OSU Dept. of Computer and Information Science Technical Report OSU-CISRC-12/03-TR70.

Robust Detection of People in Thermal Imagery

James W. Davis Vinay Sharma
Dept. of Computer and Information Science
Ohio State University
Columbus OH 43210 USA
{jwdavis,sharmav}@cis.ohio-state.edu

Abstract

We present a new contour analysis technique to detect people in thermal imagery. Background-subtraction is first used to identify local regions-of-interest. Gradient information within each region is then combined into a contour saliency map. To extract contour fragments, a watershed-based selection algorithm is used. A path-constrained A search is employed to complete any broken contours, from which silhouettes are formed. Results using thermal video sequences demonstrate the capability of the approach to robustly detect people across a wider range of environmental conditions than is possible with standard approaches.*

1. Introduction

Intelligent activity analysis systems (e.g., for surveillance and monitoring) will be required to be *persistent* (continuous 24-7 operability) and *ubiquitous* (deployed anywhere and everywhere). These requirements provide several challenges for both fundamental and applied computer vision research. In this paper, we present a new contour-based technique for robust person detection using a persistent video camera under different environmental conditions.

Color and grayscale video cameras have an obvious outdoor limitation of daytime-only operation (not persistent). *Thermal* video cameras detect the amount of thermal radiation emitted/reflected from objects in the scene, and are applicable to both day and night scenarios. Therefore, they become a prime candidate for a persistent video system. As long as the thermal properties of the person are slightly different (higher or lower) than the background radiation, the person regions are detectable. Also, cast shadows do not appear unless the person is stationary for a long duration (shadow gradually cooling the background).

Though some classic problems are alleviated with the use of thermal cameras, they have their own unique challenges, including a lower signal-to-noise ratio and the "halo effect" that appears around very hot or cold objects with pyro-electric sensors (e.g., notice the strong dark and light halos around the people in Fig. 1).

Most of the previous strategies for detection in thermal imagery use "hot-spot" algorithms, relying on the assumption that the person (object) is much hotter than the surrounding environment. Though this is common in cooler nighttime environments (or during Winter), it is not universally true throughout the day or across different seasons of the year.

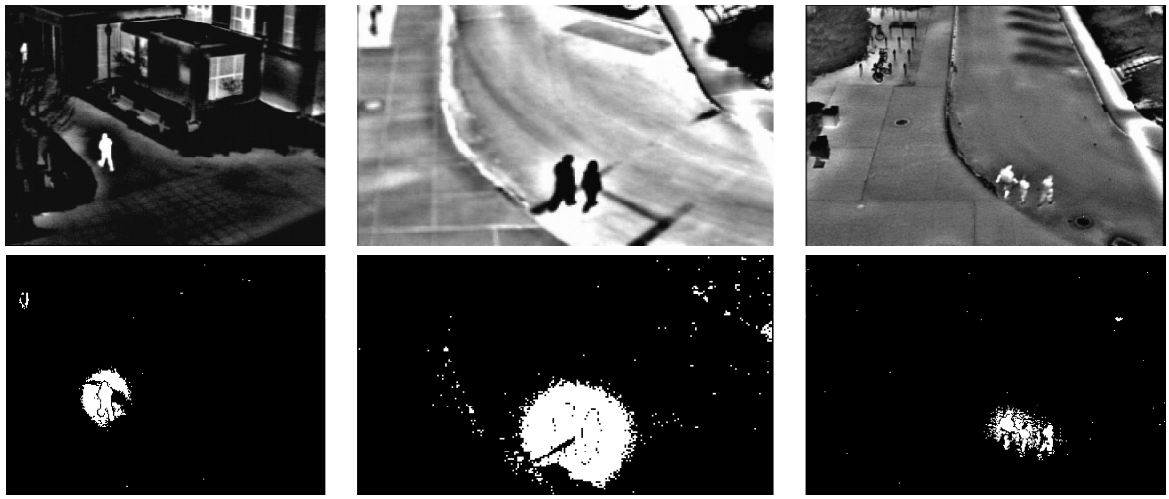
Our approach to detect people is to first use a standard background-subtraction technique to identify local regions-of-interest, each containing the person and surrounding thermal halo. The foreground and background gradient information within each region are then combined into a contour saliency map (highlighting the person boundary). Using a watershed-based algorithm, the gradients are thinned and thresholded into contour fragments. The remaining watershed lines are used as a guide for an A* search algorithm to connect any contour gaps. Finally, the closed contours are flood-filled to make silhouettes. As we will demonstrate, this approach enables silhouette extraction across a wider range of environmental conditions.

