Robust Detection of People in Thermal Imagery[*]

James W. Davis   Vinay Sharma
Dept. of Computer Science and Engineering
Ohio State University
Columbus OH 43210 USA
{jwdavis,sharmav}@cse.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, shadows do not appear unless the person is stationary for a long duration (shadow gradually cooling the background).
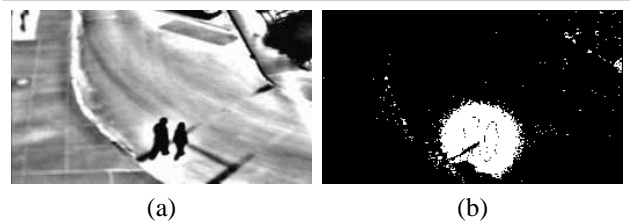
Figure 1. Thermal image (Summer afternoon) and background-subtraction results.

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, polarity inversion, and the "halo effect" that appears around very hot or cold objects with ferroelectric sensors (e.g., notice the strong bright halo around the people in Fig. 1.a).

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.

2. Related Work

Some template-based approaches have been used to detect people, but most methods employ some form of background subtraction. The most common approaches employ a single or multi-modal Gaussian formulation [10, 7]. Other approaches include the W4 method for detecting body parts and tracking [2], the three-stage (pixel/region/frame) Wallflower approach [8], a two-stage color and gradient technique [4], and a Markov chain Monte Carlo approach [11].

Recently, person detection using thermal imagery has been explored [3, 1], but these approaches rely heavily on the assumption that the person region always has a much hotter (brighter) appearance than the background. We examine a new contour analysis technique for detecting people in thermal imagery that is most related to the color/gradient approach of [4].

3. Person Detection in Thermal Imagery

One issue with the use of uncalibrated ferroelectric thermal sensors is that the polarity (black/white) and strength of the thermal intensity can change dramatically across different environment conditions. Additionally, these sensors produce thermal halos around objects having a large thermal contrast with the background (see white halo in Fig. 1.a). Hence, the use of a statistical background-subtraction techniques will be ineffective to detect the precise silhouette of the person (see Fig. 1.b).

Two key observations regarding thermal imagery 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 technique for person detection in thermal imagery that focuses on the extraction and completion of edge contours within the halo regions.

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 Gaussian background-subtraction approach to identify pixels in the foreground image that are statistically different from the background model using

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

We show the background-subtraction results ($T = 6$) for a sample image in Fig. 1. Note that a statistical background-subtraction technique alone is ineffective at detecting the precise shape of the person. We then use a 5×5 dilation operation and connected-components algorithm to extract the individual ROIs.

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 difference magnitudes

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

The CSM for the ROI in Fig. 2.a (using Gaussian derivative masks with $\sigma = .75$) is shown in Fig 2.b. Notice that the non-person gradients are suppressed in the CSM.

3.3. Watershed Analysis

Our next step is to extract the person contours from the CSM. We make use of the watershed transform [9] as a unified method to both thin the CSM and to guide the completion of any contour fragments.

After computing the watershed transform of the CSM, we employ a "contour dynamics" approach [6, 5] to assign a "strength" value to each contour segment and merge contours into groups. However, each group is assigned the dynamic of its weakest member, suppressing stronger contour segments. Therefore simple thresholding of the contour groups will not suffice.

To combat this problem, we first compute the median of the sub-contour c_i minimums (in group $\hat{c}$) as

$$\text{MED}_{\hat{c}} = \text{median}(\text{MIN}_{c_1}, \cdots, \text{MIN}_{c_n}) \quad (3)$$

Next, for each pixel in $\hat{c}$, we compare its original strength value 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 (see Fig. 2.c).

To adaptively threshold the contours, we use K-means clustering (with low/medium/high-value clusters) of the thinned contour values and select the threshold as the value between the bottom two clusters (see selected white pixels in Fig. 2.d).

3.4. Closing Contour Fragments

As there is no guarantee that the resulting binary contour image is a closed figure (required for making silhou-
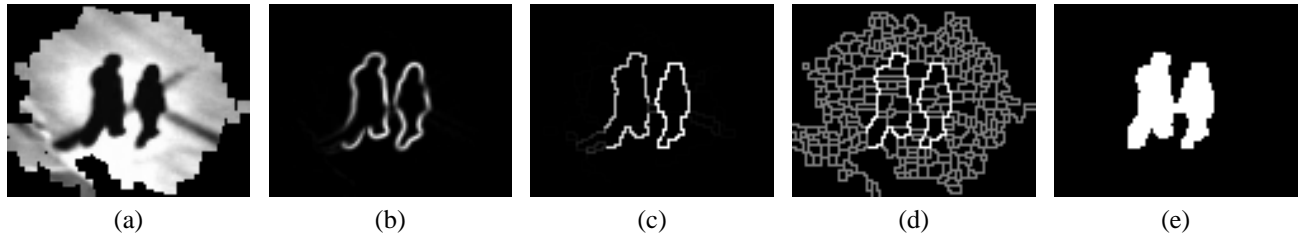
Figure 2. Detection stages. (a) Selected ROI. (b) CSM. (c) Thinned CSM. (d) Watershed with selected contours. (e) Flood-filled region of completed/closed contours.

ettes), we next identify and close any broken contour segments using the watershed lines. To ensure the best possible completion, we follow a two-stage strategy.

