

# Background Estimation under Rapid Gain Change in Thermal Imagery

Hulya Yalcin  
Robotics Institute  
Carnegie Mellon University  
Pittsburgh, PA 15213  
hulya@cs.cmu.edu

Robert Collins  
CSE Department  
Penn State University  
University Park, PA 16802  
rcollins@cse.psu.edu

Martial Hebert  
Robotics Institute  
Carnegie Mellon University  
Pittsburgh, PA 15213  
hebert@cs.cmu.edu

## Abstract

*We consider detection of moving ground vehicles in airborne sequences recorded by a thermal sensor with automatic gain control, using an approach that integrates dense optic flow over time to maintain a model of background appearance and a foreground occlusion layer mask. However, the automatic gain control of the thermal sensor introduces rapid changes in intensity that makes this difficult. In this paper we show that an intensity-clipped affine model of sensor gain is sufficient to describe the behavior of our thermal sensor. We develop a method for gain estimation and compensation that uses sparse flow of corner features to compute the affine background scene motion that brings pairs of frames into alignment prior to estimating change in pixel brightness. Dense optic flow and background appearance modeling is then performed on these motion-compensated and brightness-compensated frames. Experimental results demonstrate that the resulting algorithm can segment ground vehicles from thermal airborne video while building a mosaic of the background layer, despite the presence of rapid gain changes.*

## 1. Introduction

Thermal imaging or infra-red (IR) photography, has provided some notable gains over conventional photographic methods and it is becoming more and more important in wide variety of problems ranging from firefighting to night-vision security and surveillance systems. Although, infrared radiation penetrates many obscurant aerosols far better than visible wavelengths and thermal cameras are applicable to both day and night scenarios, these beyond-visible-spectrum imaging modalities have their own challenges such as saturation or "halo effect" that appears around very hot or cold objects, rapid automatic gain change and lack of chromatic information - thermal radiation reflected

from objects in the scene is mapped as a 1-d heat information. Hence, not all conventional computer vision techniques developed for electro-optical imagery are directly applicable to thermal video sequences. For example, vision algorithms such as optic flow estimation and background subtraction assume a pixel will maintain a constant intensity value over time – the brightness constancy assumption. Rapid adjustment of pixel sensor gain invalidates this assumption, leading to algorithm failure. When hot objects such as roads and building roofs enter the camera field of view, large regions of pixels saturate, and the sensor adjusts gain to darken the image in an attempt to avoid saturation. The resulting change in intensity can be dramatic from one frame to the next. Smaller changes in gain occur on a continual basis as the sensor adjusts to the changing contents of a moving field of view.

In this paper we show that an intensity-clipped affine model of sensor gain is sufficient to describe the behavior of our thermal sensor. Experimentation with thousands of thermal images proved that the gain change exhibits an affine behaviour, except in regions of "halo effect". We first align successive frames using a stabilization technique that uses sparse flow of corner features to compute the affine background scene motion that brings pairs of frames into alignment prior to estimating change in pixel brightness and then we fit an affine model to gain change and estimate the parameters using least mean squares estimator. Dense optic flow and background appearance modeling is then performed on these motion-compensated and brightness-compensated frames. We experiment the resulting algorithm to segment ground vehicles from thermal airborne video while building a mosaic of the background layer, despite the presence of rapid gain changes.

Section 2 reviews related work on gain compensation. In Section 3 we describe our approach to two frame stabilization, based on robust estimation of an affine transformation from sparse corner point corre-

spondences. Gain compensation and background estimation is described in section 4 and section 5, respectively. Section 6 presents experimental results on a thermal sequence, showing both gain-and-motion-compensated frames and estimated background mosaics.

## 2. Related Work

Radiometric calibration of a sensor allows one to model how the quantity of light (or heat) energy  $q$  collected at a sensor pixel during an exposure period is converted into pixel values in the image [11, 7]. The radiometric transfer function  $f(q)$  is typically nonaffine. In the present application we do not really need to know the nonaffine sensor response  $f(q)$ , but only the relation between  $f(q)$  and  $f(kq)$  after a relative change  $k$  in sensor gain or exposure time.

