Adaptive and Selective Multi-Target-Tracker

Chen-Ko Sung
Fraunhofer Institute IOSB, Fraunhoferstrasse 1, 76131 Karlsruhe, Germany
chen-ko.sung@iosb.fraunhofer.de

ABSTRACT

An adaptive and selective multi-target tracker combines various methods for object detection and tracking such as feature-based methods and trajectory-based tracking, thus providing meaningful results with variable shape, brightness and size of the target images. In the case of a high dynamic scene, dynamically varying parameter sets must be used for the detection of object images and object tracking. This requires an automatic generation of parameter sets by adaptive adjustment. In this paper, the selected process chains for the automatic and seamless tracking of multiple object images and for the interpretation of target objects in image sequences are presented.

Keywords: adaptive threshold, multi-target tracker, image or scene interpretation, motion, process chain, detection and tracking, automatic generation of parameter sets, cumulative histogram.

1. INTRODUCTION

Moving objects can be imaged, either small or large and either bright or dark, in a long image sequence such as a highway scene or an aerial scene. In this case, the development and implementation of a process chain for (automatic) object tracking [1] is a great challenge, since it requires the use of different approaches in combination, both in terms of the extraction and tracking of image features as well as the grouping of characteristics.

Object trackers use mostly threshold values to detect and track moving objects. If object images change their intensities or their sizes from time to time, some tracking algorithms may not work well for the whole time and need manual resetting of the threshold value. Therefore, we propose an adaptive and selective multi-target tracker (ASMTT) in this paper to improve the methods for the object detection and tracking. The ASMTT combines various methods such as feature-based methods and trajectory-based tracking for object detection and tracking, and provides meaningful results even when the target images change their shapes and brightness as well as sizes in an image sequence.

The proposed adaptive and selective multi-target tracker is generic, since it offers a choice of alternative methods depending on the individual tasks, on the used image sensors and on the type of motion of the target objects. In the case of a high dynamic scene, dynamically varying parameter sets must be used for the detection of object images and object tracking. This requires an automatic generation of parameter sets by adaptive adjustment.

In this paper, the selected process chains for the automatic and seamless tracking of multiple object images and for the interpretation of target objects in image sequences are presented. Detailed descriptions of objects were extracted from the images for the interpretation of target objects. The comparison with existing object descriptions in the search system permits to decide what the object image represents.

Video-based motion analysis has increasing importance for diagnosis of body functions in medicine [2], for studies on human motion [3, 4] and movement behavior of animals [5], and for research in water channels [6], etc. The proposed ASMTT has not been implemented specifically for surveillance purposes, but in the above-mentioned application areas, where the analysis of movements plays a role, it can be used.

2. SYSTEM OVERVIEW

A created process chain for the automatic and uninterrupted tracking of object images in image sequences is often called multi-target tracker, and consists mostly of the first and the second component of the following four components:

- detection of object images (segmentation),
- tracking of object images,
evaluation and classification of trajectories, and
image or scene interpretation.

Figure 1 shows a modular system concept for such a multi-target tracker with all four components mentioned above. A modern multi-target tracker is to have several alternative methods for each component depending on the task, the image sensors, or the type of motion of the target objects to choose from. Suitable methods for segmentation are contour or edge operator [7, 8], monotony operator [9], threshold operator [10, 11, and 12] and so on. The used segmentation methods should be able to produce meaningful results with different parameter sets for example for a surveillance task, because many scene-specific properties may change constantly, like shape, brightness and size of the target figures, etc., during the monitoring period. All parameter sets for the segmentation, except the initial parameter set, should be generated adaptively.

In order to illustrate how to operate the switch from one parameter set to another parameter set adaptively as well as from one method to another method automatically, the steps of one single object tracking according to the image structures were roughly divided for example into three phases (see Figure 2):

- phase 1: Detection and tracking of object images with blob like small structure,
- phase 2: Detection and tracking of object images with non-evaluable sheet like structure,
- phase 3: Detection and tracking of object images with a sheet like and well evaluable structure.

