Moving Object Detection and Localization Using Stereo Vision System

Bogdan Żak and Stanisław Hożyń

Polish Naval Academy, Gdynia, Poland
{b.zak,s.hozyn}@amw.gdynia.pl

Abstract. The aim of this study was to design an moving object detection and localization algorithm able to detect and localize especially humans, vehicles and planes. We focused on classical methods for cameras calibration and triangulation techniques to calculate the position of the detected objects in a stereo vision rig coordinates frame. Verification of a proper operation of the proposed algorithm was made by conducting series of experiments. Our results indicates that the algorithm detects objects accurately and the troublesome un-stationary background regions can be excluded from detection using the presented localization method.

Keywords: object detection, object localization, stereo vision.

1 Introduction

In the last few years, visual surveillance has become a challenging area in a computer vision, especially in a view of the growing importance of these systems for security purposes [1–4]. The ultimate target in designing smart surveillance systems is to minimize the need of continuous monitoring and analyzing visual data by an operator [5]. This goal seems to be very difficult to reach without developing trusted object detection and localization algorithms.

Automatic moving object detection and localization algorithms play a fundamental role in video surveillance [1, 2]. Moving object detection algorithms are necessary to detect threat, while localization algorithms may be used for identifying a detected danger.

The algorithms have been proposed in literature for an object detection can be categorized as optical flow, a frame difference and a background subtraction [6]. The goal of the optical flow estimation is to compute an approximation to the motion field from an time-varying image intensity. Unfortunately, this method is highly complicated and very sensitive to a noise [7]. The frame difference and the background subtraction are based on a pixel difference between a reference image and a current image [8]. For the frame difference, the reference image is the previous frame. The frame difference method is able to detect objects even though the environment changes dynamically, but it is ineffective for detection of low-speed objects. In the background subtraction method, the reference frame is reconstructed from the previous

J. Awrejcewicz et al. (eds.), *Mechatronics: Ideas for Industrial Applications*,
Advances in Intelligent Systems and Computing 317, DOI: 10.1007/978-3-319-10990-9_39

video sequence, which contains an observed scene with no moving objects [9]. The background subtraction has been reported as the most popular object detection method because of its high effectiveness and simplicity in implementation. However, the simple background methods are inadequate to handle rapid lighting and shadow changes.

As a result, many more sophisticated methods, based on the optical flow, the frame difference and the background subtraction have been developed to reduce mentioned drawbacks [1–5, 7–9]. Unfortunately, these methods strongly depend on applications and camera parameters, consequently cannot be easily adopted to use for other purpose. For example, a resolution and optics of cameras, indoor and outdoor conditions, lighting and a size or a speed of potential objects play an important role in a selection and a parameterization of an image processing technique. Therefore, for the purpose of the object detection, the novel algorithm suited for our application was elaborated. This algorithm is particularly design for military applications; it is able to detect especially humans, vehicles and planes.

For the object localization, a stereo vision method was applied. Usage of two cameras enables a calculation of localizations of various points in a scene, relative to a position of cameras [10–12]. Much research in recent years has been focus on implementation of the stereo vision in a large variety of applications [12, 13]. Most of developed algorithms are based on a disparity map calculation and a triangulation technique [6, 13]. Because of the disparity map is very time consuming, in our work the triangulation algorithm was used to calculate a position of a detected object in a stereo vision rig coordinates frame.

2 Methodology

2.1 Stereo Vision

Depth perception is one of the most important tasks of computer vision systems. A stereo correspondence by calculating localizations of various points in a scene, relative to the position of cameras, allows to perform complex tasks, such as depth measurements and an environmental reconstruction [10].