Stage-1: Gap Completion
Each contour fragment endpoint is forced to grow outward along the watershed lines towards other contour points. To find the optimal path, we employ the classic A* search algorithm 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.

Gap completion is performed using only the original contour points so that the order of gap completion does not influence the result. 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 own contour as potential targets.

Stage-2: Figure Closure
We next identify all contours that are not part of a closed loop (e.g., a line connecting two closed circles is itself not closed). For each of these contours, we perform the A* search strategy to move along other watershed lines from one endpoint to the other. 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. If no possible path exists between the endpoints, we default to a straight-line connection.

The flood-filled result for the thresholded contours in Fig. 2.d after Stage-1 and Stage-2 is shown in Fig. 2.e. In this example, the bodies were joined since there is a very small gap between the people where the contours are fragmented.

4. Experiments

We examined the proposed approach on several frames from three thermal sequences recorded at very different environmental conditions: Winter afternoon, Summer afternoon, and Summer night. Each sequence contained >300 frames and had a separate 30-frame background sequence for learning the statistical background model. 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. 3, we show a representative frame from each of the sequences and the resulting silhouettes. The images demonstrate the ability of the algorithm to separate multiple people contained within a single ROI. In Fig. 3.a-b, the person silhouettes are identified quite well. The small silhouette regions in the top left corner of Fig. 3.a are a result of the people being occluded by tree branches. Despite the very low thermal person-background differences in Fig. 3.c, the algorithm is still able to detect reasonable portions of the people. The overall results of the approach were encouraging and perform better than background-subtraction or hot-spot approaches alone.

There were some problems with the current approach that deserve mentioning. When the thermal intensity of the people is similar to the background, a reduction of the contour saliency occurs and sometimes results in the contour completion growing into the similar background region. In another situation, we noticed that a highly fragmented ROI could result in an over-completion creating false contours. Furthermore, when sequences have weak halos, the weaker gradients of the people can be deleted in the thinned CSM (see Fig. 4.a). One approach to combat this problem would be to estimate the strength of the halo and amplify the CSM when the halo is weak. We show the positive effect of an exponent amplification ($CSM^{\frac{1}{2}}$) in Fig. 4.b.
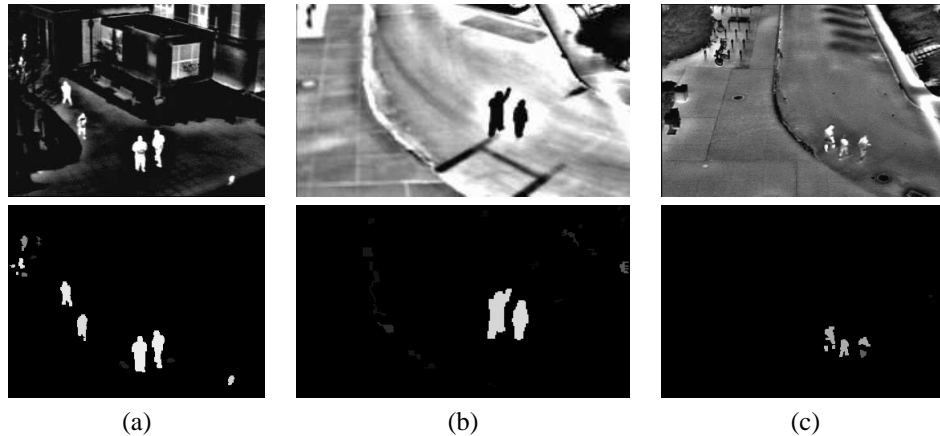
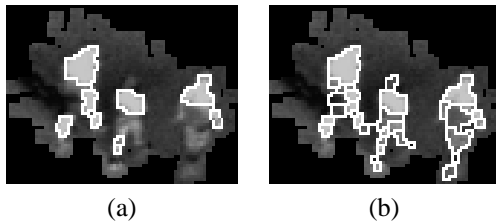**Figure 3. Thermal images (Winter, Summer) and resulting silhouette regions (contrast weighted).**

Figure 4. Problem image. (a) Using normal contours. (b) Using amplified contours.

5. Summary

We presented a new approach to person detection in thermal imagery that is applicable over a wide range of environmental conditions. Our approach is designed to handle the common problems with thermal imagery such as polarity inversion and halo effects. 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 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. In future work we will incorporate a multi-modal background model, estimate halo strengths for amplification, and include motion information.

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] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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