vol. 16, no. 5, pp. 449-459, 1995.

[5] S. S. Beauchemin and J. L. Barron, "The Computation of Optical Flow," *ACM Computing Surveys*, vol. 27, no. 3, pp. 433-467, 1995.

[6] Q. Cai, A. Mitiche and K. Aggarwal, “Tracking Human Motion in an Indoor Environment,” *Proceedings of the 2nd International Conference on Image Processing*, pp. 215-218, 1995.

[7] R. Cucchiara, M. Piccardi, A. Parati and N. Scarabottolo, “Real-time Detection of Moving Vehicles,” *Proc. Of 10th ICIAP,* Venice, Italy, pp. 618-623, 1999.

[8] D. A. Forsyth and J. Ponce, *Computer Vision A Modern Approach*, Prentice Hall, 2003.

[9] Feng, K. T. Lo, H. Mehrpour, A. E. Karbowiak, "Adaptive block matching motion estimation algorithm for video coding," *Electronics letters*, vol. 32, no. 18, pp. 1542-1543, Aug. 1995.

[10] S. U. Seol, H. B. Lee, H. S. Kim, K. G. Nam, C. H. Lee, “Algorithm for Moving Object Tracking from Moving Camera Using Histogram Projection,” *Journal of The Institute of Signal Processing and Systems,* vol. 4, no. 2, pp. 38-45, Oct. 2001.