Depth Map's 2D Histogram Assisted Occlusion Handling in Video Object Tracking

Adam Łuczak, Sławomir Maćkowiak, and Jakub Siast

Poznań University of Technology, Poland
{aluczak,smack,jsiast}@multimedia.edu.pl
http://www.multimedia.edu.pl

Abstract. The paper describes new algorithm for automatic video object tracking. Proposed architecture consists of two loops of Kalman filter. In the loop of the tracking process, the information achieved from video and from 2D histogram based on depth map is used. Two loops work simultanously and the parameters between the loops are interchanged when the occlusion occurs. The 2D histogram representation of the depth map has unique properties that can be used to improve the tracking eficiency especially in the case of occlusions of the objects in the image. Experimental results prove that the proposed system can accurately track multiple objects in complex scenes.

1 Introduction

Accurately tracking of moving objects within monitored scenes is crucial to a range of surveillance tasks. There are many effective methods of detecting and tracking objects, many analyses have been conducted to improve object tracking technique accuracy. Comprehensive literature survey of object tracking is presented in [1]. Another comprehensive survey was also produced by [2,3]. Techniques used in object tracking are categorized on the basis of the used type of objects and used motion representations. The most significant challenge in video object tracking is the frequent problem of occlusion. During occlusion, an ambiguity occurs in occluded object features. The tracking methods must be capable to resolve the individuality of the objects involved in the occlusion, when the occlusion takes place.

In real situation 3 types of the occlusion occur: *i*) a self-occlusion when one part of the object occludes another, *ii*) an inter-object occlusion when two objects being track occlude another object, *iii*) and occlusion by the background when a structure in the background occludes the tracked objects.

In this paper, a new technique to deal with the second type of the occlusions using the information contained in the depth maps is presented.

Depth maps create new opportunities to improve the algorithms of analysis of 3D scene, also in the video object tracking. New video acquisition systems often use the stereo cameras that allow the calculation of the depth map - an image that contains information about the distance of the points from the lens.

L.J. Chmielewski et al. (Eds.): ICCVG 2014, LNCS 8671, pp. 400–408, 2014.

Until recently, the depth map estimation algorithms were very complex, time-consuming and generated depth maps of poor quality. For depth estimation a global optimization algorithm like Belief propagation or Graph cuts has been used. Nowadays, the local cost aggregation methods are used instead of global optimization [5]. For example Winner-Takes-All (WTA) algorithm with low complexity gives better results than the old solutions. WTA algorithm is also suitable for hardware implementations. Moreover, increasingly are also available stereoscopic cameras with depth map estimation support. Some of the hardware or even software solutions allow obtaining a depth map in real time [6-9]. For example the Bumblebee stereo camera system offered by Point Grey Research [11] is able to produce a depth map in real time. Despite significant progress in algorithms of depth maps estimation, they still are not perfect. Main problems arise when the scene consists some semi-transparent objects, light reflections, occlusions or obscuring objects. On the other hand, even not perfect depth maps contain reach information about 3D scene. It is still possible to obtain information about the object distance from the camera lenses. Such information can help us properly measure of scaling or distance for tracked objects (e.g. car or person, etc.)[10]. The authors consider the use of depth maps to improve the efficiency of the tracking when inter-object occlusions occur.

This paper is divided into 4 main sections. Section 2 presents a novel idea of 2D histogram of depth map. Section 3 presents a new video object tracking algorithm that uses the information from 2D histogram of depth map in the tracking process. In the section 4 the assumptions of the experiments and achieved results for object tracking under the occlusions conditions are presented.

2 2D Histogram of Depth Map

Luminance of each pixel in a depth map is interpreted as a normalized disparity. Usually depth maps with 256 disparity levels are used. The proposed 2D histogram is a graphical representation of disparity values distribution in a depth map. For a depth map with resolution J x K pixels we build the 2D histogram with resolution J x 256. Each column of the 2D histogram is a 1D histogram with 256 bins corresponding to 256 disparity levels. This 1D histogram for column j ($j \subset \langle 1; J \rangle$) is calculated for j-th column of the depth map.

