

Chapter 2

Moving Object Detection Approaches, Challenges and Object Tracking

2.1 Object Detection from Video

In a video there are primarily two sources of information that can be used for detection and tracking of objects: visual features (e.g. color, texture and shape) and motion information. Robust approaches have been suggested by combining the statistical analysis of visual features and temporal analysis of motion information. A typical strategy may first segment a frame into a number of regions based on visual features like color and texture, subsequently merging of regions with similar motion vectors can be performed subject to certain constraints such as spatial neighborhood of the pixels.

A large number of methodologies have been proposed by a number of researchers focusing on the object detection from a video sequence. Most of them make use of multiple techniques and there are combinations and intersections among different methodologies. All these make it very difficult to have a uniform classification of existing approaches.

This chapter broadly classifies the different approaches available for moving object detection from video.

2.1.1 Background Subtraction

Background subtraction is a commonly used technique for motion segmentation in static scenes [1]. It attempts to detect moving regions by subtracting the current image pixel-by-pixel from a reference background image. The pixels where the difference is above a threshold are classified as foreground. The creation of the background image is known as background modeling (e.g. by averaging images over time in an initialization period). After creating a foreground pixel map, some morphological post processing operations such as erosion, dilation and closing are performed to reduce the effects of noise and enhance the detected regions. The reference background is updated with new images over time to adapt to dynamic scene changes.

There are different approaches to this basic scheme of background subtraction in terms of foreground region detection, background maintenance and post processing.

In [2] Heikkila and Silven uses the simple version of this scheme where a pixel at location (x, y) in the current image I_t is marked as foreground if $|I_t(x, y) - B_t(x, y)| > Th$ is satisfied; where Th is a predefined threshold.

The background image B_T is updated by the use of an Infinite Impulse Response (IIR) filter as follows:

$$B_{t+1} = \alpha I_t + (1 - \alpha) B_t$$

The foreground pixel map creation is followed by morphological closing and the elimination of small-sized regions.

Although background subtraction techniques perform well at extracting most of the relevant pixels of moving regions even they stop, they are usually sensitive to dynamic changes when, for instance, stationary objects uncover the background (e.g. a parked car moves out of the parking lot) or sudden illumination changes occur.

2.1.2 Temporal Differencing

In temporal differencing, moving regions are detected by taking pixel-by-pixel difference of consecutive frames (two or three) in a video sequence. Temporal differencing is the most common method for moving object detection in scenarios where the camera is moving. Unlike static camera segmentation, where the background is comparably stable, the background is changing along time for moving camera; therefore, it is not appropriate to build a background model in advance. Instead, the moving object is detected by taking the difference of consecutive image frames $t-1$ and t . However, the motion of the camera and the motion of the object are mixed in the moving camera. Therefore in some techniques the motion of the camera is estimated first.

This method is highly adaptive to dynamic changes in the scene as most recent frames are involved in the computation of the moving regions. However, it generally fails to detect whole relevant pixels of some types of moving objects. It also wrongly detects a trailing regions as moving object (known as ghost region) when there is an object that is moving fast in the frames. Detection will also not be correct for objects that preserve uniform regions.

A sample object for inaccurate motion detection is shown in Fig. 2.1. The mono colored region of the human body (portions of legs) makes the temporal differencing algorithm to fail in extracting all pixels of the human's moving body. The white region at the left outer contour of the human body represents the ghost region.

This method also fails to detect the objects that have stopped in the scene. This occurs due to the reason that the last frame of the video sequence is treated as the reference which is subtracted from the current frame. Additional methods should be



Fig. 2.1 Temporal frame differencing. (a) Present Frame (PF) (b) Previous Frame (Prev) (c) Result=PF-Prev

adopted in order to detect stopped objects. This problem may be solved by considering a background model generated taking frames that came earlier in the sequence and are temporally distant from the present frame; this will incorporate other problems in detecting recent changes in the scene).

A two-frame differencing method is presented by Lipton et al. [3] where the pixels that satisfy the following equation are marked as foreground.

$$|I_t(x, y) - I_{t-1}(x, y)| > Th$$

In order to overcome shortcomings of two frame differencing in some cases, three frame differencing can be used [4]. For instance, Collins et al. developed a hybrid method that combines three-frame differencing with an adaptive background subtraction model [5]. The hybrid algorithm successfully segments moving regions in video without the defects of temporal differencing and background subtraction.

