Pedestrian Tracking Using Thermal Infrared Imaging

Emmanuel Goubet[†], Joseph Katz, and Fatih Porikli

Mitsubishi Electric Research Laboratories, Cambridge, USA

ABSTRACT

This paper describes an experimental study of the use of thermal infrared (8 - 12μm) imaging applied to the problem of pedestrian tracking. Generally it was found that infrared images enable better image segmentation, but their tracking performance with current algorithms is poorer. Simple fusion of both types of images has produced some improvement in the segmentation step of the tracking algorithms. In addition to the specific experimental results, this paper also provides a useful set of practical factors that need to be taken into account when using thermal infrared imaging for surveillance applications under real-world conditions.

Keywords: Infrared, Thermal imagery, Pedestrian detection, Change detection

1. INTRODUCTION

Until recently, thermal imaging using infrared cameras has been mainly limited to military applications, helping operators to better see their environment and to detect humans and vehicles in very low light conditions. Driven by the emergence of new types of thermal cameras, more affordable, infrared imaging technology is starting to reach industrial, commercial, and consumer markets.

Current applications - for example surveillance systems used to detect human or vehicle intrusions - generally require the presence of a human operator for making decisions after viewing the infrared images on a monitor. This paper describes an investigation of the use of an uncooled ferroelectric infrared camera for the task of automatic pedestrian tracking. Emphasis is placed on identifying real-world image features and camera artifacts, and explaining their influence on the system performance. Initial experiments with visible/infrared image fusion point to potential performance improvements that can be achieved via hyper-spectral imaging.

The outline of the paper is as follows. In Section 2 we briefly mention some of the specific characteristics of thermal infrared images captured with ferroelectric infrared cameras. In Section 3 we review the current state of the art in visible and infrared-based autonomous pedestrian detection/tracking, and in Section 4 we present experimental results. The results are augmented with some observations on infrared images taken under different scenarios (generally classified based on the ambient temperature as either "winter" or "summer"). This allows drawing some conclusions about the potential efficacy of the different types of algorithms under real-world conditions.

2. INFRARED IMAGING CHARACTERISTICS

The following well-known characteristics impact performance of automated image processing algorithms when applied to thermal IR-acquired images. Due to the longer wavelengths involved, surface reflections for many materials tend to be more specular and less diffuse than in the visible since the same surface irregularities are relatively smaller compared with the long IR wavelengths than for the short visible ones. See, for example, the car and the pedestrian reflection on the concrete road in Fig 1. This effect can produce errors in detection systems as the objects may be detected twice in some extreme cases. A solution has been proposed by Vollmer et al.[3] that takes advantage of the partial polarization of the specularly reflected light by using a polarizer to differentiate between the direct light and the reflected light.

In our experiments we used an infrared camera based on ferroelectric detectors. One of the advantages of these sensors is their immunity to ambient temperature drifts. All uncooled thermal detector systems are unshielded, which results in about 80% of the infrared radiation that reaches the detector elements coming from the camera housing, and not from the observed scene. Any temperature change in the camera creates a DC level drift. These changes are very slow compared to

---

those due to the rotating chopper in the camera, and thus can be filtered out. Due to the AC nature of this technology, it can't be used to measure the absolute value of the scene object temperature. However, it naturally produces a contrasted signal: the colder spots in the image appear darker, and the warmer ones lighter, regardless of their absolute value. For example, human bodies can appear light on a cold winter night, and dark on a hot summer day, even if they are at the same temperature (see Fig. 2).
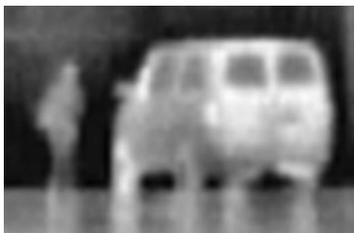

Figure 1. Example of infrared reflection.

Figure 2. Human body images produced by a ferro-electric camera: (left) cold night; (right) warm day

One of the most relevant characteristic of the images obtained with ferroelectric arrays, which can be a drawback in some applications, is the halo that surrounds objects having a large contrast with the background (see Fig. 2): bright spots on a dark background will be surrounded by a region darker than the background is and vice versa. The halo is due to the mechanical chopper failing to fully block infrared radiation: a hot source will heat the back of the chopper, and since this secondary radiation is less focused, it will heat the sensor array over a larger area than that of the actual image of the object. It is created when the system electronically subtracts the images with and without chopper obstruction.[6]

3. PEDESTRIAN DETECTION AND TRACKING: GENERAL PRINCIPLES

Detecting a pedestrian is generally realized in a two-step approach. The first phase, segmentation, consists of selecting regions of interest (ROI) likely to contain a pedestrian. During the second phase, classification, the ROI are analyzed to check correspondence with pedestrians. In contrast to blind approach, this widely used two-step method reduces computational time by reducing the number of analyzed regions. However, the boundary between the two steps is not strict, as a good segmentation process, which selects regions with high probability to contain a pedestrian, also acts as a pre-classifier. This is still an active research area, driven by broad applications in autonomous surveillance, security, and people counting, as well as other military applications. Overviews on this topic can be found at Lombardi[14] and at Gavrila.[7]