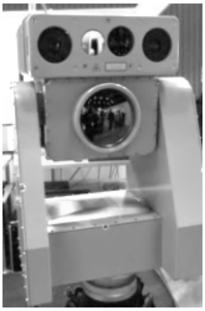

Fig. 1. Sensor head with CCD-C-Z36 TV cameras (Carl Zeiss Optronics GmbH)

For the purpose of object detection and localization, the stereo rig shown in Fig. 1 was used. It is built in the sensor head, consists of the thermal imager ATTICA, the visual daylight color TV camera and the laser range finder LDM38, stable aligned to each other. The sensor head is especially designed for an advanced surveillance task under rough environmental conditions. The main components of the stereo rig are visual cameras. They combine ¼ inch CCD detector with a powerful 36 × auto-focus zoom lens providing a wide/telescopic field of view, ideally suitable within surveillance system applications. The effective picture elements of the cameras are approximately 440,000 px (752 × 582). The distance between cameras is equal to 225 mm. For the purpose of a video stream acquiring and a real-time operation, the Matrox Morphis Family frame grabber and Matrox Imaging Library were used.

2.2 Camera Parameters

In order to obtain a reconstruction of the scene depth in the Euclidean space, it is necessary to determine the camera parameters. The classic calibration methods are based on a specially prepared calibration pattern with known dimensions and a position in a certain coordinate system [11]. For the purpose of obtaining the cameras parameters, the calibration pattern with 289 markers was used. The calibration procedure (presented in detail in [11]) was conducted for 10 different zoom levels.

2.3 Triangulation

The 3D reconstruction from two views involves extracting target features from one image, matching and tracking these features across the second image, and using a triangulation method to determine position of the target points relative to the stereo rig. In this work, an extracting target area was set using the object detection algorithm on the left camera image. Next, the correspondence problem of finding the same windows in the right image was solved using correlation-based method. The Sum of Absolute Differences used in this work is presented below [12]:

$$\mathrm{SAD}(x,y,d)=\sum_{i=-w}^{w}\sum_{j=-w}^{w}\left|I_1(x+i,y+j)-I_2(x+d_x+i,y+d_y+j)\right| \qquad (1)$$

where I_1, I_2 are left and right image pixel grayscale values; d_x, d_y are disparity ranges; w is window size; i, j are coordinates of the central pixel of the working window for which the similarity measurement is computed.

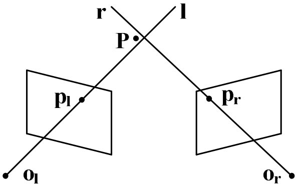

Fig. 2. Triangulation with non-intersecting rays

For the estimated central point of the windows in the left and right images (p_l, p_r), the point $\boldsymbol{P}(a, b, c)$ lies at the intersection of the two rays from o_l through p_l and from o_r through p_r respectively (Fig. 2).

Because of approximate camera calibration parameters and a target location, the two rays don't actually intersect in the space, their intersection can only be estimated as the point of minimum distance from both rays. Assuming $a\boldsymbol{p}_\mathrm{l}$ can be the ray l, $\boldsymbol{T} + c\boldsymbol{R}^\mathrm{T}\boldsymbol{p}_\mathrm{r}$ can be the ray r and $\boldsymbol{w}$ can be a vector orthogonal to both l and r, triangulation problem reduces to determining a midpoint of segment parallel to $\boldsymbol{w}$ and joins l and r. It can be computed solving the linear system of equations [12]

$$a\boldsymbol{p}_\mathrm{l} + b\boldsymbol{w} = \boldsymbol{T} + c\boldsymbol{R}^\mathrm{T}\boldsymbol{p}_\mathrm{r} \tag{2}$$

where a, b, and c are coordinates of point $\boldsymbol{P}$.

2.4 Object Detection Algorithm

One of the purpose of this study was to elaborate an reliable object detection algorithm. It means, that any changes caused by a new object should be detected, whereas un-stationary background regions, such as branches and leafs of a tree or a flag waving in the wind should be identified as a part of the background. In order to meet the above assumption, the following algorithm was proposed. To present our work in a readable way, the algorithm was divided into the main pieces and shortly described.

Grabbing 5 Consecutive Frames. Working with more consecutive frames improve detection quality, but it is very time consuming. In our experiment, there was found that 5 frames were the best choice for the further calculation. Because of the video stream consists 25 frames per second, this assumption determines that the algorithm can detect object 5 times per second. The each grabbed frame is divided then into red, blue and green channels. All operation on images were performed using Matrox Library, which represents each pixel of an image as an element of a matrix.