The proposed 2D histogram is defined as (1):

$$2D\ histogram = DH = \begin{bmatrix} L(0,1) & \cdots & L(0,J) \\ \vdots & \ddots & \vdots \\ L(255,1) & \cdots & L(255,J) \end{bmatrix} \frac{255}{K} \tag{1}$$

where $L(i,j)$ is a number of pixels in j-th column of a depth map that have disparity value of i. K is a number of pixels in a single depth map column. The 2D histogram values are normalized to range $\langle 0, 255 \rangle$. Fig.1 presents a depth map and its 2D histogram.

Considering a picture from Fig. 1 c and associated 2D histogram from Fig.1 b we can see that three people from a center part of a picture are represented in a 2D depth histogram as a three separated aggregations of lighter pixels.

Depth map with camera parameters represents information about three dimensional scene. In presented algorithm for 2D histogram calculation a depth map is treated as a two dimensional picture. No information about perspective are exploited. The presented technique has two useful properties. First is a low computational complexity. Second, horizontal coordinate of an tracked objects in a picture are the same as horizontal coordinate of its representation in 2D histogram. With this property associating an objects in a sequence and in a 2D histogram is simplified.

a. b. c.

Fig. 1. a. Depth map and b. 2D histogram associated with c. a frame 282 from *Poznan Car Park* sequence

What is very important, partly obscured objects are still easy distinguishable on a 2D histogram. Of course, object cannot be recognised properly even using depth map due to the occlusion, but the information about the position in space of not obscured part is still correct.

3 Tracking Approach Assisted by the Information from 2D Histogram of Depth Map

Occlusions are predicted by checking pairs of bounding boxes at predicted positions. Suspending the update phase for any length of time, however, is problematic since motions (particularly of people) can rapidly evolve. A recent simple but effective approach is to track the boundaries of bounding boxes separately which results in at least some updating evidence recovered for a substantial proportion of the occlusion event.

The proposed video object tracking architecture is presented in Figure 2. The tracker consists of two Kalman Filter loops. The first loop works on consecutive frames of the video. This is a typical implementation of a system for object detection and tracking based on the segmentation and classification of moving pixels in the scene. Motion detection processes locate blobs (connected regions of moving pixels) to create a candidate list of observations of the current active scene objects. Normally these blobs are recovered by pixel differencing against the reference frame of the static scene, usually attributed with their bounding box.

The Tracker module is implemented using a two-step approach: prediction and update. In the prediction step of the procedure, position of the objects tracked in previous frames are projected to the current frame according to trajectory models. Next, in the Data Association step predicted positions of objects are confrontated with list of candidate observations i.e. objects from Object Detection phase. Corresponding objects and observations are found.

The second loop is a fresh approach in the tracking systems. This loop works on 2D histogram of disparity map (a depth map). In this domain, the process locates moving regions and creates a candidate list of observations of objects. In the concept, the algorithm defines the moving blobs in two-dimensional space of the histogram. When two or more objects occlude one another it can still be possible to separate objects in 2D histogram. The only prerequsitive is that the objects have a different assiociated depth. If that condition is fulfilled the object will be represented in 2D histogram of depth map as separated blobs. This makes segmenting the object blob during occlusion easier.

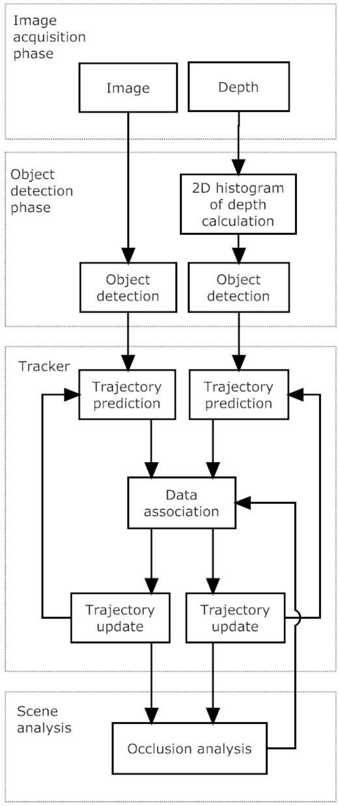

Fig. 2. Tracking architecture

In order to apply the Kalman filter, the process should be described by the following linear equations:

$$\overline{x_{I_k}} = A_I x_{\widehat{I_{(k-1)}}} + B_I u_{I_{(k-1)}} \tag{2}$$

$$\overline{x_{D_k}} = A_D x_{\widehat{D_{(k-1)}}} + B_D u_{D_{(k-1)}} \tag{3}$$

$$z_{I_k} = H_I x_{I_k} + \nu_{I_k} \tag{4}$$

$$z_{D_k} = H_D x_{D_k} + \nu_{D_k} \tag{5}$$

The equations (2) and (4) concern the loop in video domain, (3) and (5) concern the loop in the 2D histogram of the depth domain. The (2) and (3) equations are called the equations of state or process models, while the (4) and (5) are the measurement models. In the above equations, A, B, H are matrices,where A_I and A_D are the state-transition models for image and depth equations, H_I and H_D are the observation models for image and depth equations, B_I and B_D are the control-input models for image and depth equations, the vector x is called the state of the system, the vector contains information from the input system, e.g., predetermined speed of the objects. Vector z is the measured output of the system. However, u and ν mean a noise (standard deviations). However, u means the process noise whereas ν is the measurement noise. During the prediction step, based on the previous x statea new value of x is determined, and the covariance matrix Q_I for image and Q_D for the depth. These values are determined on the basis of above equations.

$$Q_I = E[u_{I_{(k-1)}}(u_{I_{(k-1)}})^T] \tag{6}$$

$$Q_D = E[u_{D_{(k-1)}}(u_{D_{(k-1)}})^T] \tag{7}$$

$$\overline{P_{I_k}} = A_I P_{I_{(k-1)}} A_I{}^T + Q_I \tag{8}$$

$$\overline{P_{D_k}} = A_D P_{D_{(k-1)}} A_D{}^T + Q_D \tag{9}$$

In the correction (update) phase we set the variable K, hereinafter referred to as the Kalman gain.

$$K_{I_k} = \overline{P_{I_k}} H_I^T (H_I \overline{P_{I_k}} H_I^T + R)^{-1} \tag{10}$$

$$K_{D_k} = \overline{P_{D_k}} H_D^T (H_D \overline{P_{D_k}} H_D^T + R)^{-1} \tag{11}$$

At the beginning, the Kalman gain is determined. If we look at the way how the K is calculated, (10) and (11), we come to the conclusion that if the measurement noises are greater which here is represented by the covariance R, the value of K is lower. Here we come to the heart of the proposal. In the case of a small value of K_{I_k} (for the object tracking in the video, it indicates the occlusion existance) and when the second parameter for the depth loop K_{D_k} is greater than δ_D (the Kalman gain does not indicate the measurement error, no occlusion exist) the R_I covariance matrix should be replaced by the R_D covariance matrix (12). The parameter δ is used to control the interchange between parameters R_I and R_D. Its value was chosen experimentally. Motion object segmentation in the 2D histogram of depth map gives more precise information about the object moving trajectory. Due to the different values and measurement representation between image and depth, the covariance matrix can not be used directly but scaling is required. The measurement in the 2D histogram of the disparity is more reliable, the standard deviation and the R_D has the lower values.

Vice versa, the covariance matrix R_I in the case of a small value K_{D_k} should be replaced by the scaled value of R_I. Only of course if the K_{I_k} is greater than δ_I.

$$\begin{aligned} \text{If } K_{I_k} \leq \delta_I \cap K_{D_k} > \delta_D \text{ then } R_I = R_D scale, \\ \text{and if } K_{D_k} \leq \delta_D \text{ then } R_D = R_I \frac{1}{scale}. \end{aligned} \tag{12}$$

From (4) and (5) the position of the detected blobs in the image and 2D histogram of the depth map (from the measurement phase) z_{I_k} and z_{D_k} are calculated. For two-dimensional space:

$$H_I = H_D = \begin{bmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \end{bmatrix} \tag{13}$$

After the measurement process the new values of the process state for image and 2D histogram domain are calculated (14)(15), the values of the covariance matrices R_I and R_D are updated.

