

Enhanced Military Security Via Robot Vision Implementation Using Moving Object Detection and Classification Methods

Shanik Tiwari, Deepa Kumari, Deepika Gupta, Raina

CSE, Mody Institute of Technology and Science, Rajasthan, India)

Abstract

Robotics have been a staple of advanced manufacturing for over half a century. As robots and their peripheral equipment become more sophisticated, reliable and miniaturized, these systems are increasingly being utilized for military and law enforcement purposes. Mobile robotics play an increasingly important role in military matters, from patrol to dealing with potential explosives, measuring flow of refugees, monitoring peace treaties, providing secure regions around bases and assisting battlefield command and control.

In this paper, we present a real time system for human detection, tracking, and verification in such challenging environments. To deliver a robust performance, the system integrates several robot vision algorithms to perform its function: a human detection algorithm, an object tracking algorithm, and a motion analysis algorithm. To utilize the available computing resources to the maximum possible extent, each of the system components is designed to work in a separate thread that communicates with the other threads through shared data structures.

Keywords- Background subtraction methods, visual surveillance, human detection algorithms

1. INTRODUCTION

Since background subtraction is being implemented on a wide range of hardware—and thus within a wide range of computational budgets—I chose to implement methods of varying complexity;

- Low-complexity, using the frame difference method,
- Medium complexity, using the approximate median method, and
- High-complexity, using the Mixture of Gaussians method

1.1 Frame Difference

Frame difference is arguably the simplest form of background subtraction. The current frame is simply subtracted from the previous frame, and if the difference in pixel values for a given pixel is greater than a threshold T_s , the pixel is considered part of the foreground(1):

$$|frame_i - frame_{i-1}| > T_s \dots\dots\dots(1)$$

1.2 Approximate median

In median filtering, the previous N frames of video are buffered, and the background is calculated as the median of buffered frames. Then (as with frame difference), the background is subtracted from the current frame and thresholded to determine the foreground pixels. Median filtering has been shown to be very robust and to have performance comparable to higher complexity methods. However, storing and processing many frames of video (as is often required to track slower moving objects) requires an often prohibitively large amount of memory. This can be alleviated somewhat by storing and processing frames at a rate lower than the frame rate—thereby lowering storage and computation requirements at the expense of a slower adapting background.

1.3 Mixture of Gaussians

Among the high-complexity methods, two methods dominate the literature; Kalman filtering and Mixture of Gaussians (MoG). Both have their advantages, but Kalman filtering gets slammed in every paper for leaving object trails that can't be eliminated. As this seems like a possible deal breaker for many applications, I went with MoG. Also, MoG is more robust, as it can handle multi-modal distributions. For instance, a leaf waving against a blue sky has two modes—leaf and sky. MoG can filter out both. Kalman filters effectively track a single Gaussian,

and are therefore unimodal: they can filter out only leaf or sky, but usually not both.

In MoG, the background isn't a frame of values. Rather, the background model is parametric. Each pixel location is represented by a number (or mixture) of Gaussian functions that sum together to form a probability distribution function F :

$$F(i_t = \mu) = \sum_{i=1}^k \omega_{i,t} \cdot \eta(\mu, \sigma) \dots\dots\dots(2)$$

The mean μ of each Gaussian function, can be thought of as an educated guess of the pixel value in the next frame—we assume here that pixels are usually background. The weight and standard deviations of each component are measures of our confidence in that guess (higher weight & lower σ = higher confidence). There are typically 3-5 Gaussian components per pixel—the number typically depending on memory limitations.

A robust video processing system should be capable of dealing with movement through cluttered areas, objects overlapping in the visual field, shadows, lighting changes, effects of moving elements of the scene (e.g. swaying trees), slow-moving objects, and objects being introduced or removed from the scene. Traditional approaches typically fail in these general situations. Our goal is to find a simple method to realize the real-time detection of the moving objects with complex background, so as to apply it to the intelligent surveillance system.