Mann [6] introduces the idea of a *comparametric plot* of  $f(kq)$  with respect to  $f(q)$ , using pairs of corresponding pixel values observed from images taken at different exposure settings. Fitting a parametric model to the observed correspondences yields a comparametric function that relates pixel intensities before and after a change in exposure. Mann points out that a commonly used model of nonaffine camera response is

$$f(q) = \alpha + \beta q^\gamma \quad (1)$$

where  $\alpha, \beta$  and  $\gamma$  are sensor-specific constants. The comparametric function relating  $f(q)$  to  $f(kq)$  is then a straight line

$$f(k(q)) = k^\gamma * f(q) + \alpha(1 - k^\gamma).$$

For thermal sensors, nonaffine film response is not an issue, and there is typically an affine relation between sensor response and radiant exitance (i.e.  $\gamma = 1$  in Equation 1). For example, Lillesand discusses the model  $N = A + B\epsilon T^4$ , where  $N$  is numeric sensor response,  $\epsilon$  is thermal emissivity at the point of measurement,  $T$  is blackbody kinetic temperature at the point of measurement, and  $A$  and  $B$  are calibration parameters to be determined [4]. Following [10], we consider a gain  $a_i$  and offset  $b_i$  mapping sensor response  $N$  to observed pixel values  $P_i$  at two different times  $i = 1, 2$ . Assuming the thermal exitance ( $\epsilon T^4$ ) stays constant between the two times, we find that an affine relationship

$$P_1 = \frac{a_1}{a_2} P_2 - \frac{a_1}{a_2} b_2 + b_1 = m P_2 + b$$

relates corresponding thermal pixel values.

Even without a physical justification, it has often been found useful to assume corresponding brightness values across different times and/or different sensors

are related by an affine intensity transformation  $P_1 = m P_2 + b$ , and to use this to facilitate image matching, change detection, and mosaicing of spatially registered views [5, 2, 3].

## 3. Affine Alignment via Sparse Corner Flow

Two-frame stabilization is achieved by establishing correspondences between adjacent video frames and estimating an affine or higher order transformation that warps the images into alignment. We estimate image alignment by fitting a global parametric motion model to sparse optic flow. We want a method that can bring frames into alignment prior to estimating gain changes. Since brightness changes are present, we use a feature-based approach. We first extract a sparse set of corner features in each frame, find potential matching pairs using normalized correlation, and then estimate the parameters of a global affine alignment using RANSAC. Higher order motion models such as planar projective could be used, however the affine model has been adequate in our experiments due to the large sensor standoff distance, narrow field of view, and nearly planar ground structure in the aerial sequences.

Our mathematical model for alignment is affine warped images with affine gain changes. That is

$$I_t(x, y) = m * I_{t-1}(a_1 x + a_2 y + a_3, a_4 x + a_5 y + a_6) + b$$

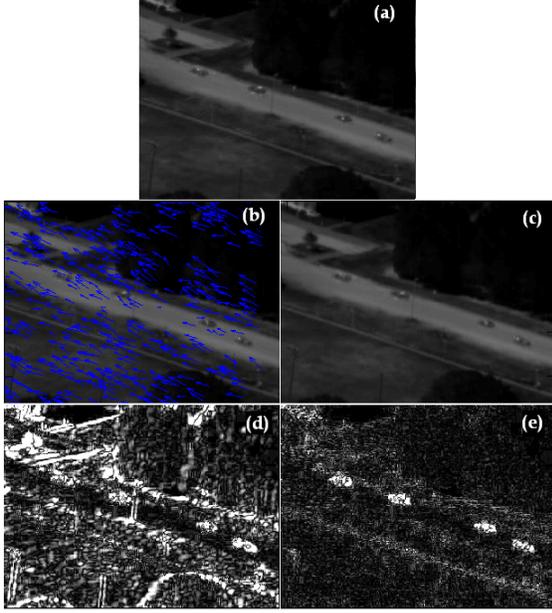
where  $a_1, \dots, a_6$  are the affine warp parameters, and  $m$  and  $b$  are the gain and offset parameters of the affine intensity change.