$$\widehat{x_{I_k}} = \overline{x_{I_k}} + K_{I_k}[z_{I_k} - H_I \overline{x_{I_k}}] \tag{14}$$

$$\widehat{x_{D_k}} = \overline{x_{D_k}} + K_{D_k}[z_{D_k} - H_D \overline{x_{D_k}}] \tag{15}$$

$$P_{I_k} = [1 - K_{I_k} H_I]\overline{P_{I_k}} \tag{16}$$

$$P_{D_k} = [1 - K_{D_k} H_D]\overline{P_{D_k}} \tag{17}$$

$$R_I = E[\nu_{I_{(k-1)}}(\nu_{I_{(k-1)}})^T] \tag{18}$$

$$R_D = E[\nu_{D_{(k-1)}}(\nu_{D_{(k-1)}})^T] \tag{19}$$

4 Experimental Results

The final versions of the proposed system architecture to object tracking have been obtained by extensive iterative experiments in the scope of this paper. Proposed two-loops Kalman filter tracking algorithm has been implemented. Here, we report the experimental results that allow to estimate the overall efficiency of the object detection under the occlusions condition. This is done using subjective and objective tests on two MPEG MVD test sequences (1920 x 1080 resolution - *Poznan Car Park*, *Poznan Street*). More extensive test for other MVD sequences are planed when depth maps will be available.

Table 1. Experimental results

Sequence	Tracking mode	Total frames	Frames with occlusions	Accuracy [frames]	Accuracy [%]
Poznan Car Park	without DM	150	98	93	62%
	with DM	150	98	138	92%
Poznan Street	without DM	130	35	98	75%
	with DM	130	35	121	93%