Image Filtration Using Median Filter. A median filter is effective against all local noise pulse, causing them to not blur in to the larger areas [13]. It eliminates those pixels of the image for which the intensity values differ significantly from the other pixel intensity values in the window. Median filtering does not introduce new values to the image, so requires no additional scaling.

Reduction of Resolution of Frames. It was presented in the literature [14], that a reduction in a resolution of an image helps decrease an influence of a noise. In the present work the resolution of frames is reduced by 50 %.

Edge Detection Using Sobel Operators. For each channel, an edge detection is carried out using Sobel operators. The advantage of using Sobel operators is that the calculated edges are very broad [15]; this feature is important for further standard deviation calculations. Working on each color channel separately allows to improve the edge detection.

Addition Red, Blue and Green Channels. In this step the edges of 5 consecutive frames are retrieved. The red, blue and green channels of each frame are added and normalized to consist the edge information in the range from 0 to 1.

Combination of 5 Consecutive Frames. This is crucial a part of the algorithm. By combination of 5 consecutive frames we achieved that the only trace of moving object was capture on the image. Number of 5 frames allows for real-time calculations with fairly well object detection. For the smaller number of frames objects were not detected properly, whereas larger number of frames was not suited for a real-time calculation. The combined frame was calculated as

$$f_c = \left| f_{k-4} + f_{k-3} + f_{k-2} \,|-|\, f_{k-2} + f_{k-1} + f_k \right| \tag{3}$$

where f_c is the combined frame and f_k is the number of the grabbed frame.

Division the Combined Frame into 10 × 10 Pixels Blocks. In this step the combined frame are divided into 10 × 10 px blocks. For each block a mean value of all pixels is calculated. Next, the matrix composed of mean values of pixels blocks is computed.

Statistical Analysis. The matrix obtained in the previous step is used for a statistical analysis. First, the standard deviation in each row and column is calculated. Then, the value of each element of the matrix is compared to the mean value of the standard deviation of its row and column. The elements with higher values are classified as foreground and label as 1. Simultaneously, each element is compared with mean value of all elements and classified as previously. Then, each elements classified as foreground in both comparisons is labeled as containing moving object.

Morphological Opening Operation. Morphological opening operation removes small objects and details, and smooths the contour of the recognized object, without changing its size [15]. This operation is able to clean the background from almost all noises in the form of short segments defined at the stage of a statistical analysis.

3 Results

It is apparent that an universal detection and localization algorithm suited for every application is impossible to elaborate. Therefore, in this study, the algorithm particularly design for military applications, that is able to detect especially humans, vehicles and planes was performed. Verification of its proper operation was made by conducting series of experiments. Fig. 4 shows sample results of executed tests.