Given two frames that we want to align, we first detect corners in each frame using the Harris corner detector. The important criterion for this step is that the corner detection be *repeatable*, meaning that many of the corners from frame 1 should also have corresponding corners detected in frame 2, despite the affine warp and gain change in frame 2. The Harris detector detects corners as spatial maxima of the function

$$f(x, y) = \det(A(x, y)) - k \text{trace}(A(x, y))^2$$

where  $A(x, y)$  a Gaussian-weighted image autocorrelation function centered at pixel  $x, y$ , and  $k$  is a constant chosen in the range 0.04 to 0.06. A study of the repeatability of many standard corner detectors was conducted in Schmid et.al.[9], where it is shown that the Harris corner detector has the best repeatability with respect to moderate affine deformations and illumination changes.

Given sets of corners detected in frame 1 and frame 2, we perform matching to determine potential corresponding pairs. We compare each corner point in frame



**Figure 1. Stabilization results: (a) Current (2095<sup>th</sup>) frame,  $I_t$  (b) previous frame,  $I_{t-1}$  and sparse flow obtained through corner detection and matching, (c) stabilized frame,  $I_{t-1}^S$  (d) absolute difference between current and previous frame (e) absolute difference between current and stabilized frame.**

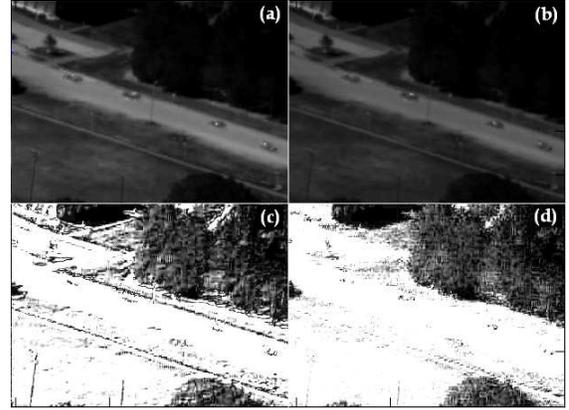
1 with corners from frame 2, using normalized cross correlation (NCC) to determine similarity. To reduce the combinatorics, the search set of corners in frame 2 is limited to lie within an area bounded by a predetermined upper bound on the magnitude of affine displacement between pixels in the two images. Letting the pixel intensities in an  $11 \times 11$  neighborhood around a corner in frame 1 be represented in vector  $P$ , and the  $11 \times 11$  neighborhood around a candidate corner match in frame 2 be vector  $Q$ , the NCC match score is computed as

$$\text{NCC}(P, Q) = \frac{1}{\sigma_P \sigma_Q} (P - \bar{P})^T (Q - \bar{Q})$$

where  $\bar{P}$  and  $\sigma_P$  is the mean and standard deviation of the values in  $P$ , and similarly for  $\bar{Q}$  and  $\sigma_Q$ . The NCC score is clearly invariant to affine gain changes, since subtracting means normalizes for the additive intensity offset  $b$  and dividing by standard deviations normalizes for the multiplicative gain factor  $m$ . Compatible corner matches are determined as pairs  $P$  and  $Q$  such that  $Q$  has lowest NCC score among the corners tested for  $P$ , and  $P$  has the lowest NCC score among the corners tested for  $Q$ . A similar “marriage-based” compatibility scheme was described in [8].

Given a set of potential corner correspondences

across two frames, a six parameter affine motion model is fit to the observed displacement vectors to approximate the global flow field induced by camera motion and a rigid ground plane. We use a Random Sample Consensus (RANSAC) procedure [1] to robustly estimate affine parameters from the observed displacement vectors. The benefit of using a robust procedure such as RANSAC is that the final least squares estimate is not contaminated by erroneous displacement vectors, points on moving vehicles in the scene, and scene points with large parallax.