a b c

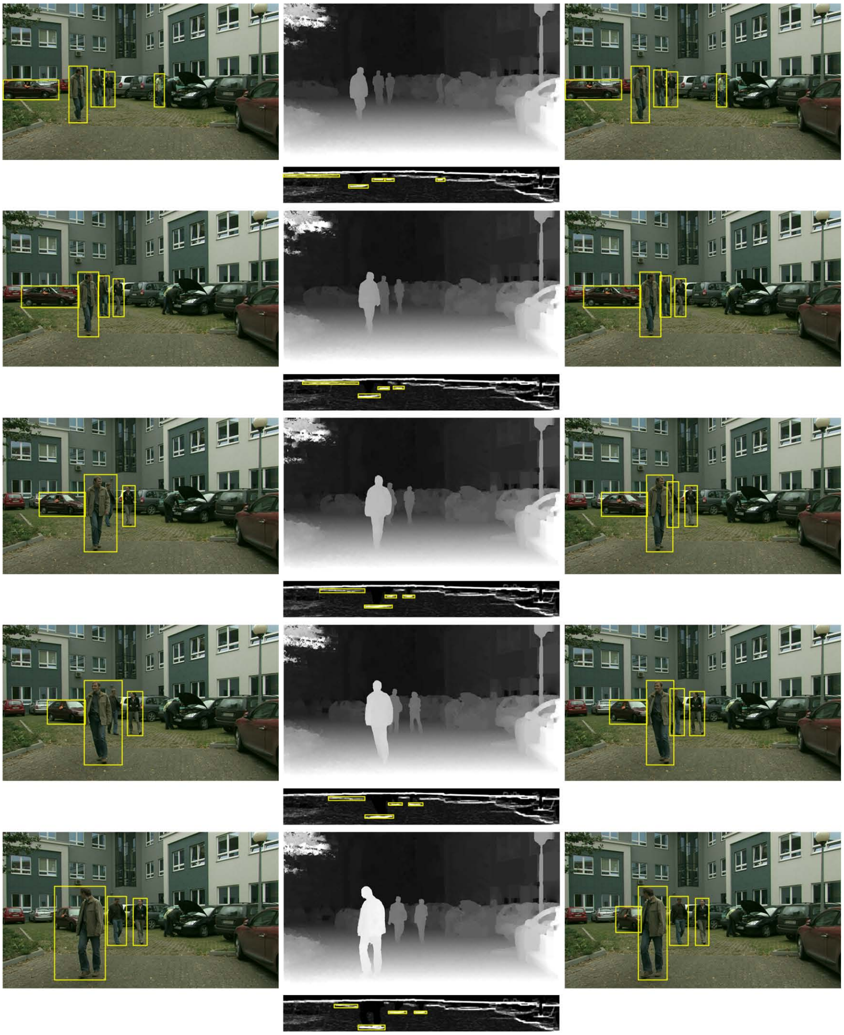

Fig. 3. Tracking algorithm results for frames numberr 223, 254, 282, 299, and 315 of *Poznan Car Park* sequence (a bounding boxes of detected moving objects achieved for the tracking system with and without information from depth)
a. the case of the tracking algorithm without the information from the depth maps.
b. the information used in the proposed solution to gain the efficiency of the detection.
c. the case of the proposed tracking algorithm (the two-loops Kalman filter solution exploiting video and 2D histogram of a depth map).

The experiments were divided into 2 steps. First test was done for architecture with one-loop Kalman filter. The information from 2D histogram of depth map has not been used at this moment. The moving object blobs are detected with HOG features and SVM classification process. The second step of the experiment use the full proposed approach (detection and tracking object on the video and detection and tracking information on the 2D histogram of the depth map). The results of the detection for both steps are presented in the Table 1. As shown on the Fig.3, also subjective tests prove that the proposed solution achieves more efficiency than classic method exploiting only video information (one-loop architecture). The average gain of the efficiency of object tracking under the occlusions is more than 24% for all frames of the sequence, moreover the gain of the efficiency for the frames when the occlusions occur only is higher than 85%.

5 Conclusions

In this paper, a novel idea of the system to track the objects in video sequence has been presented. In the paper the architecure of the tracking system, which exploits the information from the depth has been proposed. Proposed method combines well known Kalman filter based tracking method and 2D histogram of depth map. This architecture has been tested under the scenarios where different occlusion situations were present. This original approach results in significant improvement of accuracy of objects' tracking. Moreover, by adding a depth map and 2D histogram to tracking algorithms their functionality has been enriched.

Acknowledgement. Research project was supported by the National Science Centre, Poland according to the Grant no. 4775/B/T02/2011/40.

References

1. Pulford, G.W.: Taxonomy of multiple target tracking methods. IEE Proceedings Radar, Sonar and Navigation 152(5), 291–304 (2005)
2. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Computing Surveys (CSUR) 38(4), 1–45 (2006)
3. Turaga, P., Chellappa, R., Subrahmanian, V.S., Udrea, O.: Machine recognition of human activities - a survey. IEEE Transactions on Circuits and Systems for Video Technology 18(11), 1473–1488 (2008)
4. Du, Y., Chen, F., Xu, W., Li, Y.: Recognizing interaction activities using dynamic Bayesian network. In: International Conference on Pattern Recogntion, pp. 618–621 (2006)
5. Scharstein, D., Szeliski, R.: Middlebury Stereo Vision Page, `http://vision.middlebury.edu/stereo/` (online December 1, 2013)
6. Yongtae, K., Jiyoung, K., Wanghoon, S.-K.-H.: Fast Disparity and Motion Estimation for Multi-view Video Coding. IEEE Transactions on Consumer Electronics 53(2), 712–719 (2007)
7. Wu, D., Zhang, H., Li, X., Qian, L.: Depth Map Generation Algorithm for Multiview Video. In: 2013 Fifth International Conference on Computational and Information Sciences (ICCIS), Shiyang, China (2013)
8. Lee, L.S.H., Sharma, S.: Real-time disparity estimation algorithm for stereo camera systems. IEEE Transactions on Consumer Electronics 57(3), 1018–1026 (2011)

9. Han, H., Han, X., Yang, F.: An improved gradient-based dense stereo correspondence algorithm using guided filter. International Journal for Light and Electron Optics 125(1), 115–120 (accepted for publication, 2014)
10. Claps, A., Reyes, M., Escalera, S.: Multi-modal user identification and object recognition surveillance system. Pattern Recognition Letters 34(7), 799–808 (2013) ISSN 0167-8655
11. Scharstein, D., Szeliski, R.: Middlebury Stereo Vision Page, `http://vision.middlebury.edu/stereo/` (online December 1, 2013)