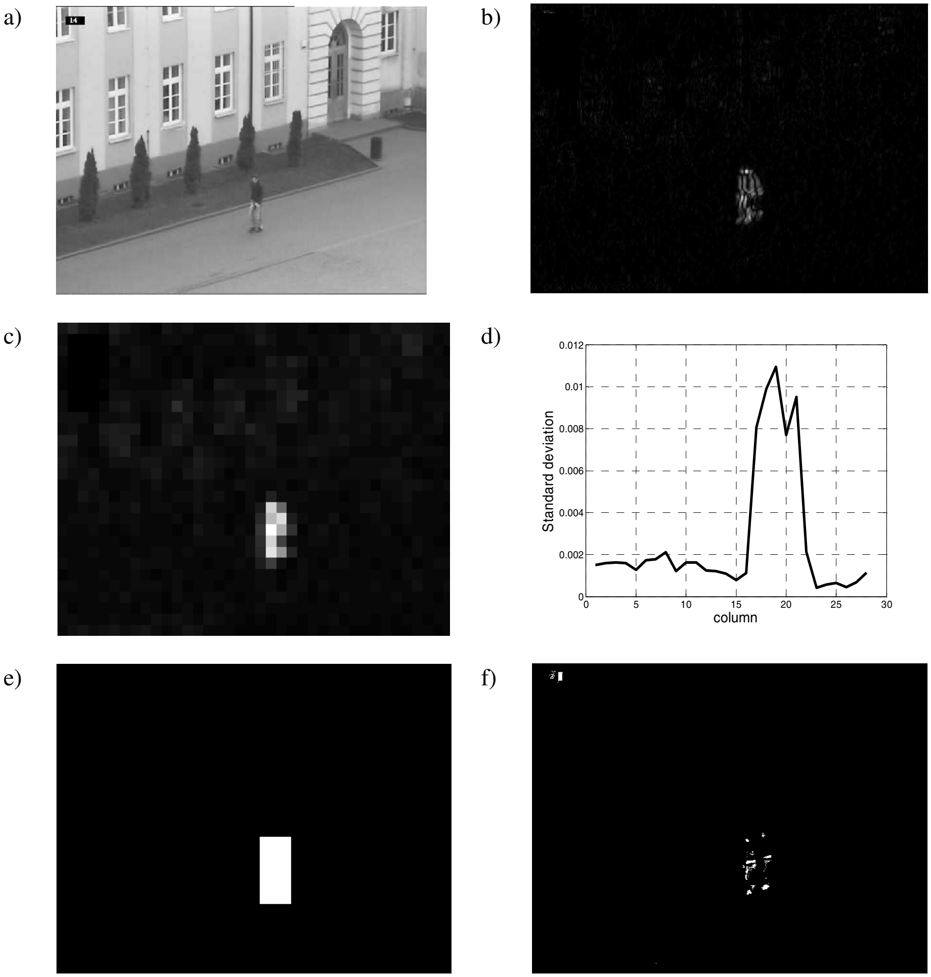

Fig. 3. Performance of the algorithm for an example video sequence

The third frame grabbed from 5 consecutive images is visible in Fig. 4a. Fig. 4b illustrates combination of the grabbed images after a median filtration, a reduction of resolution, an edge detection and an addition of red, green and blue channels. As can be seen, the only edges connected to the moving object remained on the picture.

The effect of a division the combined frames into 10 × 10 px blocks can be observed in Fig. 4c. This step is very important because identifies and connects an area where the detected object is placed.

Some results of the statistical analysis are given in Fig. 4d. By calculation of the standard deviation in each column, the accurate position of the detected object was obtained and additionally a remained noise was removed from the frame.

The standard deviation calculation is an essential step for a foreground classification and a morphological opening operation shown in Fig. 3e. The foreground classification identifies and connects an area of the detected object, whereas a morphological opening operation removes the noise and small unnecessary objects.

In order to compare our method with methods presented in the literature, basic Gaussian mixture and optical flow algorithms were implemented. For example, figure 3f illustrates the Gaussian mixture method for the example video sentence. As can be seen, the background is not updated properly and the object steel exist in its previous position. On the other hand, optical flow method appeared to be too sensitive and a lot of noise was detected as the object. It confirms that basic algorithms should not be adapted for a specific application.

The labeled area from the Morphological opening operation step is used in The Sum of Absolute Differences method for solving the correspondence problem of finding the same windows in both stereo pair images. Solving correspondence problem allows determine a position of the target relative to the stereo rig using the triangulation method. To validate the results the obtained position was compared with a handmade measurement. For the various 3D scenes observed by the cameras, the positions of the selected points were calculated using the triangulation method. Then, the handmade measurement using an laser distance meter was done. The experiment shown that the distance difference between the hand-made and the stereo measurement was less than 3 % for the close objects (up to 10 meters) and less than 10 % for the distant objects (up to 100 meters). The obtained results strictly dependent on geometric parameters of a stereo vision system; accuracy of a measurement decreases with distance increase between a stereo rig and a target. It could be unacceptable for applications based on an exact position, but for the target localization it seems to be appropriate.

In general, for the most obtained results, the algorithm detects objects accurately. It should, however, be noted that sometimes un-stationary background regions, such as branches and leafs of a tree or flags waving in the wind were detected as foreground. In this case the target localization algorithm can be easily used. It is possible to mark an unwanted object position and exclude the object from detection using the presented localization method. For example, branches of a tree would be passed over, whereas moving cars of humans in front of the tree, nearer to the cameras, would be detected.