3.1. Pedestrian Detection in Infrared

Specific features of infrared imaging: The most salient characteristic of infrared imaging is that unlike visible imaging, it doesn't depend on lighting conditions but rather varies with temperature changes. The first consequence is obviously the ability to operate even in low light or night conditions. Most of the infrared imaging applications, in military and surveillance field, take advantage of this property, allowing detection of people or vehicles in the dark. Even in daytime outdoor scenes, the independence on lighting conditions is advantageous for image processing algorithms, as lighting conditions generally change faster than temperature, especially in scenes with clouds occluding sunlight. So with the background modeling techniques, the background adaptation is generally less crucial in infrared than in visible. Foreground region detection using background subtraction is also easier in infrared, as there are no shadows in infrared scenes. Shadow removal is one of the most important issues with background subtraction in the visible.

However, due to camera technology limitations, infrared images generally have lower spatial resolution (fewer pixels) and less sensitivity than visible ones. This may limit the efficacy of the classification step based on pedestrian silhouette/shape, especially with pedestrians far from the camera. Also, due to these limitations, infrared images can not provide as much information as visible ones about objects. This is further exacerbated by the fact that there is no spectral information in infrared images as there is only one sensor, compared to the three-channel RGB sensor arrangement in visible cameras, and this may lead to some problems in the tracking algorithms.

Segmentation and classification of pedestrians in infrared imaging: Most of pedestrian detectors working in thermal infrared assume that pedestrians are warmer than the background. This is true most of the time, especially at night and

during the winter, the contrast between the background and the pedestrians is generally significant. Systems making this assumption have a less complicated (and requiring less computational load) segmentation step than some of the more sophisticated ones used in the visible. The simplest approaches consist of applying a threshold to select hot interesting regions as proposed by Cielniak and Duckett.[11] This is more applicable to bolometer-based infrared cameras able to determine the absolute temperature of the objects.

Other algorithms select regions that both are hotter than the background and have significant contrast with it. For example, as proposed by Fang et al.,[12] the pixel values in each column are summed, allowing detecting vertical bands having significant contrast with the background and likely to contain a pedestrian. Then in each interesting band the highest and lowest bright pixels location is estimated in order to completely define the ROI boundaries. Conaire et al.[13] used a different approach. A histogram of the image is performed, which has a Gaussian shape centered on the background mean value and a little bump corresponding to the ROI pixels. Then the ROI are selected by extracting the pixels which are not in the large peak that have a value higher than the background mean value and which are grouped in objects with a minimum size. These techniques, while generally fast and efficient, are unable to detect pedestrians colder than the background, a situation that can happen in hot summer days. However, at night, when infrared detection is really needed compared to the visible one, they are efficient under most conditions.

Algorithms that use techniques similar to those in the visible (background subtraction in particular), manage to deal with the summer situation. Another advantage of the background subtraction technique is that it doesn't consider background hot spots as foreground regions. For example, in Fig.2-b, the background bright spot between the car and the pedestrian will not be extracted by a background subtraction technique. The infrared systems developed (Torresan et al.[10] or Thermal-Eye,[5] for example) generally use the same background subtraction algorithm as in the visible, which requires substantial computational and memory resources.

The classification step heavily depends on specific application requirements and on the segmentation step used. In typical urban scenarios, the detection steps based on background subtraction, are already efficient preclassifiers as the objects detected will be mostly cars and pedestrians, and so even simple aspect ratio discrimination can lead to reasonably good results. However, this criterion is generally used with other classification techniques to mitigate classification errors. For example, in the detector presented by Torresan et al.[10] (background subtraction segmentation) the pedestrian step frequency is considered more than the aspect ratio criteria.

Systems using a segmentation step based on intensity thresholding, in urban scenario, may require a more involved classification step than ones using background subtraction as they detect also background hot spots. For example, Fang et al.[12] use a combination of several classifier techniques. First the ROI histograms are compared with generic pedestrian histograms. Then an inertia-based classification is also realized to determine whether the bright pixels in the ROI are closer to the centre of the ROI (pedestrian case) or closer to the edges (non-pedestrian case). Finally a criterion similar to the aspect ratio of the ROI is applied.

In conclusion, it seems that under most common conditions, where pedestrians are considerably warmer than the background, ROI segmentation in infrared is efficient: nearly all the pedestrians are considered as ROI and most of the ROI are pedestrians. Thus it is possible to use simpler and faster segmentation algorithms (threshold/contrast), even if they are less accurate than visible-like techniques as they will also detect background hot spots. Efficient segmentation makes the classification work easier, and in some cases simple classification criterion can lead to good results. More complex classification techniques suffer from the lower quality of infrared images (problem with shape analysis techniques) and from the low amount of object spectral/texture information (problem with tracking and so with algorithms using motion). However when pedestrians are not warmer than the background, the threshold-based classification becomes inefficient, and segmentation techniques like background subtraction or motion detection must be used.

