# SENSOR SIGNAL PROCESSING FOR DEFENCE - SSPD CONFERENCE 2014

# Fusion of Thermal and Visible Images for Day/Night Moving Objects Detection

**Tarek Mouats** and **Nabil Aouf**
Centre for Electronic Warfare
Cranfield University, Defence Academy of the United Kingdom

## Abstract

A background subtraction technique based on the fusion of thermal and visible imagery using Gaussian mixture models (GMM) is presented in this work. An automatic daytime/night-time detection is introduced that can be used to dynamically adapting the fusion scheme. Three fusion schemes are investigated and coined as early, late and image fusion. Most approaches improve the performance of the combined system by compensating the failures of individual sensors.

**Figure 1.** One surveillance camera for every 11 people in Britain [1].

## Introduction

In recent years, the requirements in terms of video surveillance persistence have dramatically increased. This led to extensive research activities in this field both at the academic and industrial levels. The proliferation of CCTV cameras represents a concrete example (need for monitoring). Currently, there is a multitude of sensors used for video surveillance from which the visible and infrared cameras. Their simultaneous exploitation represents an interesting research path. The motivation comes from the fact that a surveillance system relying solely on one modality (e.g. CCTV camera) might become impaired in certain conditions. The multi-spectral system can provide the sought surveillance robustness and persistence by combining advantages of each modality while accounting for individual failures. The low level core task of such systems is the detection of moving objects or foreground (e.g. pedestrians, cars, …). Background subtraction algorithms are generally used for this purpose. They serve as the founding layer for higher level tasks such as traffic surveillance and crowd monitoring.

**Figure 2.** Multi-spectral systems.

## Methods and Materials

**Automatic daytime/night-time classification**
Histogram thresholding based on the HSV or CIELab colour spaces (see Figure 3).

**Background subtraction and fusion strategies**
The most appealing statistical characteristics of the GMM are [2]:

- representing multimodal backgrounds;
- accounting for background time variations.

We consider three fusion schemes:

*Early fusion:* pixels are modelled using a 4D vector (RGBT) prior to background subtraction.

*Late fusion:* background subtraction is applied to each sensor separately. The outputs are then fused.

*Image fusion:* a hybrid image is created by fusing thermal and visible images using a linear combination of both.

## Results

The proposed fusion approaches were tested qualitatively and quantitatively against different datasets available in the literature described in **Table 1** [3-5].

Using the ground truth images, we computed the following comparison metrics:

$$R = \frac{TP}{TP + FN} \quad and \quad P = \frac{TP}{TP + FP}$$

$$F = \frac{2 * R * P}{R + P}$$

**Table 1.** Video sequences used for the experiments.

| Sequence | Description | Ground truth |
|---|---|---|
| Parking Evening (**PE**) | Evening scene including cars and a person. | yes |
| Multiple Deposit (**MD**) | Daytime scene with large illumination variations. | yes |
| Group Fight (**GF**) | Daytime scene including a group of people. | yes |
| Parking Snow (**PS**) | Daytime scene in cold weather with persons and cars. | yes |
| Intersection (**IN**) | Daytime scene with monotonous illumination conditions. | no |
| Dark Night (**DN**) | Night-time scene with pedestrians and cars. | no |

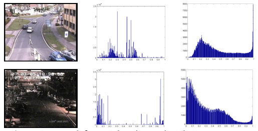

**Figure 3.** From left to right: day and night images; Hue histograms; Value histograms.

## Discussion

A first observation from **Table 2** is that foreground segmentation is very scene dependent as the results vary considerably with the imaged scene.

In general, the early and late fusions provide quite similar results. However, the early fusion is more robust in daytime scenarios (**GF** and **PS**).

Another observation is that early fusion is more accurate than image fusion with all its variations. One exception is when the scene exhibits large illumination variations (**MD**).
The **IN** sequence represents a daylight scenario where there is no influence from weather conditions as the scene remains relatively monotonous throughout. In this type of scenario, the early and late fusion perform similarly. The hybrid image composed from 15% thermal and 85% visible provides the best results in this category that are comparable to the other two.

Other mixtures of thermal and visible images deteriorated the performance of BS. This is due to the accumulation of the weaknesses of both sensors resulting in an overall performance degradation.

**Table 2.** Averaged F metric for each test sequence.

| Method | GF | MD | PS | PE |
|---|---|---|---|---|
| **VS only** | 0.679 | 0.483 | 0.846 | 0.752 |
| **IR only** | 0.758 | **0.719** | 0.871 | 0.733 |
| **Early fusion** | **0.852** | 0.649 | **0.900** | 0.812 |
| **Late fusion** | 0.835 | 0.613 | 0.883 | **0.833** |
| **Fus15IR85VS** | 0.669 | 0.603 | **0.839** | 0.641 |
| **Fus30IR70VS** | 0.643 | 0.657 | 0.835 | 0.600 |
| **Fus50IR50VS** | 0.609 | 0.645 | 0.819 | 0.629 |
| **Fus70IR30VS** | 0.703 | 0.798 | 0.691 | 0.623 |
| **Fus85IR15VS** | **0.819** | **0.802** | 0.717 | **0.694** |

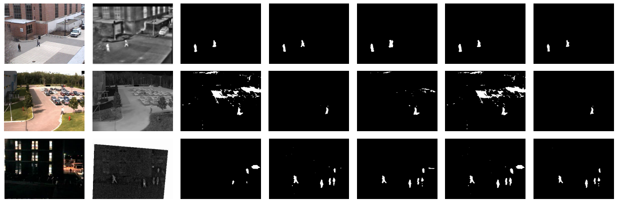

**Figure 4.** From left to right: visible images; corresponding thermal images; BS outputs from visible images; BS outputs from thermal images; BS outputs using early fusion; BS output using late fusion; BS output using image fusion.

## Conclusions

Thermo-visible fusion schemes for background subtraction are presented. They take advantage of optical and thermal sensors to provide better assessment of the foreground. To ensure an automatic 24/7 persistent operation, a daytime/night-time image classification algorithm is introduced. The early and late fusion schemes yielded mixed results with more robustness attributed to the early fusion. Image fusion is more appealing in scenes with large illumination variations as it was shown to cope better. This work represents the basis for higher level video surveillance tasks such as the continuous tracking of moving pedestrians or cars.

## Contact

Tarek Mouats
Cranfield University
Defence Academy of the United Kingdom
Email: t.mouats@cranfield.ac.uk
Website: www.cranfield.ac.uk/

## References

1. http://www.telegraph.co.uk/technology/10172298/One-surveillance-camera-for-every-11-people-in-Britain-says-CCTV-survey.html
2. C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in Computer Vision and Pattern Recognition, 1999, pp. 246–252.
3. J. W. Davis and V. Sharma, "Background subtraction using contour-based fusion of thermal and visible imagery," Comput. Vis. Image Underst., vol. 106, no. 2–3, pp. 162–182, May 2007.
4. C. O'Conaire, N. O'Connor, E. Cooke, and A. Smeaton, "Comparison of Fusion Methods for Thermo-Visual Surveillance Tracking," 9th Int. Conf. Inf. Fusion, pp. 1–7, 2006.
5. L. St-Laurent, D. Prevost, and X. Maldague, "Combination of thermal and color images for accurate foreground / background segmentation in outdoor environment," in IEEE International Conference on Image Processing, 2013, pp. 3431–3435.