The remainder of this paper is described as follows. We begin with a review of related work (Sect. 2). Next we describe the contour approach for person detection (Sect. 3). Then we present experimental results (Sect. 4). Lastly, we conclude with a summary of the research and discuss future work (Sect. 5).

2. Related Work

Several methods have been proposed for identifying people in images without background-subtraction methodologies, including the direct use of wavelets [11], coarse-to-fine edge matching [4], and motion differencing [17, 9].

Most of the remaining person detection methods employ some form of background-subtraction using a single Gaussian background model [18] or a multi-modal Gaussian for-



(a) (b) (c)

Figure 1. Thermal imagery at different environmental conditions and background-subtraction results. (a) Winter afternoon. (b) Summer afternoon. (c) Summer night.

mulation [14]. Other approaches include the W4 method for detecting body parts and tracking [5], the three-stage (pixel/region/frame) Wallflower approach [15], a two-stage color and gradient technique [7], and a Markov chain Monte Carlo approach [20].

Recently, person detection using thermal imagery has been explored [6, 1], but these approaches rely heavily on the assumption that the person region always has a much hotter (brighter) appearance than the background (hot-spot techniques are commonly employed in thermal-based detection schemes [2, 3, 19]). We examine a new contour analysis technique for detecting people in thermal imagery that is most related to the color/gradient approach of [7].

3. Person Detection in Thermal Imagery

One issue with the use of pyro-electric thermal sensors is that the polarity (black/white) and strength of the thermal intensity in the person, halo, or background region can change dramatically across different environment conditions (see top row of Fig. 1). Clearly, hot-spot techniques or statistical background-subtraction techniques alone will be ineffective to detect the precise shape of the person (silhouette) under different conditions.

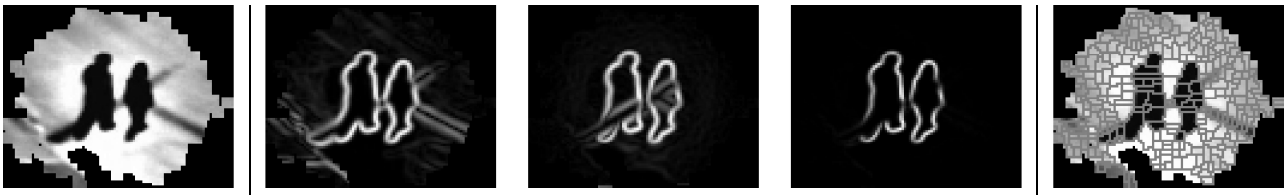
Two key observations are that 1) thermal halos fade smoothly into the image, and 2) stronger halos cause the edge/contour information of the person to become more pronounced. Based on these observations, we propose a new contour-based technique for person detection in thermal imagery that focuses on the extraction and completion of edge contours within the halo regions. Because the approach relies on contours, the method is expected to be more stable and robust across very different environmental conditions (including intensity polarity switches and different halo strengths).