3.2. Visible-Infrared Fusion in Pedestrian Detection

One often distinguishes among several fusion levels of images coming from several cameras.[8] In low-level fusion, raw images obtained from the sensors are fused to a new image before applying any algorithm extracting information. In medium- (feature-) level fusion, extracted features from each raw image are fused. Finally, in high- (decision-) level fusion, decision algorithms are applied independently to each sensor output, and then these decisions are combined to reach a global decision. In pedestrian detection applications, the first level corresponds to direct fusion of the images coming from the different sensors. The second level corresponds to a fusion at a higher level, for example fusion of

Figure 3. The two-camera system used in the experiments

the segmentation results (foreground regions), and the high level fusion corresponds to the combination/verification of pedestrian localizations detected independently (segmentation + classification) by the different cameras.

The fusion of infrared and visible data to detect pedestrians seems promising as each sensor provides independent and complementary information. However, only a few research efforts have been carried out on this subject, especially for the medium- and high-level fusion. Several results have been reported on raw data fusion, but generally not for autonomous detection systems.

Therrien et al.[9] described a system designed to help land troops to better see their environment. Infrared images, useful in low light conditions, are combined with visible images that contain better texture information. The fusion is adaptive. For each local region of the two images, a spatial filtering is applied in order to measure and compare the amount of details. If one of the two images presents more details in a region then it will be locally preferred during the fusion process. The goal is to retain details (i.e. information) where they are present in either the visible or infrared image. A simpler (non adaptive) fusion imaging system is used by Iwao et al.[1] for medical applications.

Conaire et al.[13] show an example of the infrared/visible medium-level fusion potentiality for autonomous pedestrian detection. A preliminary fast ROI segmentation, based on image histograms, is applied on the IR images. Then a final segmentation process using background subtraction is performed on images having four channels (R, G, B, IR). The first segmentation helps the background subtraction in its initialization stage. The background model is initialized only in areas considered as background by the first segmentation process. This helps in dealing with initially occluded backgrounds. The final background subtraction is also improved by including the infrared image as a fourth image channel.

Torresan et al.[10] presented a pedestrian detector using a medium/high level fusion. The segmentation is realized independently on the infrared and visible data. Then the objects (defined by grouping closest ROI) are tracked, so that temporal information can be used in the next classification module (step frequency analysis). Fusion occurs at the tracking step. The detection quality of a same object is compared in infrared and visible images and the tracking in the channel having the stronger object (master) helps the tracking in the other channel (slave object). This leads to increased tracking continuity.

4. EXPERIMENTAL RESULTS

4.1. Data Acquisition

The camera used in the experiments is the Thermal-eye 2000B from Raytheon (Currently available from L-3 Communications[4]). It has resolution of 320×240 pixels, and its sensitivity is specified to be better than 100 mK. To be able to study visible and infrared fusion, visible sequences are simultaneously acquired. For the visible camera, we used the Sony Handycam PC105, with a 640×480 resolution. The two cameras were placed in parallel, with an 8-cm inter-optical axis separation (see Fig. 3). As typical scene objects are usually located at least several meters away from the camera, and since pixel-wise alignment is not required, having two different optical axes is adequate.

We recorded sequences on tapes at a rate of 30 frames/sec, and then digitized and aligned the image sequences afterwards (using Batch it Ultra software). A second camcorder is used to record on tape the signal from the NTSC infrared camera video output. To ensure proper alignment, the visible and infrared images are superposed, and the resultant images are displayed in real time. Each pixel value of a fusion image is the weighted sum of the values of the two corresponding pixels in the original images. Then a normalization step is performed so that the fusion pixel with the highest value corresponds to 255 and the one with the lowest value corresponds to 0.

<center>(a)   (b)</center>

<center>**Figure 4.** Typical winter day (a) and night (b) scenarios</center>

4.2. Examples of Infrared and Fusion Images

"Winter" infrared images The image of Fig. 4-a is a typical example of cold winter night scenario; the pedestrian is highly contrasted with the background, which is uniformly dark except for some artificial light sources. Fig. 4-b photograph represents the same location during the day. Due to sun heating, the background is less dark and less uniform. But the pedestrian is still relatively well discernable (and detectable with threshold/contrast segmentation). The halo-driven phenomenon of the ferroelectric camera is clearly observed at the boundary between the sky (very dark) and the scene objects (grey trees and buildings). Similarly, the background surrounding the pedestrian appears darker, and the pedestrian appears brighter than he should be.