**Figure 2. Stabilization results: (a) Current (2099<sup>th</sup>) frame,  $I_t$  (b) stabilized frame,  $I_{t-1}^S$  (c) absolute difference between current and previous frame (d) absolute difference between current and stabilized frame.**

Figure 1 illustrates the stabilization for one of the frames (2095<sup>th</sup> frame) in our database. Stabilization of the frames compensates for gross affine background motion and the residual error is high only in regions exhibiting different motion statistics than background layer as long as the gain doesn’t change. Unfortunately, gain changes rapidly in thermal imagery due to the nature of the physics of the thermal cameras acquiring these sequences. For example, Figure 2 illustrates stabilization results for 2099<sup>th</sup> frame in the same sequel, where a drastic change in gain occurs and the foreground objects can not be distinguished from background layer, despite correct alignment of the frames.

#### 4. Compensating Gain Change

Many computer vision techniques, such as optic flow estimation, background subtraction techniques are based on the assumption that brightness is conserved between successive frames. Rapid adjustment of pixel sensor gain invalidates this assumption, lead-

ing to algorithm failure. We are interested in estimating the function  $g$  that relates the registered previous frame,  $I_{t-1}^S$  to the current frame,  $I_t$ , namely

$$I_t = g(I_{t-1}^S) \quad (2)$$

Since  $I_{t-1}^S$  is registered towards  $I_t$  by an affine transformation, each pair at every pixel location  $(x, y)$  is supposed to have very close values, if the automatic gain control is not active. To understand the nature of this function, we built a 2-D histogram of the intensity values of stabilized frame versus current frame pairs:  $(I_{t-1}^S(x, y), I_t(x, y))$ , as illustrated in Figure 3(c). Horizontal axis corresponds to intensity bins for  $(I_{t-1}^S(x, y))$  ranging from 0 to 255, and vertical axis corresponds to intensity bins for  $I_t(x, y)$ . By examining thousands of thermal image pairs, we verified that the gain relationship between registered successive frames exhibits an affine behaviour.

We use least mean squares estimator (LMSE) to compute the parameters of this affine model, considering the pairs of observables,  $(I_{t-1}^S(x, y), I_t(x, y))$  to be related by

$$I_t = m I_{t-1}^S + b + \epsilon \quad (3)$$

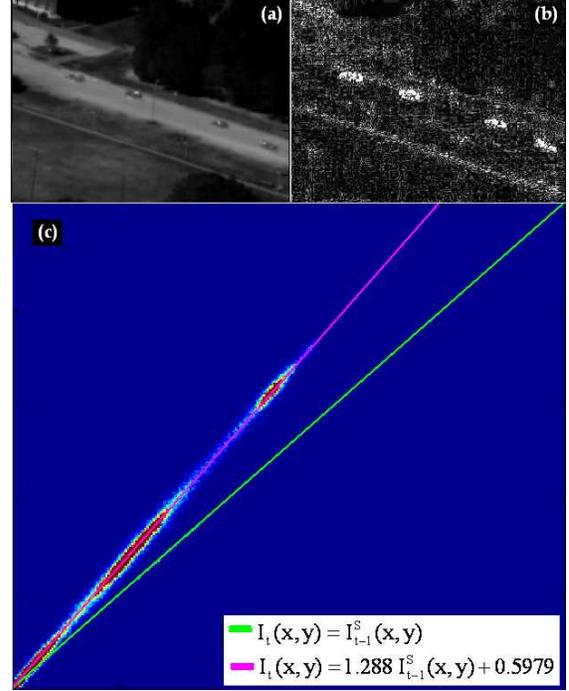
where  $\epsilon$  values at every pixel are independent and identically distributed. The formulas for the least squares estimator are provided in the Appendix.

The regions that appear either as too hot or too cold, namely regions with "halo effect" are excluded from the model estimation since they violate the affine model. For such regions, we construct confidence maps with low probability values, essentially meaning that the affine model for the automatic gain control does not hold in those regions.