3.1. Halo Detection

To find the regions-of-interest (ROIs) that contain the person (or people) and the surrounding halo, we apply a standard background-subtraction approach. We currently employ a univariate Gaussian model for each pixel location (derived from a collection of background images) to identify pixels in the foreground image that are statistically different from the background model using the Mahalanobis distance

$$\mathrm{D}(x,y) = \begin{cases} 1 & \frac{|I(x,y)-\mu(x,y)|}{\sigma(x,y)} > T \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

We show the background-subtraction results for three different environmental conditions in Fig. 1. Note that a statistical background-subtraction technique alone is ineffective at detecting the precise shape of the person. We then dilate and region-grow D to select the connected-component regions (the ROIs). A statistical threshold of 6 st. dev. (from 30 background images) and a 5×5 dilation mask were employed for all results in this paper.



(a) (b) (c) (d) (e)

Figure 2. Contour saliency and watershed. (a) ROI. (b) Foreground gradients. (c) Foreground-background gradient differences. (d) Contour saliency map. (e) Watershed overlay.

3.2. Contour Saliency Map

We next examine each ROI individually to separate the person (or people) from the surrounding halo. From the earlier observations regarding thermal halos, the gradient/edge strengths within the ROI can be used to identify the person boundary. For each ROI, we form a *contour saliency map* (CSM) by multiplying the normalized foreground gradient magnitudes with the normalized foreground-background gradient differences

$$\text{CSM} = \frac{\|\langle I_x, I_y \rangle\|}{\max} \times \frac{\|\langle (I_x - BG_x), (I_y - BG_y) \rangle\|}{\max} \quad (2)$$

For the ROI in Fig. 2.a, we show the normalized foreground gradient magnitudes in Fig. 2.b and the normalized foreground-background gradient differences in Fig. 2.c. To calculate the gradients, Gaussian derivative masks with $\sigma = .75$ were employed. We present the corresponding contour saliency map in Fig. 2.d. Notice that the non-person foreground gradients are suppressed (having small foreground-background gradient differences), and the non-person foreground-background gradient differences are also reduced (having low foreground gradients).

3.3. Watershed Analysis

Our next step is to extract the person contours from the saliency map. We make use of the watershed transform [16] as a unified method to both thin the saliency map into contours and to guide the completion of any contour fragments.

In the watershed technique, the image (saliency map) is considered as a topological surface (each pixel value corresponding to an elevation) being immersed into a lake. The water progressively floods basins corresponding to regions of local minima in the image. Regions separated by low ridges (weaker saliencies) merge earlier than others separated by higher ridges (stronger saliencies). When two regions merge (at a minima along the connecting contour), a watershed line (dam) is drawn along the flooding contour ridge. After the entire surface is immersed (flooded), the result is a partition of the original image given by the constructed watershed lines (see Fig. 2.e).

Using the method of "contour dynamics" [10], we next assign a *strength* value to each contour in the watershed image. Consider two regions A and B separated by a watershed contour c. Let $\text{MIN}_X = \min(\text{CSM}(X))$ be the region (or contour) minimum of X. The dynamic d of contour c is then defined as

$$d_c = \text{MIN}_c - \max(\text{MIN}_A, \text{MIN}_B) \quad (3)$$

The dynamic for a contour between two regions is assigned when the regions merge across that contour. However, as two regions could merge indirectly through a connecting region between them, not all contours will be assigned a dynamic. These remaining contour dynamics can be computed using the hierarchical technique of [8]. The approach produces sets of contours that are associated with the same dynamic value. The result tends to merge the strong outer contours of the object (person) region.

The contour dynamic values are usually thresholded to identify the object boundaries in the image. This approach works well in many cases, but when the object (person) has both strong and weak outer contours (see Fig. 2.d), the approach tends to suppress the set of strong object contours. To combat this problem, we individually examine each set of contours and assign pixel values related to the original saliency values.