There is yet another consequence of the halo/contrast-enhancement effect. Figs. 5-a and 5-b show a couple of pedestrians walking away from the camera in a winter night. We can see by comparing the two images the significant decrease in the textural information of the pedestrians. This is due of course to the fact that in Fig. 5-b they are defined by fewer pixels. But this is also a consequence of the white halo present inside the pedestrian area, which further suppresses the details. As the pedestrian images become smaller, the white halo covers their entire image. It can also be clearly seen in Fig. 4-a: the pedestrian is totally uniform, lacking any textural information (we have inspected the pixel values to confirm it was not a saturation artifact.)
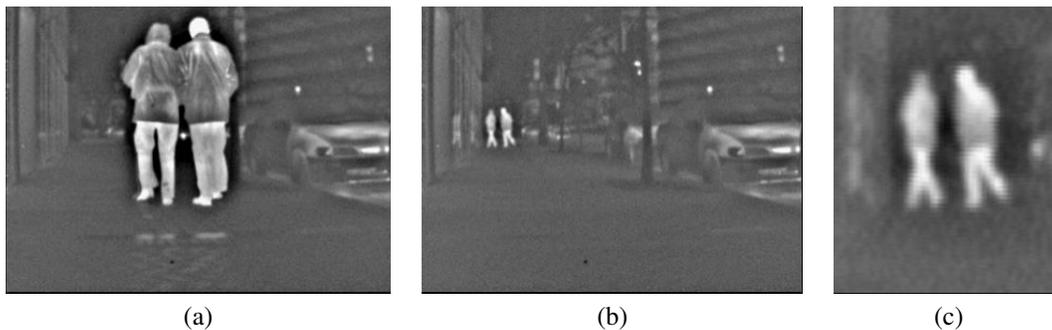

<center>(a)   (b)   (c)</center>

Figure 5. (a) Pedestrians close to the camera (high resolution). Pedestrians further from the camera (low resolution): (b) overall image; (c) enlarged view of area near the pedestrians in (b)

So the halo effect has several consequences. Depending on the detection algorithms and the scene, some are beneficial, but some are not. As it increases the contrast, the halo effect seems to aid threshold/contrast-based segmentation: the warm pedestrians can be more easily extracted. But in some scenarios, the contrast augmentation leads to some of the grey objects which are close to a very dark area to become as bright as the pedestrians and so they will also be extracted, thus necessitating a more complicated classification step (see, for example, the tree in the upper center of Fig. 4-b).

The halo effect may also facilitate ROI segmentation by background subtraction (also due to a higher contrast), but the dark halos surrounding the objects, darker than the real background, will be also extracted. So the ROI shapes will not correspond to the object shapes, and this may involve some problems with shape analysis classification.

Finally, whatever segmentation technique used, by reducing the pedestrian image details (leading to less rich histograms), the halo effect complicates the tracking processes and the use of classifiers based on these features.

An innovative segmentation algorithm, presented by Davis and Sharma,[5] attempts to take advantage of the halo effect. First ROI segmentation is realized by background thresholding. Then inside the ROI, which include the halos, the real boundary of the object is localized by looking for high contrast regions. So it is possible to have an efficient ROI extraction (due to the contrast enhancement of the ferroelectric technology) and to obtain the exact object boundaries.

Figure 6. Difference in brightness of pedestrian images.

Figure 7. Very hot summer day scenario.

Another problem can be seen in Fig. 6, which depicts two pedestrians. The one on the left has just left a warm car and thus appears very bright, contrary to the other one who wears cold clothes. The amount of infrared light coming from a pedestrian depends heavily on the temperature and infrared absorption coefficient of his/her clothes. This explains why it is not possible to use thermal signature alone to detect humans, even with microbolometer cameras that are able to measure the absolute temperature value of an object. So a classification algorithm using histogram comparison between ROI and a generic pedestrian template may lead to some errors.

Summer infrared images: Fig.7, taken in a hot summer day, presents a very complicated scene, totally different from the previous ones and with significant background detail (see again the significant halo effect at the sky boundary). The pedestrian is colder than the background. This often happens in warm conditions and this is the reason why background subtraction or motion detection segmentation algorithms should be preferred over the ones assuming hot pedestrians and cold background in order to have an efficient detection in all seasons/conditions.

Fig. 8 shows another interesting situation: a portion of the pedestrian image is colder, and the other portion is warmer than the background. We can also see another consequence of the contrast enhancement effect: the pedestrian upper part is brighter in Fig. 8-a than in Fig. 8-b. This is due to the fact that in Fig. 8-a the upper portion of the pedestrian image is surrounded by a darker background (the tree), so the contrast enhancement is stronger. Once again, this may be a problem in tracking processes as the extracted ROI histogram temporally change as the objects move in front of different background brightness (good tracking algorithms are adaptive, but the question is whether they are adaptive enough.)
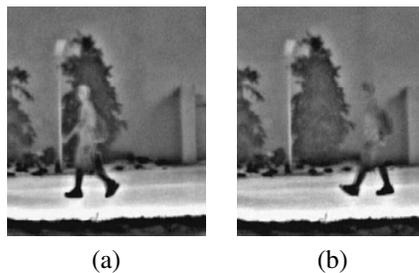

(a)   (b)

Figure 8. Contrast of pedestrian image; compared with the background, upper part is brighter in (a), and darker in (b)