Figure 3(c) illustrates this affine relationship for the 2099<sup>th</sup> frame of Figure 2. We built the 2-D histogram of the intensity values of  $I_{2098}^S$  (stabilized frame) versus  $I_{2099}$  (current frame) pairs. The intensity pairs  $(I_{t-1}^S(x, y), I_t(x, y))$  are mainly clustered along the line with parameters  $(m, b) = (1.288, 0.5979)$ . If the the automatic gain control of the thermal camera was inactive, these pairs were supposed to be clustered around  $I_t = I_{t-1}^S$  line. Note that the cars become visually distinctive after gain compensation as shown in Figure 3(b).

## 5. Application : Background Estimation

Affine alignment via sparse corner flow and modeling automatic gain control between successive frames enables reliable dense optical flow computation and background appearance modeling.

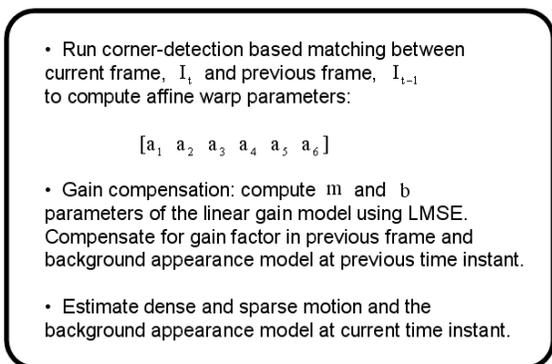


**Figure 3. Stabilization results: (a) Gain-corrected frame (b) absolute difference between current and stabilized-and-gain-corrected frame (c) histogram of  $(I_{t-1}^S(x, y), I_t(x, y))$  pairs and gain relationship plot between current and stabilized frame.**

We employ our background mosaicking technique [12] on these motion-compensated and brightness-compensated frames. Our approach to background layer modeling is based on applying robust optical flow algorithm to stabilized and gain-compensated image pairs. Stabilization of the frames compensates for gross affine background motion prior to running robust optical flow to compute dense residual flow. Based on the flow and the previous background appearance model, the new frame is separated into background and foreground occlusion layers using an EM-based motion segmentation. Estimated variance of the residual flow is used as a statistical test to determine the likelihood that each pixel is from the background or foreground, providing an ownership weight to a layer segmentation process. This ownership weight along with the gain confidence map (construction of which is explained in previous section) are then used to build a background mosaic from previous background appearance model and also new information from the observed image.

## 6. Experimental Results

The flow of our approach is illustrated in Figure 4. First, affine warp parameters are computed using corner-detection based matching and RANSAC and then previous frame and previous background appearance model is registered towards current frame. Second, parameters of affine model for automatic gain control is computed using least mean squares estimator and then registered frames are compensated for the gain factor. Finally, we run our algorithm that uses sparse and dense motion statistics to estimate background appearance model on these motion and gain-compensated frames.



**Figure 4. Our approach.**

We experimented with our approach on the task of detecting moving vehicles in airborne thermal video imagery. Figures 5 - 6 illustrate the background ownership weight, the absolute difference between current and stabilized frame and stabilized-and-gain-corrected frame, background mosaic and the gain relationship plot between current and stabilized frame. The frames displayed here are picked particularly from the ones where a severe gain change occurs between successive frames. Video clips of the experimental results presented in this section are provided as additional material.

Initially, we assume that the steady regions after stabilization belong to the background layer. Hence, we set the regions with no motion as the initial background appearance model and some regions in the background appearance model appear as blacked out. As further frames are processed, these regions are gradually recovered since the occluded regions (regions with low background ownership weight) are disoccluded and filled in with the warped appearance model from the previous time instant. Regions with high background ownership weight are updated from the current image.

Gain compensation under severe automatic gain control change makes the continuous execution of the algorithm possible without breaking and it also allows the

algorithm to carry along the previously estimated background appearance model which would otherwise be impossible.