A target object image can have any size and shape in the initial phase and may change from time to time from one phase to the adjacent phase. For the determination of a phase change, the structural changes of the target object image were evaluated.
Image structures such as blobs, contour segments or line segments, etc. are elements of image features that can be extracted from images by a segmentation method, and are often used for the representation of the object images. An object detection and tracking with contours or line segments in phases 1 and 2 is very difficult to implement, if the images contain highly structured background and extreme short straight lines from small targets. On the one hand, a merger occurs frequently by the lines of a small target and the lines of the background by the segmentation, on the other hand, a confusion of lines in the object tracking is quite possible, because the object either has no or an unspecified evaluable image structure. Moreover, the length and the orientation of the line segments of a small target object in the phases 1 and 2 could be changed considerably by an incorrect segmentation from image to image. The same happens with the contours. For this reason, the contour segmentation or line segmentation is generally used only for the extraction of meaningful image structures for the interpretation of object images.

Peaks (local maxima) and valleys (local minima) in a gray scale image can be extracted with a so-called monotony operator as image features. Blobs are formed by grouping adjacent image features which have the same characteristics (local maximum or local minimum). On average, the monotony operator extracted thousands of blobs per image. The local stable relationships between the blobs over several subsequent images determine the clustering of blobs. The relative arrangement (constellation) of blobs in a cluster, which is known as an artificial mesh, can be used to characterize an individual object and is therefore used for target tracking [13]. The sizes of the blobs are dependent on the mask size of the selected band pass filter. By using different band-pass filtering, a hierarchical constellation of blobs of a target image can be built. This constellation provides additional valuable information for the target tracking. This blob-based method is consistently well useable for object detection from phase 1 to phase 3. The artificial constellation of blobs is appropriate for the recognition of learned artificial constellations, but is not sufficient for the interpretation of a target object, because still no CAD data of target objects is present in this form.

For many tasks in image analysis, a gray-scale image is first binarized using the method of foreground-background separation. It creates a black and white image (blob image) in the expectation that the white areas represent target images and the black areas the backgrounds. The binarization is performed using one or more gray value thresholds. Crucial for a good binarization according to the respective task is an appropriate threshold setting. Such an adapted threshold determination is generally not trivial, if the gray value ranges for the relevant image regions are not known in advance. The bright and dark image regions may be close to each other and their gray values may have an arbitrary distribution. If such an image is binarized with a lower gray value threshold, the foreground gets too many pixels. If the image is binarized with a higher threshold, the target loses its signature or pixels. A new method which is introduced in the next chapter for the foreground-background separation has been developed and successfully implemented to adaptively find the desired threshold value.

3. ADAPTIVE THRESHOLD OPERATION FOR TARGET DETECTION

A new method for the extraction of (moving) target images in image sequences was developed through an evaluation of gray level cumulative histograms. The data sets for the extraction of dark target images were prepared as follows by the well-known calculation formulas for cumulative histogram which is a so-called “forwards cumulative histogram (FCH)” method:

$$M_f = \sum_{c=G_{\text{ms}}}^{f} h_c$$

(1)

Figure 2: Airplane signatures in phase 1 (left), in phase 2 (middle) and in phase 3 (right).
where $h_c$ is the frequency of the gray value $c$, $G_{\text{min}}$ denotes the minimum gray value in image as the lower bound, and $M_f$ represents the cumulative number of observations from the minimum gray value ($G_{\text{min}}$) up to the specified gray value $f$.

Using the same idea, a so-called “backwards cumulative histogram (BCH)” method was defined for the generation of data sets for the extraction of bright target images and has been used as follows:

$$M_b = \sum_{c=b}^{G_{\text{max}}} h_c$$  \hspace{1cm} (2)

where $h_c$ is the frequency of the gray value $c$, $G_{\text{max}}$ denotes the maximum gray value in image as the upper bound, and $M_b$ represents the cumulative number of observations from the maximum gray value ($G_{\text{max}}$) down to the specified gray value $b$.

Depending on the individual task, the two approaches, FCH and BCH, can be used to prepare data sets for the extraction of (moving) target images. In order to explain how the thresholds were calculated, the method with “backwards cumulative histogram” is used in this paper. The proposed threshold operator can be directly applied to the whole image or only on selected areas of the image after a pre-segmentation. There is no restriction with regard to the application of those threshold operators, FCH and BCH, to images that have been obtained by picture-giving sensors.