Infrared-visible fusion images A fusion image example is presented in Fig. 9, which shows the visible (Fig. 9-a, in grey levels) and corresponding infrared (Fig. 9-b) images, respectively. These two images have been fused using the weighted sum algorithm, and Fig. 9-c and Fig. 9-d show some of the fusion results for differing weighting factors. We can see that this technique allows producing images having both types of information: texture/details information and heat localization information. This is an improvement example, as the top of the pedestrian in Fig. 9-a is hardly discernable, contrary to the situation in the fused images.
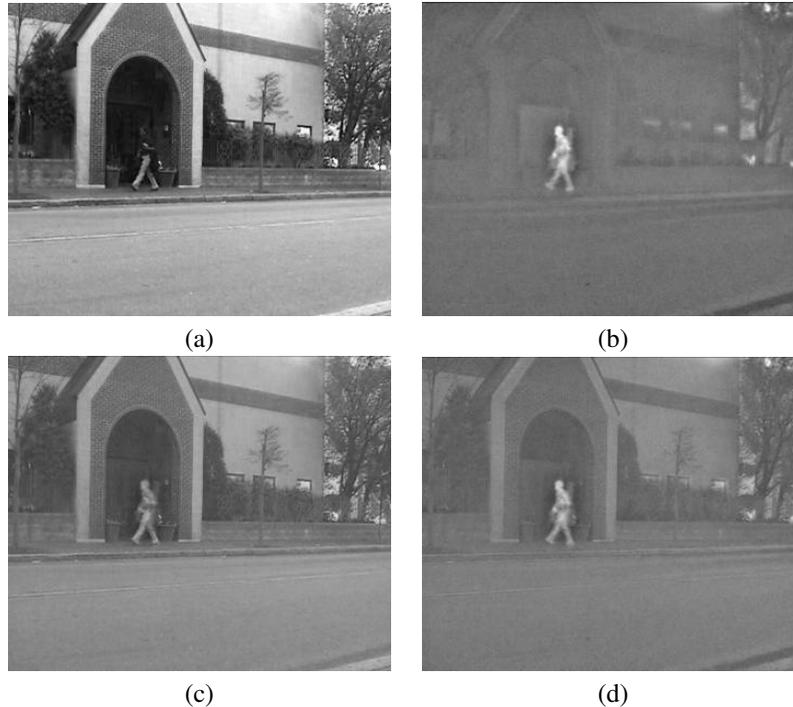
(a)   (b)

(c)   (d)

Figure 9. Fusion example of visible and infrared images: original visible (a) and infrared (b) images; (c) and (d) show two fusion examples with different weighting factors

4.3. Performance of Change Detection

We utilized a background subtraction based change detection algorithm developed by Porikli and Tuzel[2] to evaluate detection performance with infrared data. Basically, a mixture of statistical models is learned for every pixel, and a background map is obtained by selecting the most confident model in the mixture. By comparing the current pixel values to the background map, foreground regions are determined. Then, objects are initialized using morphological filtering and connected component techniques. Finally objects are tracked in the following frames using their color histograms. Although there is not a pedestrian classification step, analyzing this system with infrared data is important as many pedestrian detectors rely on background subtraction for segmentation step and tracking for classification step, which is often associated with motion analysis.

We analyzed the performance of change detection with both infrared and visible data. Our six spatially aligned, synchronous test sequences can be seen in Fig. 13. These sequences are between 20 to 60 seconds long. Except for sequence 2, they all have been recorded outdoors in summer days, which is the worst-case scenario for temperature magnitude based detection. For each frame of these sequences, we manually annotated pedestrian regions. We computed true and false positive rates by comparing the estimated foreground regions to the ground truth. The true positive rate is defined as the ratio of the number of correctly estimated foreground pixels to the number of foreground pixels in the ground truth. The false positive rate is calculated as the ratio of the number of incorrect foreground pixels that are not within the ground truth regions to the total number of actual background pixels, which equals to the total number of pixels subtracted by the number of ground truth pixels. These rates depend heavily on one parameter: the threshold value in the foreground extraction step. A low threshold value implies that all the pixels slightly different from the background should be considered as foreground pixels, meaning high true and false positives rates. On the other hand, a high threshold value leads that only pixels very different from the background are considered as foreground pixels, thus, low true and false positives rates.