## 7. Conclusions

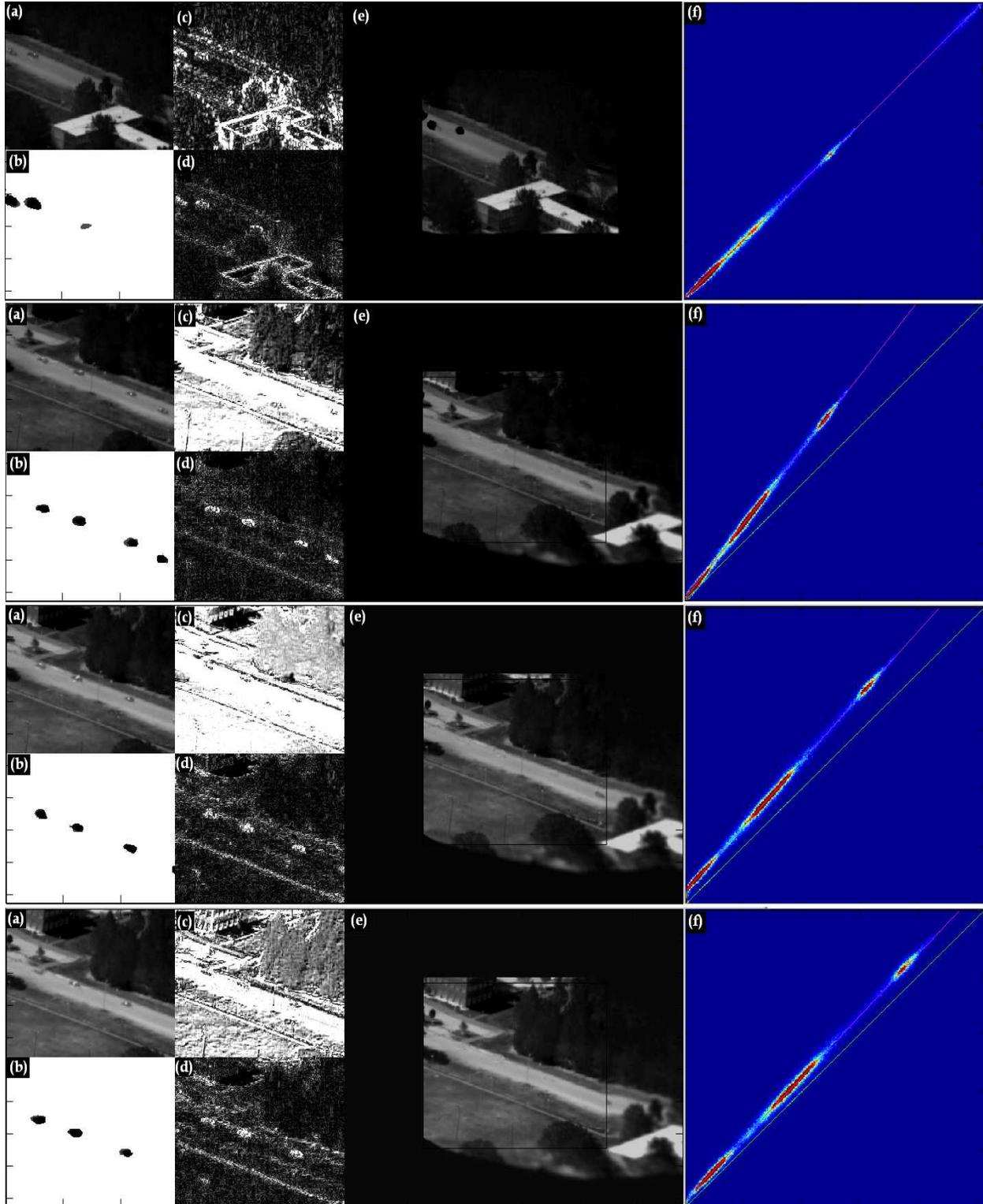
In this paper, we considered the detection of moving ground vehicles in airborne sequences recorded by a thermal sensor with automatic gain control. We exploited an affine model that describes the behaviour of gain change between successive frames which proved to be a very useful preprocessing step in combination with affine alignment procedure, for using an approach that integrates dense optic flow over time to maintain a model of background appearance and a foreground occlusion layer mask.