Thermal infrared sensors record temperature distributions of a scene. Heat radiating objects such as mammals, mobile high value targets and industrial plants produce bright spots in the IR image. IR cameras provide good images even in darkness and have numerous applications in all kinds of surveillance tasks. The threshold operator as a detector for the detection of (moving) targets in IR image sequences is particularly powerful. To demonstrate the function of the proposed threshold operators IR images were used in the present study.

![Flowchart of image processing](image.png)

**Figure 3**: The flowchart of image processing through the use of the backwards/forwards cumulative histogram for the foreground-background separation (binarization).

How can we find the suitable threshold values from the data sets to extract the target images or to separate the foreground and background pixels? What kind of information is hidden behind the data? A lot of methods for the foreground-background separation have been developed recently and studied intensively [14]. Our answer to the questions was found by the analysis of the abrupt change in the FCH and BCH. Figure 3 shows the flowchart of image processing for the foreground-background separation. The calculated adaptive threshold value $S$ from the image at the time $t$ was used to separate the image at the time $t$ into foreground (targets) and background. The targets were assigned to the maximum gray value and the background to the minimum gray value. An adaptive threshold value $S(t)$ was calculated using the new method for the separation of foreground and background through the following steps:

1. Let $i(t) = G_{\text{max}}(t)-1$, $i(t) \in [G_{\text{min}}(t), G_{\text{max}}(t)]$  \hspace{1cm} (3)

   where $G_{\text{min}}(t) < G_{\text{max}}(t)$, $G_{\text{max}}(t) \neq 255$, and $G_{\text{min}}(t) \neq 0$.

2. Calculate the sample mean at gray value $i(t)$, as below:

   $$\overline{h}_i(t) = \frac{1}{G_{\text{max}}(t)-i(t)+1} \sum_{c=i(t)}^{G_{\text{max}}(t)} h_c(t)$$  \hspace{1cm} (4)
where \( h_c(t) \) is the frequency at gray value \( c \).

3. Calculate the standard deviation at gray value \( i(t) \), as below:

\[
 s_i(t) = \frac{1}{n(t) - 1} \sum_{c=i(t)}^{G_{\text{max}}(t)} \left[ h_c(t) - \bar{h}_c(t) \right]^2
\]

where \( n(t) = G_{\text{max}}(t) - i(t) + 1 \), \( n(t) > 1 \) is the number of samples.

4. Calculate the estimated frequency at the gray value \( i(t) \), as below:

\[
 \tilde{h}_i(t) = \bar{h}_i(t) + f \cdot s_i(t)
\]

where \( f \) is a given parameter.

5. If \( \tilde{h}_i(t) \leq \bar{h}_i(t) \), then \( i(t) = i(t) - 1 \) and go back to step 2

If \( \tilde{h}_i(t) > \bar{h}_i(t) \), then step 6

6. The adaptive threshold value \( S(t) = i(t) + 1 \).

The given parameter \( f \) with the mean \( \bar{h}_i(t) \) and standard deviation \( s_i(t) \) determines the size of the estimated frequency at the gray value \( i(t) \). This parameter was chosen so that an increase of blob size may show faster than the expected blob size at the gray value \( i(t) \).

Adjacent foreground pixels were clustered according to segmentation as blobs. Blobs which have less or more than a predicted number of pixels are not considered for further evaluation. The calculation of the prediction will be described in the next chapter. The center of gravity, the size and the shape of a blob were calculated for the description of the blob. The shape of a blob was represented by an ellipse or a rectangle. The length of major axis and the length of the minor axis of an ellipse were used as one of the measures for the pre-selection of target types.

4. SELECTIVE OBJECT TRACKING BY EVALUATION OF TRAJECTORIES

The automatic tracking of object images was performed by comparing the description of the extracted image features like blobs between image pairs. Image features between consecutive image pairs, which lie in a predefined image area and have a similar description, were assigned to each other and were marked graphically by connecting lines. Because of the previous movement and the current size of the object, one or several possible continuations can be detected; they are passed into the next iteration of the blob tracking. The ambiguity of the assignment can be eliminated by long-term evaluation, for example, about 3 or more image pairs. Each object image was processed individually. Size-changes of a target object image were evaluated (see Figure 1):

- for determination of its phase change,
- for prediction and reduction of its search area,
- for elimination of noise.