This algorithm is simple and of low computational complexity. However, the object is hard to be precisely detected when both of the background and the foreground are complicated. Holes come with the object detected as well. Background subtraction is very adaptive to stable environments, but is extremely sensitive to dynamic scene changes due to lighting and extraneous events. Pfister [2] uses a multi-class statistical model for the tracked objects, but the background model is a single Gaussian distribution of color at each pixel. Haritaoglu [3] models the background scene by representing each pixel by three values: its minimum and maximum of intensity values and the maximum intensity difference between consecutive frames observed during this training period. This method can adapt to slow changes in the scene by recursively updating the model.

Moving object detection from video sequences is an important research portion, since it can be used in

many regions such as intelligent video surveillance, motion analysis, human-machine interface applications, and so on. Moving object detection is the basis of moving object identification and tracking.

Currently, the main detection algorithms include frame difference method, background subtraction method, optical flow method and statistical learning method. Optical flow method is the most complex algorithm. It spends more time than other methods, and statistical learning method needs many training samples and also has much computational complexity. These two methods are not suitable for real-time processing. Background subtraction method is extremely sensitive to the changes of light. Frame difference method is simple and easy to implement, but the results are not accurate enough, because the changes taking place in the background brightness cause misjudgment [4,5,6,7].

2. IMPLEMENTATIONS

2.1. Fast, median and slow moving objects

In this paper, we divide the relative speed into three levels, which are slow, median and fast. If the same object appears twice in the overlapping area between the consecutive frames, the relative speed is slow, as Fig.1(a) shows. The shadow represents overlapping region. If the detected object moves regularly, it will not appear again in next frame. If the first appearance is in the overlapping area and the second in the next overlapping region, we define this as fast, as Fig.1(c) shows. The case among the above two situations is median. The second appearance is at the place not overlapped with fore-and aft frames, as Fig.1(b) shows.

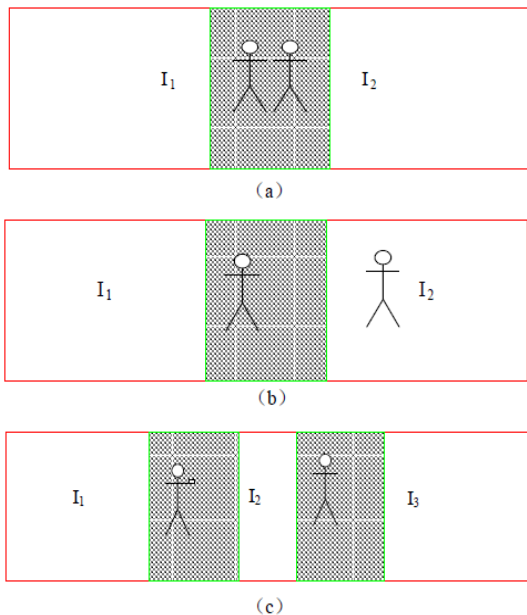


Fig 1. three level of relative speed: (a) slow, (b) median, (c) fast.

2.2. Human Detection Algorithm

The human detection algorithm used in our implementation was introduced in [1]. This algorithm searches for humans in the image by matching its edge features to a database of templates of human silhouettes. Examples of these templates are shown in Fig.2. The matching is done by computing the average Chamfer distance [8] between the template and the edge map of the target image area. The image area under consideration must be of the same size as the template. Let the template T be a binary image that is 0 everywhere except for the silhouette pixels where the value is 1, and let the Chamfer distance transform of the edge map of the target image area be denoted by C . The distance between a template T and the target image area I can be computed by (3):

$$D(I, T) = \frac{1}{|T|} \sum_i C_i T_i \dots\dots\dots(3)$$

where $|T|$ is the number of silhouette pixels in T , T_i is the pixel number i in T , and C_i is the Chamfer distance value at pixel number i in I . The smaller the value of the distance between the template and the target image area, the better the match between them. For efficient computations, a hierarchal structure,

which contains selected templates from the database, is built off- line.



Fig 2. example human silhouette templates

That limits the comparison to a few number of templates and accelerates the search in the database. Details of building the hierarchy and matching the edge features to templates are explained in [1].