We provide receiver operating characteristic (ROC) curves, which plots the true/false positives rates for different threshold values, of our change detection algorithm in Fig. 10. The upper-right corner of these graphs is the desired operating performance that means detection of all ground truth regions while making no wrong estimations. We tested the performance for 3 cases: infrared only, visible only, and joint representation, in which we considered the data has 4-channels including color and temperature values.
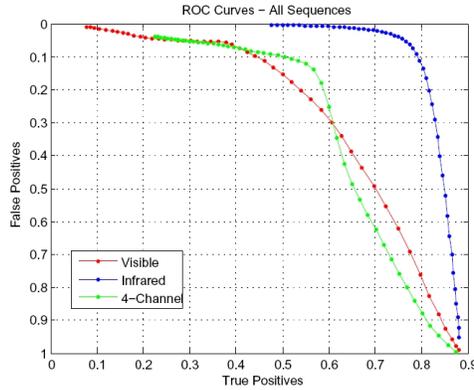
Figure 10. True positive versus false positive graphs for visible only (red), infrared only (blue), and joint data vector (green).
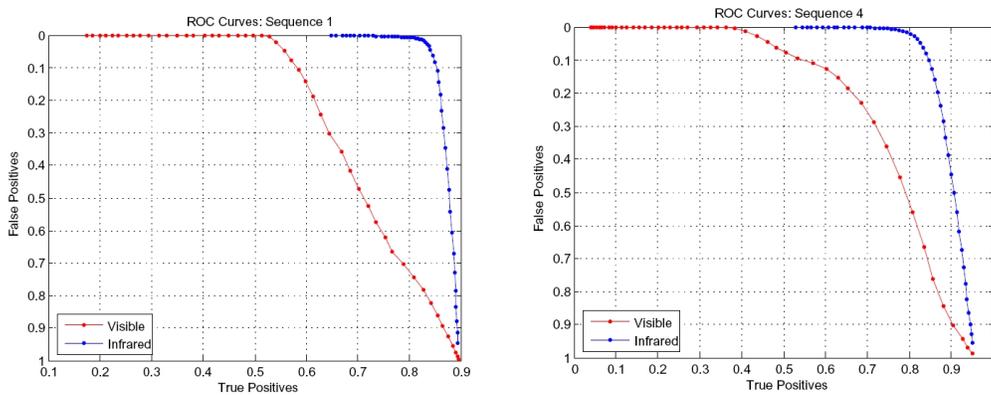

Figure 11. ROC curves for sequence 1 (left) and sequence 4 (right)

It is clearly seen that the infrared data yields more accurate foreground extraction; the infrared ROC curve is closer to the ideal upper-right corner point of the graph (maximum true positives, minimum false positives) than the visible and joint representation. This is a prevalent outcome; even if some differences can be noticed between the ROC curves of individual sequences (see for example Fig. 11 for ROC curves of sequence 1 and 4), the infrared data leads higher performance in background subtraction based change detection.

There are several factors influence these results. One issue is the reliability of the visible data. In color images, the background spectrum seems to have a higher likelihood of sharing similar properties with the pedestrians. For instance, most pedestrians have dark colors on their body, which can be easily confused with the darker parts of the background in pixel-level analysis. Another concern is the moving cast shadows created by the pedestrians. Such shadows cause inaccurate foreground detections and distort the pedestrian boundaries. Moreover, static shadows in the scene significantly change the appearance of pedestrians whose enter those regions. The background subtraction may fail to detect pedestrians in static shadow regions due to the lack of sufficient color dissimilarity since the intensity of the object colors are suppressed. Yet other issues of visible data are the amount of image noise due to small aperture sizes at fast frame rates, and artifacts caused by the compression of relatively greater amount of color information.

On the other hand, infrared data comes with its own concerns, for instance, in summer, where contrary to the winter scenario, pedestrians have temperature values similar to that of background. Besides, infrared data has only one dimension compared to the multiple channels in color and this may restrict the saliency of the infrared based features (or we can say that there is no spectral information in infrared). Also due to the specific camera technology used in our experiments, the contrast is enhanced by the halo effect.

