Evaluation of the Fusion of Visible and Thermal Image Data for People Detection with a Trained People Detector

Achim Königs and Dirk Schulz
Unmanned Systems Group, Fraunhofer FKIE, Fraunhoferstr. 20, 53343 Wachtberg, Germany

Keywords: People Detection, Thermal Imaging, Image Processing, Data Fusion.

Abstract: People detection surely is one of the hottest topics in Computer Vision. In this work we propose and evaluate the fusion of thermal images and images from the visible spectrum for the task of people detection. Our main goal is to reduce the false positive rate of the Implicit Shape Model (ISM) object detector, which is commonly used for people detection. We describe five possible methods to integrate the thermal data into the detection process at different processing steps. Those five methods are evaluated on several test sets we recorded. Their performance is compared to three baseline detection approaches. The test sets contain data from an indoor environment and from outdoor environments at days with different ambient temperatures. The data fusion methods decrease the false positive rate especially on the outdoor test sets.

1 INTRODUCTION

There are several reasons to detect people in the sensor data of mobile unmanned systems. One is to navigate more cooperative in the proximity of moving people. This requires a high detection rate but does not suffer much from false positives. Other applications like service robots where one might want to control a nearby robot using gestures or want an unmanned system to follow a certain person through a crowded area need to reduce false positives to be useful. In other scenarios where unmanned systems might be deployed in the future, like disaster response or surveillance tasks, false positives should be avoided as much as possible without lowering the detection rate.

There exist a lot of approaches to people detection in the computer vision community that work on images from the visible spectrum. Not all of them can be deployed on unmanned systems. Some cannot handle moving cameras whereas others are to slow to produce results in a close to real-time fashion which is required in the scenarios described above. The remaining approaches to people detection, however, cannot deliver the required high detection rates and low false positive rates at the same time. For all detection systems a low false positive rate means also a lower detection rate. In order to overcome this we propose to fuse visible spectrum images with thermal images. In thermal images background structures which would cause false positive detections in the visible spectrum are usually only barely visible because their temperature signature is not very different from the remaining background. All in all, in thermal images the person tends to be clearly visible even if for outdoor scenarios the assumption that persons are a hot spot not always holds (see Figure 2).

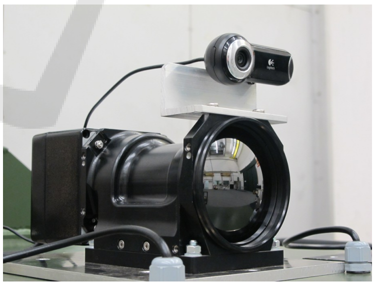

Figure 1: Camera setup. The USB camera is mounted on top of the thermal camera so they both capture images of the same scene.

In this work we use the Implicit Shape Model (ISM) (Leibe et al., 2008) object detector trained on people and propose five methods to fuse visible spectrum and thermal data during the detection process. These different methods influence the detection process in the feature and decision levels. At the same time we use unmodified ISM as baseline algorithm on both visible spectrum and thermal images. Additionally we implement a thermal threshold based blob detector (short blob detector) as second baseline. For

Königs A. and Schulz D..
Evaluation of the Fusion of Visible and Thermal Image Data for People Detection with a Trained People Detector.
DOI: 10.5220/0004484203450352
In *Proceedings of the 10th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2013), pages 345-352
ISBN: 978-989-8565-71-6

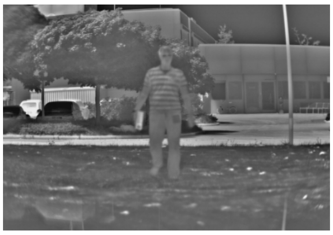

Figure 2: On this thermal image, the person has a very low contrast to the background, because of high temperatures in the surroundings. This makes it impossible to find a fixed temperature threshold to separate person and background reliably.

the evaluation we recorded three datasets, one from indoors and two from outdoors on two different days with different ambient temperatures.