3. PERSONAL IDENTIFICATION FOR VISUAL SURVEILLANCE

The problem of “who is now entering the area under surveillance” is of increasing importance for visual surveillance. Such personal identification can be treated as a special behavior-understanding problem. Human face and gait are now regarded as the main biometric features that can be used for personal identification in visual surveillance systems. In recent years, great progress in face recognition has been achieved. The main steps in the face recognition for visual surveillance are face detection, face tracking, face feature detection and face recognition. Usually, these steps are studied separately. Therefore, developing an integrated face recognition system involving all of the above steps is critical for visual surveillance.

3.1. Model-Based Methods

In model-based methods, parameters, such as joint trajectories, limb lengths, and angular speeds, are measured. Biomechanical models of walking and running are used to form a type of new anatomical model called a dynamically coupled oscillator, for the hip motion, and the structure and motion of the thigh and the lower leg.[9]

3.2. Statistical Methods

Statistical recognition techniques usually characterize the statistical description of motion image sets, and have been well developed in automatic gait recognition . A set of moment-based region features is used to recognize people and to predict the gender of an unknown person by his/her walking appearance.[10]

3.3. Physical-Parameter-Based Methods

Physical-parameter-based methods make use of geometric structural properties of a human body to characterize a person's gait pattern. The parameters used include height, weight, stride cadence and length, etc.[11]

4. CONCLUSION

Robotics help to meet challenges posed by the specter of urban terrorism. "Instead of having people get close to hazards such as unattended objects or car bombs, robots are used. If an operator concludes a dangerous object might explode, the robot could neutralize that object by shooting to detonate it. Vision systems enable a vehicle-mounted articulated arm to pick randomly oriented objects from its environment to monitoring motion in security applications. Mobile systems use a variety of sensory inputs including laser, sonar, as well as tactical inputs to enable a vehicle to dynamically interact with its environment. Vision is used for both navigation and by the end-of-arm tool to pick and place product in warehouses as well as in tending equipment in remote service applications. The need for robotics performing tasks in military applications will continue expanding. Military applications are a growing market for mobile robotics, and will continue to grow. The military is always pushing for new ways to keep their people out of harm's way.

REFERENCES

- [1] D. Gavrilu. Pedestrian detection from a moving vehicle. In *ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part II*, pages 37–49, London, UK, 2000. Springer-Verlag.
- [2] Wren C, Azabayejani A, Darrell T. and Pentland A. Pfunder: Real-time Tracking of the Human Body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997 : 780-785.
- [3] Haritaoglu I, Harwood D and Davis L. W4: Real-Time Surveillance of People and their Activities, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2000 : 809-830.
- [4] Wang Junqing, Shi Zelin, and Huang Shabai, "Detection of Moving Targets in Video Sequences". (Opto-Electronic Engineering, Dec 2005, pp. 5-8).
- [5] Ren Mingwu, and Sun Han, "A Practical Method for Moving Target Detection Under Complex Background". (Computer Engineering, Oct 2005, pp. 33-34).
- [6] Milan Sonka, Vaclav Hlavac, and Roger Boyle, "Image Processing, Analysis, and Machine Vision (Second Edition)", Posts & Telecom Press, Beijing, Sep 2003.
- [7] Zhang Yunchu, Liang Zize, Li En, and Tan Min, "A Background Reconstruction Algorithm Based on C-means Clustering for Video Surveillance", (Computer Engineering and Application, 2006, pp. 45-47).
- [8] H. G. Barrow. Parametric correspondence and chamfer matching: two new techniques for image matching. In *International Joint Conference on Artificial Intelligence*, pages 659–663, 1977.
- [9] D. Cunado, M. S. Nixon, and J. N. Carter, "Using gait as a biometric: via phase-weighted magnitude spectra," in *Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication*, 1997, pp. 95–102.
- [10] P. S. Huang, C. J. Harris, and M. S. Nixon, "Human gait recognition in canonical space using temporal templates," *Proc. Inst. Elect. Eng. (IEE) Vision Image and Signal Processing*, vol. 146, no. 2, pp. 93–100, 1999.
- [11] A. Johnson and A. Bobick, "A multi-view method for gait recognition using static body parameters," in *Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication*, 2001, pp. 301–311.