2.1.3 Statistical Approaches

Statistical characteristics of individual pixels have been utilized to overcome the shortcomings of basic background subtraction methods. These statistical methods are mainly inspired by the background subtraction methods in terms of keeping and dynamically updating statistics of the pixels that belong to the background image process. Foreground pixels are identified by comparing each pixel's statistics with that of the background model. This approach is becoming more popular due to its reliability in scenes that contain noise, illumination changes and shadows [4].

The statistical method proposed by Stauffer and Grimson [6] describes an adaptive background mixture model for real-time tracking. In this approach, every pixel is separately modeled by a mixture of Gaussians which are updated online by incoming image data. In order to detect whether a pixel belongs to a foreground or background process, the Gaussian distributions of the mixture model for that pixel are evaluated.

The W4 [7] system uses a statistical background model where each pixel is represented with its minimum (Min) and maximum (Max) intensity values and

maximum intensity difference (Diff) between any consecutive frames observed during initial training period where the scene contains no moving objects. A pixel in the current image I_t is classified as foreground if it satisfies:

$$|Min(x, y) - I_t(x, y)| > Diff(x, y) \text{ or } |Max(x, y) - I_t(x, y)| > Diff(x, y)$$

After thresholding, a single iteration of morphological erosion is applied to the detected foreground pixels to remove one-pixel thick noise. In order to grow the eroded regions to their original sizes, a sequence of erosion and dilation is performed on the foreground pixel map. Also, small-sized regions are eliminated after applying connected component labeling to find the regions. The statistics of the background pixels that belong to the non-moving regions of current image are updated with new image data.

2.1.4 Optical Flow

Optical flow methods [8–10] make use of the flow vectors of moving objects over time to detect moving regions in an image. In this approach, the apparent velocity and direction of every pixel in the frame have to be computed. It is an effective but time consuming method. Background motion model, which serves to stabilize the image of the background plane, can be calculated using optic flow. Independent motion can also be detected by this approach as either in the form of residual flow or by the flow in the direction of the image gradient which is not predicted by the background plane motion. This method can detect motion in video sequences even from a moving camera and moving background, however, most of the optical flow methods are computationally complex and cannot be used in real-time without specialized hardware.

2.2 Challenges

Object detection and tracking remains an open research problem even after research of several years in this field. A robust, accurate and high performance approach is still a great challenge today. The difficulty level of this problem highly depends on how one defines the object to be detected and tracked.

If only a few visual features (e.g. color) are used as representation of an object, it is not so difficult to identify the all pixels with same color as the object. However, there is always a possibility of existence of another object or background with the same color information. Moreover, the change of illumination in the scene does not guarantee that the color will be same for the same object in all the frames. This leads to inaccurate segmentation based on only visual features (e.g. color). This

type of variability changes is quite obvious as video objects generally are moving objects. The images of an object may change drastically as it moves from one frame to another through the field of view of a camera. This variability comes from three principle sources namely variation in target pose or deformations, variation in illumination and partial/full occlusion of the target [11].

The typical challenges of background subtraction in the context of video surveillance have been listed below:

2.2.1 Illumination Changes

It is desirable that background model adapts to gradual changes of the appearance of the environment. For example in outdoor settings, the light intensity typically varies during day. Sudden illumination changes can also occur in the scene. This type of change occurs for example with sudden switching on/off a light in a indoor environment. This may also happen in outdoor scenes (fast transition from cloudy to bright sunlight). Illumination strongly affects the appearance of background, and cause false positive detections. The background model should take this into consideration.

2.2.2 Dynamic Background

Some parts of the scenery may contain movement (a fountain, movements of clouds, swaying of tree branches, wave of water etc.), but should be regarded as background, according to their relevance. Such movement can be periodical or irregular (e.g., traffic lights, waving trees). Handling such background dynamics is a challenging task.

2.2.3 Occlusion

Occlusion (partial/full) may affect the process of computing the background frame. However, in real life situations, occlusion can occur anytime a subject passes behind an object with respect to a camera.

2.2.4 Clutter

Presence of background clutter makes the task of segmentation difficult. It is hard to model a background that reliably produces the clutter background and separates the moving foreground objects from that.

2.2.5 *Camouflage*

Intentionally or not, some objects may poorly differ from the appearance of background, making correct classification difficult. This is especially important in surveillance applications. Camouflage is particularly a problem for temporal differencing methods.