The remainder of this paper is organized as follows. First we give an overview of other approaches to people detection on thermal or fused data in section 2. Then we describe the two baseline algorithms in section 3. The fusion methods are explained in section 4. Our datasets and the experiments we performed are presented in section 5 followed by a short conclusion in section 6.

2 RELATED WORK

People detection is an important and very active research topic. People detection is done with a lot of different sensors and different techniques. In order to stay in scope, we will only discuss a subset of literature available on this broad topic. For a broader survey about people detection methods in the visible and thermal spectrum see (Enzweiler and Gavrila, 2009) or (Gero andnimo et al., 2010).

First we will talk briefly about people detection in visible spectrum images and explain why we did choose ISM as our baseline algorithm and as basis for modifications. Then we will give a short overview of people detection in thermal images. In the end we talk about approaches that use the thermal and visible spectrum for people detection at the same time.

2.1 Visible Spectrum Images

People detection in the visible image space is still a very active research topic. Recent advances in people detection (Dalal and Triggs, 2005; Leibe et al., 2008; Leibe et al., 2005; Schiele et al., 2009) show good detection rates and are quite time efficient. In this work the FastPRISM detector from Lehmann (Lehmann et al., 2010) is used. It is an advanced variant of Leibe's implicit shape model (ISM) detector (Leibe et al., 2008). On the one hand, it improves stochastic reasoning and allows discriminative learning, and on the other hand, it proposes a necessary acceleration of the ISM method. The acceleration is achieved by applying the Efficient Subwindow Search (ESS) method proposed by Lampert (Lampert et al., 2008) to the feature centric view of ISM by Lehmann (Lehmann et al., 2009).

The feature centric view makes ISM (and FastPRISM) very robust against missing data. Missing data can occur for partly occluded people. But the robustness against missing data also helps during the fusion, if one of the fused spectra delivers only few features.

2.2 Thermal Images

One very popular approach to people detection on thermal images is the detection of hot spots, in most cases by applying a fixed temperature threshold. Often after that step only heuristics over the human size are applied to distinguish people from non-human heat sources (Cielniak and Duckett, 2004; Serrano-Cuerda et al., 2011; Li et al., 2010). All those approaches are based on the assumption that people are generally warmer than the background. For indoor scenarios this assumption holds, but for outdoor scenarios this often is wrong as Goubet shows in (Goubet, 2006). We can confirm this observation from our own experience; Figure 2 shows an example where the person is actually colder than parts of its environment, making constant temperature thresholds unfeasible. There are more sophisticated methods like (Davis and Sharma, 2004). However, they often heavily rely on statistical background subtraction. For statically mounted surveillance cameras this makes perfect sense, but on mobile robots this approaches does not work. As soon as the robot starts moving the previously calculated statistical model of the background is worthless, because the whole background has changed and the subtraction does not work anymore.

The use of a Histogram of Gradients (HOG, by Dalal and Triggs in (Dalal and Triggs, 2005)) like detector on thermal images is proposed by Miezianko and Pokrajac in (Miezianko and Pokrajac, 2008). Which is similar to the work of Zhang et al. (Zhang et al., 2007). They compare the performance of HOG like detectors on thermal data to the performance on visible spectrum images. The evaluation is carried out on thermal images and images from the visible spectrum. However, the images do not show the same scenes, which hinders the comparison of the results on thermal data against the results on images from the visible spectrum. A fusion of the thermal and visible spectrum is not possible with their data and does not seem to be in their interest.

2.3 Image Fusion

Goubet et al. are one of the first to propose the fusion of thermal and visible spectrum images (Goubet, 2006) for people detection. They fuse the two images on the pixel level using a weighted sum creating a new gray scale image in which changes are detected and heuristics about the size of people are used filter out people. A much more sophisticated method is proposed by Han et al. in (Han and Bhanu, 2007), which works on silhouettes from both spectrum images. Both approaches rely heavily on background subtraction, though, which forbids the usage on mobile robots. Another low level fusion approach is presented in (San-Biagio et al., 2012). They fuse the images at pixel level and calculate covariance matrix descriptors that they use as features for later stages. So the features contain information from both spectra. For that they need a very good registration of thermal and visible spectrum camera. They circumvent the lack of a suitable interest operator for such features by using a dense feature map which is very expensive to calculate and therefore not feasible for applications on mobile robots.