For a contour set (of the same dynamic) $\hat{c} = \{c_i\}$, we first compute the median of the sub-contour c_i minimums as

$$\text{MED}_{\hat{c}} = \text{median}(\text{MIN}_{c_i}, \cdots, \text{MIN}_{c_n}) \quad (4)$$

Next, for each pixel in $\hat{c}$, we compare its original saliency value (maximum within a 3×3 neighborhood to deal with an 8-connected watershed construction) to $\text{MED}_{\hat{c}}$, and retain the minimum of the two values. This method results in a final thinned contour image suitable for thresholding, without the aforementioned contour dynamics suppression problem (see Fig. 3.a). To adaptively threshold the contours, we use K-means clustering (with low/medium/high-value clusters) of the thinned contour values and select the

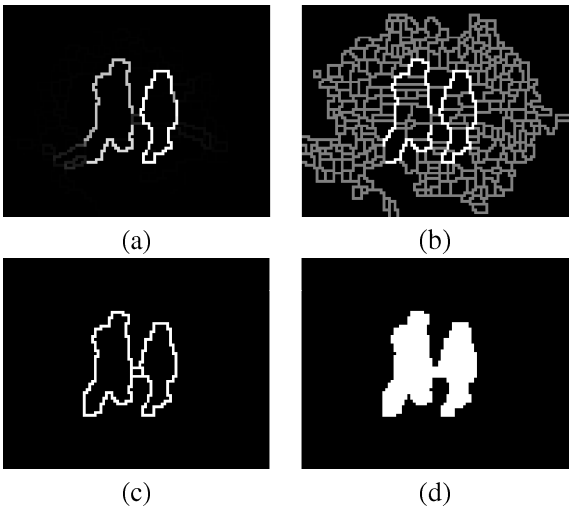


Figure 3. Silhouette forming process. (a) Thinned contour image. (b) Thresholded contours. (c) A* contour completion. (d) Flood-filled silhouettes.

threshold as the value between the bottom two clusters (see selected white pixels in Fig. 3.b).

3.4. Closing Contour Fragments

As there is no guarantee that the resulting binary contour image is complete (required for making silhouettes), we next identify and close any broken contour segments. Our approach is to employ a search algorithm at each gap to find another contour pixel along the outgoing watershed lines to close the gap. To ensure the best possible completion, we follow a two-stage strategy.

Stage-1:
In the first stage, we attempt to seal all contour gaps. Each contour fragment endpoint (found with a 3×3 neighborhood check) is forced to grow outward along the watershed lines to find the closest contour point. To find the optimal path, we employ the A* search algorithm [13] that minimizes the expected cost *through* the current pixel location to reach a contour point. The Euclidean distance from the current location to the closest contour point is employed as the heuristic cost function.

To minimize short "loop backs" and to force the path to grow outwards, we do not consider any of the points belonging to the endpoint's contour as a potential target point. Each gap completion is performed using only the original contour points (minus the current contour) so that the order of gap completion does not influence the result.

Stage-2:
In the second stage, we ensure that every completed contour segment in the image is part of a closed loop (for flood-filling). First, we region-grow along all contours to identify any contours that do not form a closed loop (e.g., a line connecting two closed circles is itself not closed). For each of those contours, we perform the A* search strategy to move along other watershed lines from one endpoint to the other[1]. To find solutions that create the minimum number of new contour pixels on the watershed lines, we give no penalty (step cost) in the A* algorithm for moving along existing contour pixels on the watershed (allowing a "free glide" along the contour). If no possible path exists between the endpoints, we default to a direct straight-line connection.

The result for the thresholded contours in Fig. 3.b after Stage-1 and Stage-2 is shown in Fig. 3.c. In this example, the bodies were joined since there is a very small gap between the people where the contours are fragmented. After the contour completion, a simple flood-fill operation can be employed to create silhouettes (see Fig. 3.d).

4. Experiments

We examined the proposed approach on several frames from the three thermal sequences shown in Fig. 1. The sequences were recorded at very different environmental conditions: Winter afternoon, Summer afternoon, and Summer night. Each sequence had a 30-frame background sequence for learning the statistical background model to identify the ROIs. For each of the sequences, we used the same parameter/threshold settings to demonstrate the applicability of the approach to different conditions.