4 Conclusions

The problem of smart visual surveillance for an automatic object detection and localization was studied. We have developed the algorithm particularly design for military applications that is able to detect especially humans, vehicles and planes. For the purpose of the object detection and localization, the stereo rig consist of CCD-C-Z36 TV (Carl Zeiss Optronics GmbH) cameras have been used. We have focused on a classic method of cameras calibration and triangulation technique to calculate a position of a detected object in a stereo vision rig coordinates frame.

Our results show that the algorithm detects objects accurately and the troublesome un-stationary background regions can be excluded from the detection using the presented localization method. However, one positive feature of a 3D reconstruction using a stereo vision system have not been utilized; usage of two cameras enables not only localization, but also calculation of detected object dimensions. This advantage is very important for a classification problem. Therefore, the future work will focus on the classification problem based on dimensions and shapes of detected objects.

References

1. Micheloni, C., Foresti, G.L.: A robust feature tracker for active surveillance of outdoor scenes. In: Electronic Letters on Computer Vision and Image Analysis, pp. 21–34 (2003)
2. Cucchiara, R., Prati, A., Vezzani, R.: Advanced video surveillance with pan tilt zoom cameras. In: Proc. of the 6th IEEE International Workshop on Visual Surveillance, pp. 334–352 (2006)
3. Hu, W., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behaviors. IEEE Trans. on Systems, Man, and Cybernetics 34, 334–352 (2004)
4. Cohen, I., Medioni, G.: Detecting and tracking moving objects for video surveillance. In: Proc. IEEE Computer Vision and Pattern Recognition, Fort Collins CO, pp. 1–7 (1999)
5. Czyżewski, A., Szwoch, G., Dalka, P., Szczuko, P., Ciarkowski, A., Ellwart, D., Merta, T., Łopatka, K., Kulasek, Ł., Wolski, J.: Multi-Stage Video Analysis Framework. In: Video Surveillance, pp. 161–216. In-Tech, Rijeka (2011), doi:10.5772/625
6. Jain, R., Kasturi, R., Schunck, B.: Machine Vision. McGraw-Hill Inc., New York (1995)
7. Zhang, D., Lu, G.: An edge and color oriented optical flow estimation using block matching. In: Int. Conf. Signal Processing, Beijing, vol. 2, pp. 1026–1032 (2000)
8. Lien, C.: Targets Tracking in the Crowd. In: Video Surveillance, pp. 232–246. In-Tech, Rijeka (2011), doi:10.5772/625
9. Ince, E.A., Naraghi, N.S., Ebrahimi, S.G.: Background Subtraction and Lane Occupancy Analysis. In: Video Surveillance, pp. 175–199. In-Tech, Rijeka (2011), doi:10.5772/625
10. Żak, B., Hożyń, S.: Distance Measurement Using a Stereo Vision System. In: Advances in Mechatronic Systems. Mechanics and Materials, vol. 196, pp. 189–197. Trans Tech Publications Ltd., Zurich (2013), doi:10.4028/www.scientific.net/SSP.196.189
11. Li, M., Lavest, J.: Some Aspects of Zoom-Lens Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1105–1110 (1995)
12. Cyganek, B., Siebert, P.: An Introduction to 3D Computer Vision Techniques and Algorithms. John Willey & Sons, Chippenham (2009)
13. Trucco, E., Verri, A.: Introductory Techniques for 3D Computer Vision. Prentice-Hall, New Jersey (1998)
14. Sugandi, B., Hyoungseop, K., Tan, J.K., Seiji, I.: A Block Matching Technique for Object Tracking Based on Peripheral Increment Sign Correlation Image. In: Object Tracking, pp. 1–21. In-Tech, Rijeka (2011)
15. Żak, B., Hożyń, S.: Segmentation Algorithm Using Method of Edge Detection. In: Advances in Mechatronic Systems, Mechanics and Materials, vol. 196, pp. 206–211. Trans Tech Publications Ltd., Zurich (2013), doi:10.4028/www.scientific.net/SSP.196.206