In contrast to the very low level fusion that Goubet et al. propose, Serrano-Cuerda et al. propose a decision level fusion in (Serrano-Cuerda et al., 2011). An detector is applied to the visible spectrum image and the thermal image separately and the results are joined in the end. The detection in the color image is done by background subtraction and color segmentation. On thermal data a simple blob detector is used. A simple blob detector for the thermal data is also used in (Cielniak and Duckett, 2004). The detected blobs are transferred to the visible spectrum image and a feature vector is calculated from both images and fed into different classification systems. Other works concentrate on the fusion of other modalities, like visible spectrum images and depth data in (Spinello and Arras, 2012; Ikemura and Fujiyoshi, 2011) or visible spectrum and near infrared images (Lietz et al., 2009).

2.4 Contribution

Previous work in (Königs and Schulz, 2012) evaluates a blob detector on thermal data against the more advanced ISM detector applied on thermal data. This evaluation is extended here to include different data fusion approaches. Novel, to our knowledge, is the direct comparison of fusion approaches on the feature level and on the decision level. Furthermore we compare the performance of the proposed methods on indoor and outdoor datasets which include both thermal and visible spectrum images from the same scene which allow a direct comparison of the algorithms. So the main contribution of this work is an in-depth evaluation of different fusion methods against detectors working on unfused thermal and visible spectrum data.

3 BASELINE ALGORITHMS

In this section we briefly introduce the algorithms used as baseline to compare the fusion methods against. First we describe the implementation of a simple *Blob Detector* that is commonly used for people detection in thermal images. Second we describe the ISM detector and especially its voting mechanism, which we modified in different ways to incorporate thermal data into the detection process.

3.1 Blob Detector

We implemented a blob detector to work on the thermal images in order to have a widely used baseline algorithm for people detection on thermal images. The *Blob Detector* first applies a thresholding to the thermal image I with a fixed thermal threshold. The result is a binary image. After that a morphological closing is applied. The result is a smoothed binary image I_m.

Now we search for contours in the image I_m (see (Suzuki and Abe, 1985)). The result is a set of blobs B that depict connected components in the binary image. For each of this blobs $b = (x, y, w, h) \in B$ the distance to the average width w_p and height h_p of persons in our test datasets is calculated and incorporated into a confidence value b_c:

$$b_c = \min(1 - \frac{|w - w_p|}{w_p}, 1 - \frac{|h - h_p|}{h_p})$$

This confidence value b_c is then tested against the detection threshold.

3.2 Implicit Shape Models

The second baseline algorithm we use is FastPRISM by Lehmann et al. (Lehmann et al., 2010). It is a speed up variant of Implicit Shape Models (ISM) originally developed by Leibe et al. (Leibe et al., 2008). The detector operates on image features; SURF (Bay et al., 2008) in the case of this work. It is based on a codebook that is created during training by clustering the observed features. For each of these codebook entries the detector stores votes for the person centers in vote images. During detection new features are associated to the codebook entries by nearest neighbor search in the feature space. Then they vote for the possible person centers from the vote image of the associated codebook entry. The votes are cast into a voting space in which space local maxima of accumulated votes are located and depicted as person centers.

The improved algorithm of Lehmann et al. applies a branch and bound search strategy for local maxima in the voting space without considering all possible person hypotheses. The algorithm starts with the complete set of all possible hypotheses, splits it along the dimension (row, column or size) with the biggest extend and calculates scores for the two new sets. These sets are inserted into a priority queue. In the next step the set with the highest score is taken from the queue and split again along the dimension with the biggest extend. This is continued until a set is found that contains only one hypothesis. With the right score function $S(\Omega)$, only few sets of hypothesis need to be checked which makes the detection very fast. The score function that Lehmann develops for FastPRISM is explained in great detail in (Lehmann et al., 2009).