Since we process each ROI separately, we additionally weighted each resulting silhouette in the image with a contrast measurement calculated from the ratio of the maximum foreground-background intensity difference within the silhouette region to the full intensity range of the background model. A final sensitivity threshold could easily be used to remove the minimal-contrast (noise) regions.

In Fig. 4, we show selected frames from the sequences and the resulting silhouettes using our approach. The images demonstrate the ability of the algorithm in many cases to separate multiple people contained within a single ROI. The small regions in the top left corner of each image pair in Fig. 4.a are a result of people being partially occluded by tree branches. Despite the very low thermal person-background differences (and low gradients) in Fig. 4.b, the algorithm is still able to detect reasonable portions of the people. Additionally, a small animal was detected moving

[1]For an un-closed contour with multiple endpoints (e.g., a three-prong connected contour), we compute a priority matrix [12] to select which two endpoints should be closed first, and then re-estimate the remaining un-closed contours.

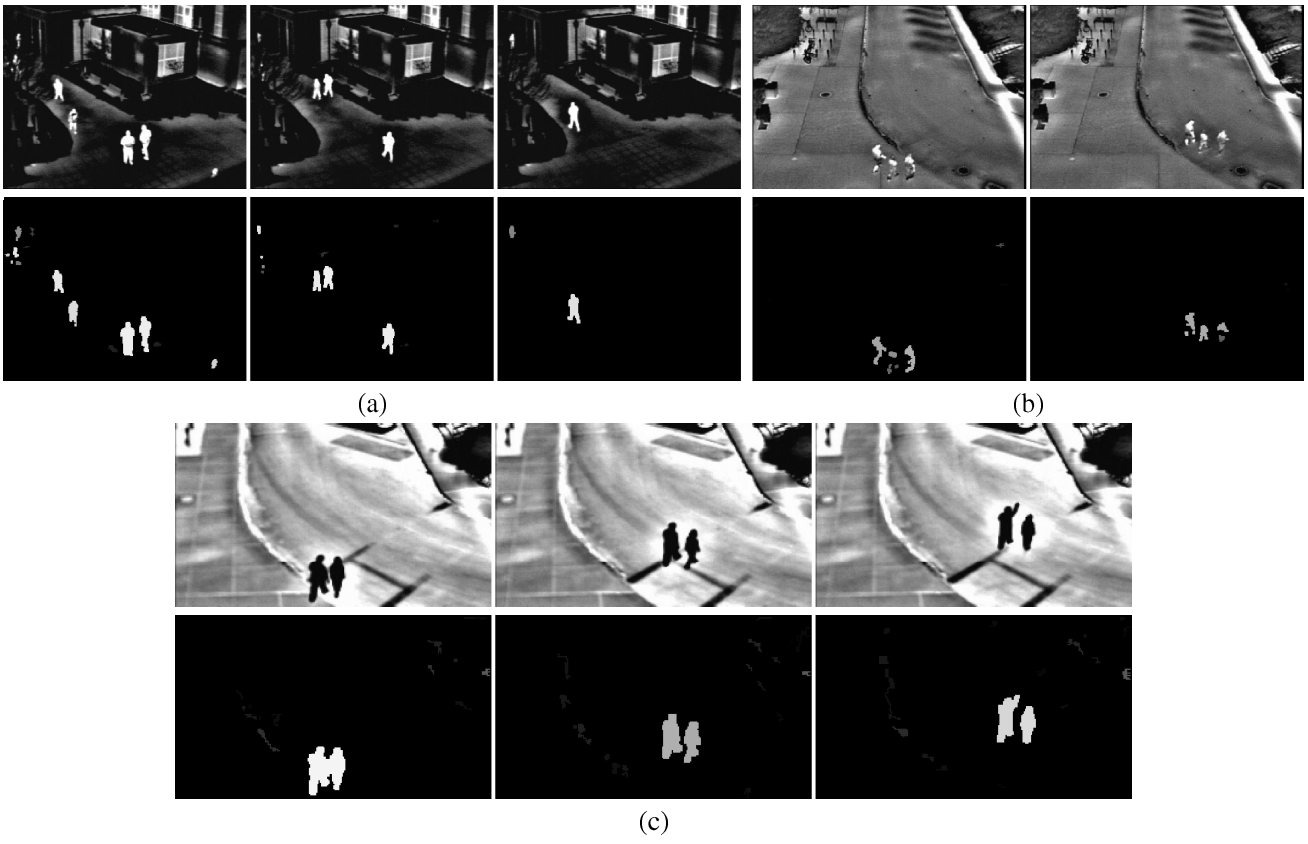


Figure 4. Example thermal images and resulting silhouette regions (contrast weighted).

down the stairs in the top-right corner of the left image pair. In Fig. 4.c, the two silhouettes are identified quite well, though in the first image the silhouettes are joined. This result is not unreasonable as the bodies are extremely close to one another. The overall results of the approach were encouraging and perform better than background-subtraction or hot-spot approaches alone. However, there were some problems that deserve mentioning.

As shown in Fig. 5.a, the thermal intensity of the people is similar to the background cross-walk line on the pavement. This causes a reduction of the contour saliency at the overlapping pixels and therefore sometimes resulted in the contour completion growing into the similar background region (see Fig. 5.b). This can be expected with similar foreground-background intensities. We could potentially employ shape-based tracking approaches to better estimate the silhouette of the person (but those approaches still need initialization).