## 8. Acknowledgements

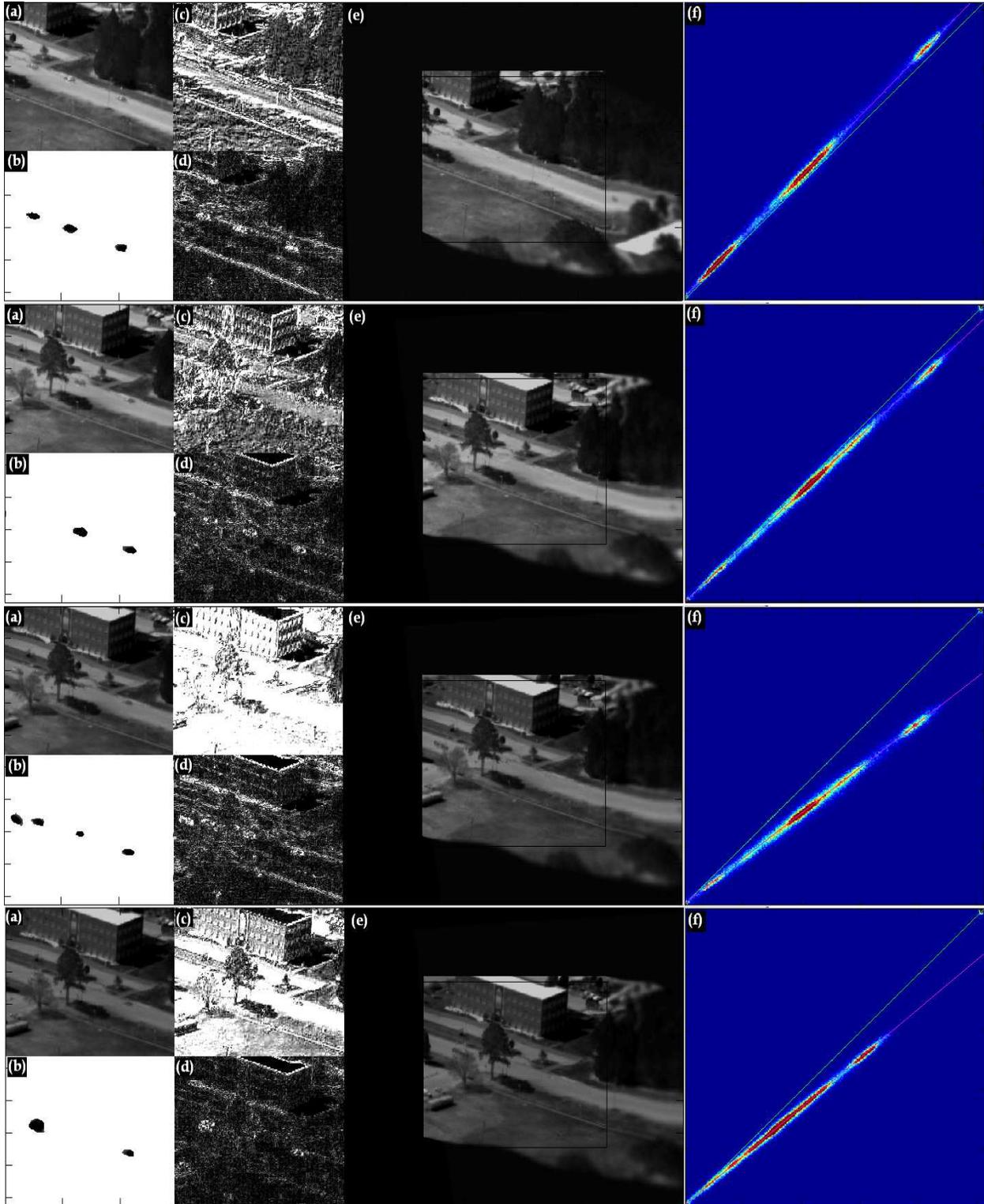
This work was sponsored by DARPA Grant NBCH1030013. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