4 FUSION OF VISIBLE AND THERMAL DATA

There are different possibilities to fuse visible and thermal images for people detection. The lowest level we look at in our evaluation is the feature level. We propose to calculate SURF features from both the visible spectrum image and thermal image and present them to our people detector as a combined set of image features. This method will be called *Merge Features* in the remainder of this work.

On a higher level, i.e. *mid level*, the detector itself incorporates both, visible and thermal data. We propose two modifications to the voting step of FastPRISM which draw hints from information in the thermal image and otherwise still rely on visible data for the detection. This fusion should be categorized somewhere between feature level and decision level.

The highest level of fusion, i.e. *decision level*, is to apply a people detector to the visible image and another detector to the thermal image and then fuse the results. This has the advantage that each detector can be developed and trained independently, and can be used on systems with only one of the data sources. The issue here is that the detectors need to transport their certainty about their detections to a fusion module. This module then somehow needs to figure out which detector to trust more, if they are disagreeing about a certain detection. It also has the responsibility to fuse detections, if they describe the same person.

4.1 Merge Features - Feature Level Fusion

The lowest level we look at is the feature level. This is, because the people detector we want to employ acts on image features only. For this fusion step we draw SURF features from both the visible spectrum and thermal image independently and add them into the same data structure. The SURF feature extractor works on greyscale images and can be applied to thermal images. The combined set of features from both spectra is presented to the people detector which is ignorant of where the single features came from.

4.2 Influencing the Voting Process - Mid Level Fusion

For the *mid level* fusion we assume that people are above a certain temperature. Because our thermal camera does only deliver intensities and not temperature readings, this translates to people being above a certain intensity value. This is close to the assumption that people are warmer than their surroundings, but not the same. The biggest difference is we use the intensity values to only guide the detector that otherwise is working on visible spectrum data. The goal here is to guide the branch and bound process away from image areas that are to cold to be a person and might be hard to distinguish from a person in the visible spectrum image. We propose two different ways to influence the voting of the ISM detector which modify the score function for a set of person hypotheses to guide the detector.

4.2.1 Maximum Search

First the original score $S(\Omega)$ for the hypothesis set Ω is calculated as in the original method from the features of the visible spectrum image. Then the rectangular region $R_\Omega = \bigcup_{h\in\Omega} R_h$ of the thermal image T is searched for the maximum value. R_Ω corresponds to the region that contains all rectangles R_h for person hypothesis $h \in \Omega$. If that maximum is below the thermal threshold t we assume that there cannot be a person in this region of the image and we lower the score for the hypothesis set. The equation is as follows:

$$S_{max}(\Omega) := \begin{cases} S(\Omega) & \max_{p\in R} T(p) \geq t \\ S(\Omega)*0.9 & \max_{p\in R} T(p) < t \end{cases}$$

4.2.2 Combination with Blob Detector

The *Blob vote* modification is the second *mid level* fusion technique we propose. It combines the *Blob Detector* and the FastPRISM detector at vote level. At first, again the score $S(\Omega)$ is calculated for the visible spectrum image as before. Then, additionally, the set of person rectangles $\Omega_{blob} := \{R_{blob} \in T\}$ is calculated with the *Blob Detector* from the thermal images. We again want to avoid image regions where persons are unlikely, because the *Blob Detector* did not find persons there. But we still want the trained person detector to filter out false detections from the *Blob Detector*. Therefor the image area of the hypothesis set is calculated as $R_\Omega = \bigcup_{h \in \Omega} R_h$ and compared to the rectangles in Ω_{blob}:

$$S_{blob}(\Omega) := \begin{cases} S(\Omega) * 1.1 & \exists R \in \Omega_{blob} : R \cap R_\Omega \neq \emptyset \\ S(\Omega) * 0.9 & \forall R \in \Omega_{blob} : R \cap R_\Omega = \emptyset \end{cases}$$

4.3 Join Detections - Decision Level Fusion

As indicated before the highest level of fusion is to have two detectors working independently on the visible spectrum and thermal image and then join the results. For the visible spectrum image we apply the FastPRISM detector, as usual. For the thermal image we have two choices. The first choice is to use the FastPRISM detector also for thermal images. The second choice is to use the *Blob Detector* we described in section 3.1. We will evaluate both, later on. The methods will be called *Join Detections* and *Join with Blobs*, respectively, and summarized as *decision level* fusion approaches.