One way to improve the true positive rate is to increase the detection sensitivity, i.e. to lower the threshold value.
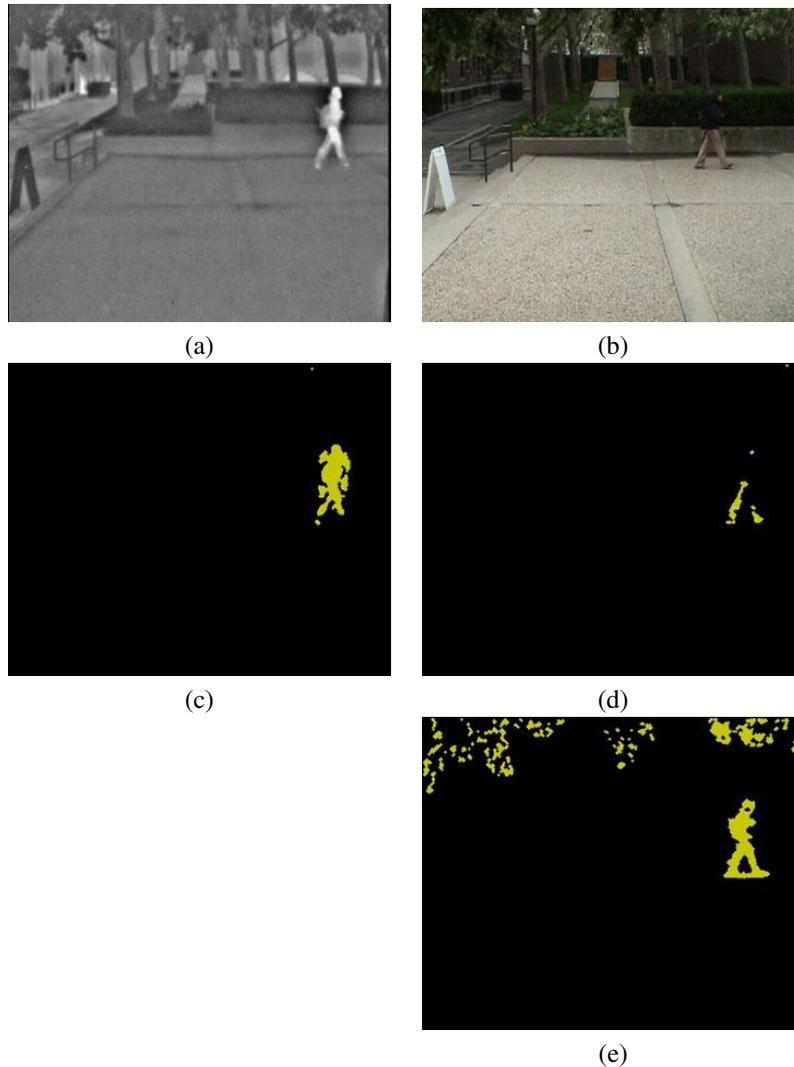
Figure 12. Background segmentation in visible and infrared images. (a) original infrared image; (b) original visible image; (c) foreground regions of (a); (d) foreground regions of (b) - normal threshold; (e) foreground regions of (b) - low threshold

However, we observed that this rapidly increases the amount of false positive regions distributed randomly in the entire scene in comparison to infrared. As the visible camera has a good resolution, little changes in the scene, e.g. moving leaves, are easily detected. Also, the scene itself seems to be noisier in visible, probably due to the absence of fast light variations in infrared, which makes background subtraction more robust. This phenomenon is shown in Fig. 12. These limitations, the halo effect around the pedestrian in the infrared data and shadows between the pedestrian feet in the visible data can be seen in Figures 12-c and 12-e respectively. If the shadows can be efficiently removed, the halo effect removal would imply either using a microbolometer camera (where there would no longer be contrast enhancement effect) or using adaptive algorithms.[5]

Although this example is not a representative of the worst-case scenario (where the pedestrian is warmer than the background), and of the best-case visible extraction scenario (where the pedestrian has different color than the background), it is helpful to understand the underlying dynamics; if we want to have a similar true positive rate in the visible data as in the infrared, we need to significantly increase the false positive rate. This can be seen clearly on the ROC graphs in Fig. 10.

We examined whether fusing the visible and infrared data would improve the results even though we observed the infrared background subtraction is more accurate. We have plotted the ROC curves averaged for all the sequences as presented in Fig. 13. We separately evaluated the likelihood of a pixel being a part of the foreground region for both
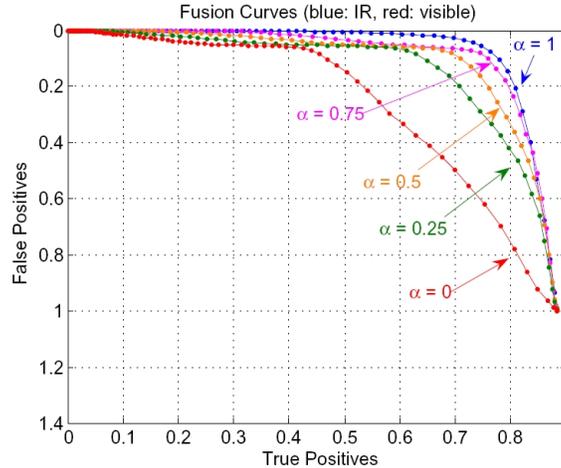
Figure 13. ROC curves for fusion: Color of the graph indicates the value of the weighting coefficient, i.e. red corresponds to $\alpha = 0$, blue to $\alpha = 1$ and other colors to $0 < \alpha < 1$. Blending coefficient α weights the likelihoods of the infrared I_R and visible I_V as $\alpha I_R + (1 - \alpha)I_V$.

infrared and visible data, then computed a weighted average as the pixel combined likelihood. Different curves correspond to different weights given to infrared and visible information during the fusion. It is interesting that Fig. 13 shows that no improvement is achieved by alpha blending of the infrared and visible data.

We qualitatively compared the performance of pedestrian tracking with infrared data to the tracking with visible. We found out the accuracy of both methods are similar. However, in the visible data, there were slightly fewer tracking errors. In the infrared data, the spatial resolution is low, the halo effect destroys the fine texture within the pedestrian regions, besides, the histogram of temperature values is less discriminatory. As we stated before, these factors complicate tracking and motion-based classification algorithms.