In another situation, we noticed that a highly fragmented ROI could result in an over-completion of the contours. In Fig. 5.c, two people were highly-occluded by tree branches, which resulted in the contour fragments falsely connecting all three people in the ROI.

We also found in the Summer night sequence with a weak halo (see Fig. 1.c) that many gradients of the people are low and were deleted in the thinned contour image. One approach to combat this problem would be to estimate the strength of the halo in the ROI and boost the saliency map when the halo is weak. We show the positive effect of an exponent boost for the contour saliency map ($\text{CSM}^{\frac{1}{2}}$) in Fig. 5.e-f.

5. Summary

We presented a new approach to person detection in thermal imagery that is applicable over a wide range of environmental conditions (including day and night scenarios). Our approach is designed to handle the common problems with thermal imagery such as polarity switches and halo effects at different environmental settings. These issues render classic background-subtraction and hot-spot detection methods ineffective by themselves.

We first use a statistical background-subtraction technique to identify local regions-of-interest. The foreground and background gradient information within each region are

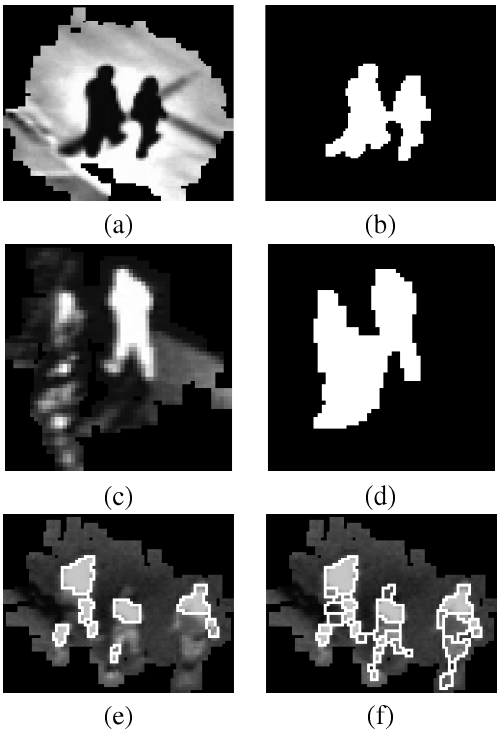


Figure 5. Problem images. (a) Person/background similarities (b) Contour extension into background. (c) Occlusion ROI. (d) Over-completion of contours. (e) Non-boosted contours. (f) Boosted contours.

then combined into a contour saliency map. A watershed-based algorithm is used to extract contours of the person from the saliency map. To close any contour fragments, an A* method constrained to the watershed paths is used. Lastly, the contours are flood-filled to produce silhouettes.