For both methods the fusion of the results is done straight forward. The result rectangles are compared pairwise. If a pair is overlapping by more than 50% only the result with the better score is used and the score is set to the sum of the individual scores. Results without a correspondence are just copied with their original score.

5 EXPERIMENTS

The thermal images are taken from a thermal camera of type EYE R640 from OPGAL with a resolution of 640x480 pixels. It delivers intensity values that are automatically scaled to the currently observed temperature range, i.e. it is impossible to interpret the intensities as temperatures. For the images from the visible spectrum we mounted a common Logitech Quick Cam 9000 USB camera on top of the thermal camera, see Figure 1. The registration of the cameras was done manually by selecting corresponding points and calculating a transformation between the images. This is only an approximate solution, but sufficient for our needs.

We recorded a large indoor dataset in our lab with multiple persons moving in front of the camera. The robot with the mounted camera was placed in different locations so that different backgrounds are visible in the dataset. The dataset amounts to roughly 2500 images from each spectrum. Additionally, we recorded data outdoors on two different days with temperatures of 15°C and 25°C and sunny weather conditions. During recording the robot was moving. The outdoor datasets amount to 3500 images from each spectrum. Figure 3 shows some examples of our datasets.

5.1 Experiment Setup

For automatic evaluation of the detector results people were manually annotated as bounding boxes in the datasets. A result is counted as correct if the cut and join of the detection rectangle and annotated rectangle overlap by at least 50%.

We evaluate the *Merge Features* fusion, both proposed *mid level* fusion variants and both *decision level* fusion variants. As baseline algorithms we use the *Blob Detector* and our ISM variant on thermal and visible spectrum images, called *ISM Visible* and *ISM Thermal* respectively. These baseline algorithms perform very different on the three datasets and their performance issues help to understand the results of the proposed fusion variants.

5.2 Evaluation

We present the results of the evaluation in three graphs for the three different datasets. The graphs plot recall rate (i.e. correct detections of people) against the false positives per image (FPPI) rate while the detection threshold is altered. We cut off the graphs at one false positive per image. Our main goal with the fusion of thermal and visible spectrum data was not to raise the recall rate but to get less false positives, i.e. make the individual curves stay at a higher recall rate for lower FPPI rates.

From the indoor experiments, see Figure 4, we learned that the *Blob Detector* yields very good results. But the other proposed methods are quite good as well. All in all, the results on the indoor set are quite similar for all the approaches and the set seems

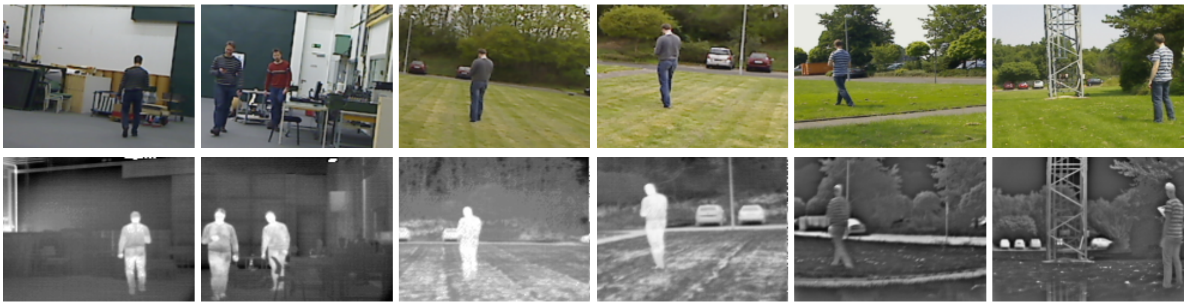

Figure 3: Sample images from the recorded datasets. The left two columns are from the indoor dataset, the middle columns from the outdoor dataset on the 15°C day and the right two columns from the outdoor dataset on the 25°C day.