## References

- [1] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24:381–395, 1981.
- [2] C. Fuh and P. Maragos. Motion displacement estimation using an affine model for image matching. *Optical Engineering*, 30(7):881–887, 1991.
- [3] J. Jensen. Urban/suburban land use analysis. In *R. Collwell ed., Manual of Remote Sensing*, Vol 2, 2nd edition:1571–1666, 1983.
- [4] T. Lillesand and R. Kiefer. *Remote Sensing and Image Interpretation*, 3rd edition, John Wiley and Sons, New York, 1994.
- [5] B. Lucas and T. Kanade. An iterative image registration technique with an application in stereo vision., *Seventh International Joint Conference on Artificial Intelligences (IJCAI-81)*, pages 674–679, 1981.
- [6] S. Mann. Comparametric equations with practical applications in quantigraphic image processing. *IEEE Trans on Pattern Analysis and Image Processing*, 9(8):1389–1406, 2000.
- [7] T. Mitsunaga and S. Nayar. Radiometric self calibration. pages 1374–1380, 1999.
- [8] D. Nister, O. Naroditsky, and J. Bergen. Visual odometry. 1:652–659, 2004.
- [9] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *IJCV*, 37(2):151–172, 2000.
- [10] J. Schott. *Remote Sensing, the Image Chain Approach*, Oxford University Press, New York, 1997.
- [11] Y. Tsin, V. Ramesh, and K. Kanade. Statistical calibration of ccd imaging process. pages 480–487, 2001.



**Figure 5.** Results of our approach for several frames. (a) Current frame. (b) Background layer ownership weight. (c) Absolute difference between current and stabilized frame. (d) Absolute difference between current and stabilized-and-gain-corrected frame. (e) Background mosaic. (f) Histogram of  $(I_{t-1}^S(x, y), I_t(x, y))$  pairs and gain relationship plot between current and stabilized frame.



**Figure 6.** Results of our approach for several frames. (a) Current frame. (b) Background layer ownership weight. (c) Absolute difference between current and stabilized frame. (d) Absolute difference between current and stabilized-and-gain-corrected frame. (e) Background mosaic. (f) Histogram of  $(I_{t-1}^S(x, y), I_t(x, y))$  pairs and gain relationship plot between current and stabilized frame.

- [12] H. Yalcin, R. Collins, M. Black, and M. Hebert. A flow-based approach to vehicle detection and background mosaicking in airborne video. *Technical Report CMU-RI-TR-05-11*, March, 2005.

## 9. Appendix

The formulas for the least squares estimator are

$$b = \frac{SS_{I_{t-1}^S, I_t}}{SS_{I_{t-1}^S, I_{t-1}^S}}$$

$$m = \bar{I}_t - b \bar{I}_{t-1}^S$$

where

$$SS_{I_{t-1}^S, I_t} = \sum_{x=1}^N \sum_{y=1}^M (I_{t-1}^S(x, y) - \bar{I}_{t-1}^S) \cdot (I_t(x, y) - \bar{I}_t)$$

$$SS_{I_{t-1}^S, I_{t-1}^S} = \sum_{x=1}^N \sum_{y=1}^M (I_{t-1}^S(x, y) - \bar{I}_{t-1}^S)^2$$

$$\bar{I}_{t-1}^S = \frac{1}{NM} \sum_{x=1}^N \sum_{y=1}^M I_{t-1}^S(x, y)$$

$$\bar{I}_t = \frac{1}{NM} \sum_{x=1}^N \sum_{y=1}^M I_t(x, y)$$

Here  $N$  and  $M$  are the horizontal and vertical sizes of the frames respectively.