2.2.6 *Presence of Shadows*

Shadows cast by foreground objects often complicate further processing steps subsequent to background subtraction. Overlapping shadows of foreground regions for example hinder their separation and classification. Researchers have proposed different methods for detection of shadows.

2.2.7 *Motion of the Camera*

Video may be captured by unstable (e.g. vibrating) cameras. The jitter magnitude varies from one video to another.

2.2.8 *Bootstrapping*

If initialization data which is free from foreground objects is not available, the background model has to be initialized using a bootstrapping strategy.

2.2.9 *Video Noise*

Video signal is generally superimposed with noise. Background subtraction approaches for video surveillance have to cope with such degraded signals affected by different types of noise, such as sensor noise or compression artifacts.

2.2.10 *Speed of the Moving Objects and Intermittent Object Motion*

The speed of the moving object plays an important role in its detection. If the object is moving very slowly, the temporal differencing method will fail to detect the portions of the object preserving uniform region. On the other hand a very fast moving object leaves a trail of ghost region behind it in the detected foreground mask.

Intermittent motions of objects cause ‘ghosting’ artifacts in the detected motion, i.e., objects move, then stop for a short while, after which they start moving again. There may be situations when a video includes still objects that suddenly start moving, e.g., a parked vehicle driving away, and also abandoned objects.

2.2.11 Challenging Weather

Detection of moving object becomes a very difficult job when videos are captured in challenging weather conditions (winter weather conditions, i.e., snow storm, snow on the ground, fog), air turbulence etc.

2.3 Object Tracking

Object detection in videos involves verifying the presence of an object in a sequence of image frames. A very closely related topic in video processing is possibly the locating of objects for recognition – known as object tracking.

There are a wide variety of applications of object detecting and tracking in computer vision—video surveillance, vision-based control, video compression, human-computer interfaces, robotics etc. In addition, it provides input to higher level vision tasks, such as 3D reconstruction and representation. It also plays an important role in video databases such as content-based indexing and retrieval.

Popular methods of object tracking are summarized below.

2.3.1 Mean-shift

Mean-shift is an approach [12] to feature space analysis. This is an iterative approach which shifts a data point to the average of data points in its neighborhood similar to clustering. It has found its application in visual tracking [13, 14] and probability density estimation.

Mean Shift tracking uses fixed color distribution. In some applications, color distribution can change, e.g., due to rotation in depth. Continuous Adaptive Mean Shift (CAMSHIFT) [15]. CAMSHIFT can handle dynamically changing color distribution by adapting the search window size and computing color distribution in a search window.

2.3.2 KLT

The Kanade–Lucas–Tomasi (KLT) feature tracker is basically a feature extraction approach. It is based on the early work of Lucas and Kanade on an iterative image registration technique [16] that makes use of spatial intensity gradients to guide the search towards the best match. The method was developed fully by Tomasi and Kanade [17].

2.3.3 Condensation

A new approach the Condensation algorithm (Conditional Density Propagation) [18] which allows quite general representations of probability. Experimental results show that this increased generality does indeed lead to a marked improvement in tracking performance. In addition to permitting high-quality tracking in clutter, the simplicity of the Condensation algorithm also allows the use of non-linear motion models more complex than those commonly used in Kalman filters.

2.3.4 TLD

TLD [19] is an award-winning, real-time algorithm for tracking of unknown objects in video streams. The object of interest is defined by a bounding box in a single frame. TLD simultaneously tracks the object, learns its appearance and detects it whenever it appears in the video. The result is a real-time tracking that often improves over time. Tracking objects through highly cluttered scenes is difficult. Tracking becomes a challenging task under the following agile moving objects, in the presence of dense background clutter, probabilistic algorithms are essential. Algorithms based on Kalman filter, have been limited in the range of probability distributions they represent.

2.3.5 Tacking Based on Boundary of the Object

Boundary-based approaches are also referred to as edge-based approaches rely on the information provided by the object boundaries. It has been widely adopted in object tracking because the edges (boundary-based features) provide reliable information which is not dependent upon the motion type or the shape of the objects. Usually, the boundary-based tracking algorithms employ active contour models like snakes and geodesic active contours. These models are based on minimization of energy or geometric features by evolving an initial curve under the influence of external potentials, while being constrained by internal energies.