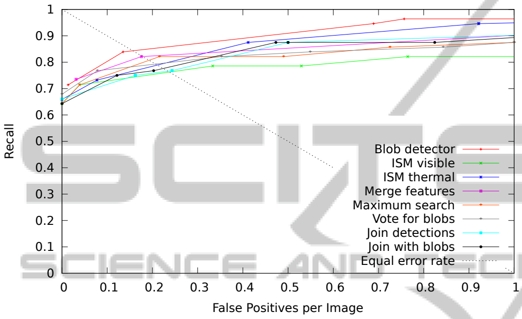

Figure 4: Results on the indoor test set. *Blob Detector* scores best, but all results are very similar.

to be easy enough. Nonetheless, we can see some effects. *ISM Visible* performs worst in terms of the recall rate, because it is affected by some background structures like shelves and racks which it tends to confuse with people. Those structures are less visible in the thermal images, because they are of a homogeneous temperature which is why *ISM Thermal* performs better. A look at the *mid level* fusion shows that they raise the recall rate a bit above the level of *ISM Visible*, because they prevent some of the false positive detections that still occur with a low threshold. This is what we wanted them to achieve. The *decision level* fusion cannot avoid false positives very good, which can be observed by the lower recall rates in the range between 0 and 0.4 FPPI. Above 6 FPPI the recall values of *Join Detections* are even higher than the one of the *Blob Detector*, which stays stable at 0.96 above 0.8 FPPI. But this are too many false positives to make a useful detector. Maybe a more advanced strategy to select the right detections from the single detectors would help here, but this is left as future work.

The outdoor experiments show a different picture as seen in Figure 5 and 6. The performance of the *Blob Detector* is not so good on the 15°C day and really bad on the 25°C day, because there are many things in the background that are heated up by the sunlight. *ISM Visible* has severe issues with false positives. There are many structures like trees and lampposts in the background that get confused with people. The *mid level* fusion variants show that they achieve our goal of reducing false positives. Their recall value goes down at a lower FPPI rate, i.e. the range between 0 and 0.4 FPPI. Especially for *Maximum Search* on the 15°C day the recall rate raises quite fast compared to *ISM Visible*, if one looks at 1.0 FPPI.

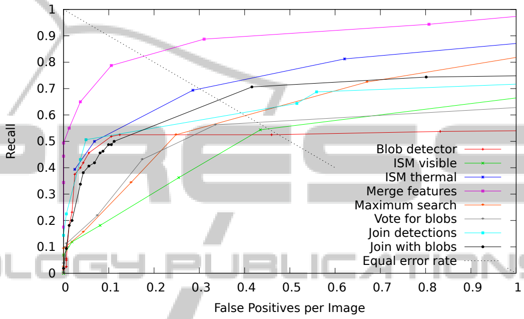

Figure 5: Results on the outdoor dataset from the 15°C day. *Merge Features* performs best. *Blob Detector* performance shows that the person now is not always warmer than the background.

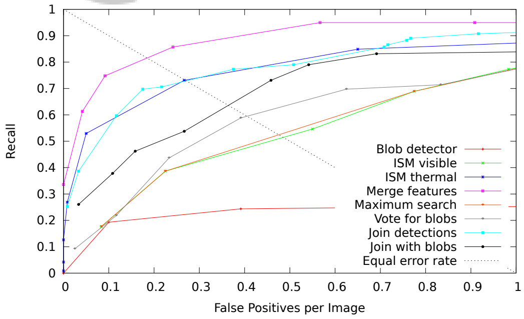

Figure 6: Results on the outdoor dataset from the 25°C day. *Merge Features* performs best. *Blob Detector* performance now is real bad.