All combined, a higher true positives rate is easily achievable in infrared. We observed that the temperature difference provides more accurate and confident information to detect pedestrians. The main reasons are; 1) temperature of background scene does not change as rapidly as the background illumination, 2) people have more likely to have similar background colors on their clothes than similar background temperature, 3) motion of the trees, leaves and other dynamic background bodies generate less false positives in infrared, and 4) imaging noise is more apparent in color images.

5. CONCLUSIONS

Autonomous pedestrian detection in infrared is different from pedestrian detection in visible as the two types of images have different characteristics. Performance of visible cameras does not depend (first order) on the ambient temperature, and thus it remains constant in all the seasons. They do depend, however, on having some minimal level of ambient light, so their performance is usually unacceptable at night (unless auxiliary illumination is provided.) In contrast, performance of infrared cameras strongly depends on the season, i.e. on the scene temperature. Due to polarity inversion between backgrounds and pedestrians observed in summer, infrared segmentation algorithms assuming warmer pedestrians against colder background shouldn't be used for a system that needs to operate reliably in summer scenario. However we have shown that visible imaging-like techniques (background subtraction in particular) results in efficient segmentation in infrared images even in summer, due to the generally high contrast between pedestrians and backgrounds.

In contrast, classification in infrared images, based on shape or on human body feature analysis is more difficult than in visible images due to the lower amount of object information provided in the former (mainly due to their lower resolution and lack of spectral and texture information.) This is also a problem for motion-based classifiers that indirectly use this information in the tracking step. We have also seen that thermal signature should not be used to classify ROI (regions of interest) in pedestrian/non-pedestrian scenes, even with microbolometer cameras that are able to determine absolute temperature. This is due to the high variation of clothes temperature.
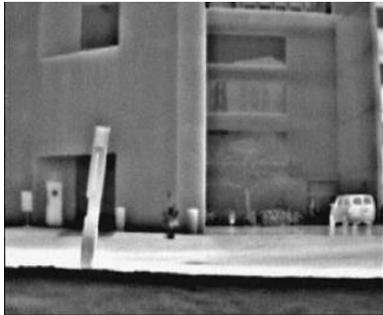
Figure 14. Sample test sequences

Overall, during daytime, pedestrian detection in infrared seems to be as efficient as pedestrian detection in visible: problems in the classification step are compensated by a better segmentation step. Fusion of infrared and visible data is believed to improve detector efficiency as both data types present complementary information. However, no noticeable improvements have been produced by basic low-level fusion technique.

REFERENCES

1. I. Fujimasa, A. Kouno, H. Nakazawa, Development of a new infrared imaging system: an infrared image superimposed on the visible image, 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol 20, No 2, 1998.
2. F. Porikli, O. Tuzel, Fast Object Tracking by Adaptive Background Models and Mean-Shift Analysis, IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2003.
3. M. Vollmer, S. Henke, D. Kartadt, K. Mollmann, F. Pinno, Identification and Suppression of Thermal Reflections in Infrared Imaging, InfraMation, 2004.
4. http://www.thermal-eye.com/productcatalog/prodItem19.html
5. J. Davis, V. Sharma, Robust Background-Substraction for Person Detection in Thermal Imagery, IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum, CVPR, 2004.
6. N. Pandya and J. Van Anda, Across the Spectrum, SPIE Magazine of Photonics Technologies and Applications, 2004.
7. D. Gavrila, The Visual Analysis for Human Movement: A Survey, Computer Vision and Image Understanding, Vol 13, No 1, 82-98, 1999.
8. D. Borghys, P.Verlinde, C. Perneel, M. Acheroy, Multi-Level Data Fusion for the Detection of Targets using multi-spectral Image Sequences, SPIE Optical Engineering's special issue on Sensor Fusion, Vol 37, No 2, 477-484, 1998.
9. C. Therrien, J. Scrofani, W. Krebs, An adaptive Technique for the Enhanced Fusion of Low-Light Visible with Uncooled Thermal Infrared Imagery, International Conference on Image Processing, 1997.
10. H. Torresan, B. Turgeon, C. Ibarra-Castanedo, P. Hebert, X. Maldague, Advanced Surveillance Systems: Combining Video and Thermal Imagery for Pedestrian Detection, SPIE Thermosense XXVI, 2004.
11. G. Cielniak, T. Duckett, People Recognition by Mobile Robots, Journal of Intelligent and Fuzzy Systems, Vol 15, No 1, 21-27, 2004.
12. Y. Fang, K. YUamada, Y. Ninomiya, B. Horn, I. Masaki, A shape-Independent Method for Pedestrian Detection With Far-Infrared Images, IEEE Transactions on Vehicular Technology, Vol 53, No 6, 2004.
13. C. Conaire, E. Cooke, N. O'Connor, N. Murphy, A. Smeaton, Background Modelling in Infrared and Visible Spectrum Video for People Tracking, Proc. of IEEE Int'l workshop on Object Tracking & Classification Beyond the Visible Spectrum, 2005.
14. P. Lombardi, A Survey on Pedestrian Detection for Autonomous Driving Systems, Technical Report, University of Pavia, 2001.