- i. **Snakes:** Snakes introduced by Terzopoulos et al. [20] is a deformable active contour model used for boundary tracking. Snakes moves under the influence of image-intensity forces, subject to certain internal deformation constraints. In segmentation and boundary tracking problems, these forces relate to the gradient of image intensity and the positions of image features. One advantage of the force-driven snake model is that it can easily incorporate the dynamics derived from time-varying images. The snakes are usually parameterized and the solution space is constrained to have a predefined shape. So these methods require an accurate initialization step since the initial contour converges iteratively toward the solution of a partial differential equation. Considerable work has been done by several researchers to overcome the numerical problems associated with the solution of the equations of motion and to improve robustness in the presence of clutter and occlusions in the scenes.
- ii. **Geodesic Active Contour Models:** These models are not parameterized and can be used to track objects that undergo non-rigid motion. Caselles et al. presented [21] a three step approach which start by detecting the contours of the objects to be tracked. An estimation of the velocity vector field along the detected contours is then performed. Subsequently, a partial differential equation is designed to move the contours to the boundary of the moving objects. These contours are then used as initial estimates of the contours in the next image and the process iterates.

Bibliography

1. A. M. McIvor; "Background subtraction techniques", Proc. of Image and Vision Computing, 2000.
2. J. Heikkila and O. Silven, "A real-time system for monitoring of cyclists and pedestrians", Proc. of 2nd IEEE Workshop on Visual Surveillance, pp. 74–81, 1999.
3. A. J. Lipton, H. Fujiyoshi, and R.S. Patil; "Moving target classification and tracking from real-time video", Proc. of Workshop Applications of Computer Vision, pp. 129–136, 1998.
4. L. Wang, W. Hu, and T. Tan; "Recent developments in human motion analysis", Pattern Recognition, Vol. 36 (3), pp. 585–601, 2003.
5. R. T. Collins et al. A system for video surveillance and monitoring: VSAM final report. Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000.
6. C. Stauffer, W. E. L. Grimson; "Adaptive background mixture models for real-time tracking", IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Vol. 2, 1999.
7. I. Haritaoglu, D. Harwood, L. Davis; "W4: real-time surveillance of people and their activities", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22 (8), pp. 809–830, 2000.
8. N. Paragios, R. Deriche; "Geodesic active contours and level sets for the detection and tracking of moving objects", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22 (3), pp. 266–280, 2000.
9. L. Wixson, "Detecting Salient Motion by Accumulating Directionally-Consistent Flow", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22 (8), 2000.

10. Robert Pless, Tomas Brodsky and Yiannis Aloimonos, "Detecting Independent Motion: The Statistics of Temporal Continuity", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22 (8), 2000.
11. Gregory D. Hager and Peter N. Belhumeur, "Efficient Region Tracking With Parametric Models of Geometry and Illumination", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20 (10), pp. 1025–1039, 1998.
12. Y. Cheng, "Mean shift, mode seeking, and clustering", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 17 (8), pp. 790–799, 1998.
13. D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift", *IEEE Proc. on Computer Vision and Pattern Recognition*, pp. 673–678, 2000.
14. D. Comaniciu, V. Ramesh, and P. Meer, "Mean shift: A robust approach towards feature space analysis", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24 (5), pp. 603–619, 2002.
15. G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface", *Intel Technology Journal*, 2nd Quarter, 1998.
16. Bruce D. Lucas and Takeo Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", *International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.
17. Carlo Tomasi and Takeo Kanade, "Detection and Tracking of Point Features. Carnegie Mellon University Technical Report", CMU-CS-91-132, 1991.
18. Michael, Isard; D.Phil. Thesis, "Visual Motion Analysis by Probabilistic Propagation of Conditional Density", Oxford University, 1998.
19. Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-Learning-Detection," *Pattern Analysis and Machine Intelligence*, 2011.
20. M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active Contour Models. *Int'l J. Computer Vision*, Vol. 1, pp. 321–332, 1988.
21. V. Caselles and B. Coll, Snakes in Movement. *SIAM J. Numerical Analysis*, Vol. 33, pp. 2, 445–2, 456, 1996.

Moving Object Detection Using Background Subtraction

Shaikh, S.H.; Saeed, K.; Chaki, N.

2014, X, 67 p. 32 illus., Softcover

ISBN: 978-3-319-07385-9