Experiments with three thermal video sequences recorded at very different environmental conditions showed promising results. To address the problems with the approach, we will examine methods for incorporating a multi-modal background model, estimating halo strengths for contour boosting, including motion into the saliency map, and separating multi-person silhouettes. As the approach is not limited to detecting only people, we will also examine the method with other objects of interest (e.g., vehicles).

References

[1] B. Bhanu and J. Han. Kinematic-based human motion analysis in infrared sequences. In *Proc. Wkshp. Applications of Comp. Vis.*, pages 208–212, 2002.

[2] B. Bhanu and R. Holben. Model-based segmentation of FLIR images. *IEEE Trans. Aero. and Elect. Sys.*, 26(1):2–11, 1990.

[3] A. Danker and A. Rosenfeld. Blob detection by relaxation. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 3(1):79–92, 1981.

[4] D. Gavrila. Pedestrian detection from a moving vehicle. In *Proc. European Conf. Comp. Vis.*, pages 37–49, 2000.

[5] I. Haritaoglu, D. Harwood, and L. Davis. W4: Who? When? Where? What? A real time system for detecting and tracking people. In *Proc. Int. Conf. Auto. Face and Gesture Recog.*, pages 222–227, 1998.

[6] S. Iwasawa, K. Ebihara, J. Ohya, and S. Morishima. Real-time estimation of human body posture from monocular thermal images. In *Proc. Comp. Vis. and Pattern Rec.*, pages 15–20. IEEE, 1997.

[7] O. Javed, K. Shafique, and M. Shah. A hierarchical approach to robust background subtraction using color and gradient information. In *Wkshp. on Motion and Video Computing*, pages 22–27. IEEE, 2002.

[8] C. Lemaréchal and R. Fjørtoft. Comments on geodesic saliency of waterhed contours and hierarchical segmentation. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 20(7):762–763, 1998.

[9] A. Lipton, H. Fujiyoshi, and R. Patil. Moving target classification and tracking from real-time video. In *Proc. Wkshp. Applications of Comp. Vis.*, 1998.

[10] L. Najman and M. Schmitt. Geodesic saliency of watershed contours and hierarchical segmentation. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 18(12):1163–1173, 1996.

[11] M. Oren, C. Papageorgiour, P. Sinha, E. Osuma, and T. Poggio. Pedestrian detection using wavelet templates. In *Proc. Comp. Vis. and Pattern Rec.*, pages 193–199. IEEE, 1997.

[12] K. Rangarajan and M. Shah. Establishing motion correspondence. *CVGIP: Image Understanding*, 54(1):56–73, 1991.

[13] S. Russell and P. Norvig, editors. *Artificial Intelligence: A Modern Approach.* Prentice Hall, 2003.

[14] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. Comp. Vis. and Pattern Rec.*, pages 246–252. IEEE, 1999.

[15] K. Toyama, B. Brumitt, J. Krumm, and B. Meyers. Wallflower: principals and practice of background maintenance. In *Proc. Int. Conf. Comp. Vis.*, pages 49–54, 1999.

[16] L. Vincent and P. Soille. Watershed in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 13(6):583–598, 1991.

[17] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In *Proc. Int. Conf. Comp. Vis.*, pages 734–741, 2003.

[18] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: real-time tracking of the human body. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 19(7):780–785, 1997.

[19] A. Yilmaz, K. Shafique, and M. Shah. Target tracking in airborne forward looking infrared imagery. *Image and Vision Comp.*, 21(7):623–635, 2003.

[20] T. Zhao and R. Nevatia. Stochastic human segmentation from a static camera. In *Wkshp. on Motion and Video Computing*, pages 9–14. IEEE, 2002.