Depending on the task of surveillance, a selective tracking of object images is sometimes very important; not only in order to suppress noise, but also for the tracking of abnormal movements of the target objects. An alarm signal should be triggered, for instance, when a traffic offender is driving on a one-way street in the wrong direction. Thus, a global movement analysis is just as important as a local motion analysis, especially for debunking the intentions of the object motion [13].

The long-term changes of the blob size of a target object image were evaluated empirically for the determination of the phase change. For example, phase 1 begins with a very small blob size (blob1) after the first foreground-background separation. The condition for a phase change is:
where \( \text{fac} = 10 \) is used in the present study, and \( v = t, t + 1, ..., n \). The expression \( [(\text{blob}_v) / (\text{blob}_t \times \text{fac})] \) can be assigned for instance with values in respect to the observer-target distance.

Due to the relative motion between the camera and moving objects as well as perspective distortions, the movement patterns could change abruptly. This circumstance leads to the fact that some estimation techniques fail in such situations. Therefore, a neighborhood classifier with prediction was used in the present work. Figure 4 shows the complete trajectories of the object tracking from point target to target with well evaluable structure. The calculated trajectories contain the following derived information:

- the temporal description of the tracked object images like positions, size, shape and so on,
- the dynamic evolution of the sizes of the tracked object images,
- the descriptions of the movement like directions, type and so on,
- trigger signal, for example, for the parallel use of contours or line-based methods to obtain descriptions of target object images.

Figure 4: Object tracking by analysis of trajectories in phase 1 (red), phase 2 (cyan) and phase 3 (yellow). The start phase is corresponding to the red line in figure 4. The end phase is corresponding to the yellow line in figure 4. The complex form of the trajectory is based on the additional movements of the camera-carrying platform.

5. INTERPRETATION OF OBJECT IMAGES

A target object could be imaged by its approach to the sensor platform, an evaluable image structure with a larger expansion on the sensor plane in phase 3. This allows object recognition by means of obtained object descriptions. Object recognition is one of the most investigated tasks in the image processing. Different approaches have been developed and implemented [15, 16]. A model-based identification, which is based on skeletons of object images, is exemplarily shown in this paper with three examples from a data base of a search system. Model data for example "aircraft" can be called from the search system. A part of the extensive data base is the section of fixed-wing aircrafts, which are divided in the subclasses such as straight-wing, swept-wing, and delta-wings and so on. Further subdivisions are possible, for instance with distinctive details such as the number of aircraft engines.

The following steps were used to prepare the interpretation of an airplane image:

1. skeletonization [17, 18] of the segmented image of the airplane,
2. concatenation of the skeleton points, and
3. calculation of the description of the airplane from the concatenations with the aid of the straight-line regression
Thus, reference models were found from the search system according to the calculated descriptions. Possible intersections of the straight line segments and angles between two adjacent line segments were calculated (see Figure 5 and Figure 6).

![Symbolic description of the reference models using the straight line segments (red) and intersections (green). Right: swept-wing (Boeing B52); middle: delta-wing (MIG-21); left: straight-wing (Anatonov-70).](image1)

By comparing the points of intersection and the angle of the straight segments of a tracked aircraft with the above-mentioned reference models, a pre-selection for the interpretation of the target object was carried out. The search space was further reduced significantly by further comparison with detailed object descriptions, for example, the number and position of the aircraft engines. The movement behavior of a tracked target object was derived from the tracking results and can be used as additional information for the interpretation of the target object, when dynamic information such as motion models of the target objects is known.

![The left image shows a currently tracked fixed-wing aircraft. The right image shows the extracted target image, which was segmented using adaptive threshold from BCH, including their "skeleton" (red).](image2)

6. CONCLUSIONS

In this paper, we have presented a modular system concept for tracking object images in image sequences and the possibilities for the selection of methods. The cumulative histogram is used to find out forwards or backwards the adaptive threshold value for the detection of target object images in all phases. Using the deduced information of the calculated trajectories, the segmentation is more precise and goal-oriented. The tracked objects can be identified by using a model-based classification. The presented adaptive and selective multi-target tracker is suitable for all application areas where analysis of movements plays a role. In the future, the proposed method will be used for color images.

REFERENCES