On the 25°C day the performance difference between *Maximum Search* and *ISM Visible* is close to non existent, though. On that dataset only *Blob Vote* shows a little higher resistance against false positives,

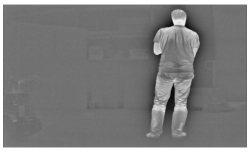

Figure 7: The image shows a clearly visible thermal structure of a person which helps people detection, because more features can be extracted on the person.

which is somewhat unexpected, because the *Blob Detector* itself performs extraordinarily bad. Most probably it detects some persons that ISM cannot see at all. The performance of *ISM Thermal* is quite good on these datasets. That is, because trees appear as homogeneous blobs in thermal data and do not generate many features. This helps to avoid them as false positives. Not surprisingly *Join Detections* seems to mostly take the detections from the thermal data into account and performs very similar. *Join with Blobs* works quite well here, too. The results of the combination are far better than the individual results which shows that *Blob Detector* and *ISM Visible* have very few right positives in common.

On both outdoor datasets the low level merge of features works best. This shows that the fusion of thermal and visible data makes sense for outdoor scenarios, especially on days where the assumption that people are the hottest spots in the images does not hold. This is, because the feature set incorporates features from both spectra. At the same time the ISM detector is very good at handling incomplete information, i.e. if one spectrum fails considerably, like the visible spectrum does in our setup, the other spectrum still can be sufficient to produce good detections. But the additional information from the visible spectrum seems to help the detector to even outperform *ISM Thermal* on this datasets.

We identified two reasons for the improved performance with thermal data. On the one hand, background clutter is often not as visible in thermal images as in images from the visible spectrum. That is because background clutter often has a common temperature which yields in big homogeneously colored blobs which are easy to filter out. On the other hand, the thermal signature of persons usually contains much more structure than the clothing, see Figure 7 for an example. This helps the SURF feature detector to extract more features for the observed persons and, therefore, improves the results. The low level fusion of data, i.e. merging of feature sets, works very good with the ISM detector. We think that this is because of the voting mechanism employed by ISM. Features from both spectra are voting for the right positives whereas for the false positives often only features from one spectrum vote.

6 CONCLUSIONS AND FUTURE WORK

We evaluated how to apply the ISM people detector to data from the thermal and visible spectrum at the same time. We identified three levels of fusion that are applicable for this detector. The lowest level just combines features from thermal and visible spectrum images into a combined feature set and applies the detector to this set. On a higher level we apply ISM to the visible data and use the knowledge from the thermal image to guide the voting process away from false positives and towards possible persons but still let the trained detector make the final decision. On the highest level we apply the ISM detector to images from the thermal and visible spectrum separately and try to join the results into a combined set of detections. These fusion techniques are evaluated on three datasets, one recorded indoors in a lab environment and two outdoor datasets on sunny days with ambient temperatures of 15°C and 25°C.

The conclusion of our evaluation is that the lowest level of fusion shows the best performance. It offers a low false positive rate with a high recall rate at the same time. The performance is much better than the one of the ISM detector applied to visible images only. It also is better than ISM applied to thermal images only, which already yields a good improvement. The other fusion approaches, namely *Merge Features*, *Maximum Search*, *Blob Vote*, *Join Detections*, and *Join with Blobs* perform similar to ISM on thermal data which suggests that the information in the thermal data is very dominant during the detection process. One reason is that people are easily distinguishable from the background in thermal images most of the time, because the background clutter in the visible domain generally has homogeneous thermal signature.

One aspect that was not investigated in this work is the fusion of thermal and visible spectrum images on the pixel level. It should be possible to design a combined feature vector for that and implement a smart feature detection method which avoids unlikely features already in the feature detection step. In future work we want to evaluate how the use of thermal imaging impacts the performance of people tracking and if persons can be distinguished using their thermal signature.

REFERENCES

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Surf: Speeded up robust features. *Computer Vision and Image Understanding (CVIU)*, Vol. 110:pp. 346–359.

Cielniak, G. and Duckett, T. (2004). People recognition by mobile robots. *Journal of Intelligent and Fuzzy Systems*, 15:21–27.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

Davis, J. W. and Sharma, V. (2004). Robust background-subtraction for person detection in thermal imagery. In *In IEEE Int. Wkshp. on Object Tracking and Classification Beyond the Visible Spectrum*.

Enzweiler, M. and Gavrila, D. (2009). Monocular pedestrian detection: Survey and experiments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31:2179–2195.

Gero andnimo, D., Lo andpez, A., Sappa, A., and Graf, T. (2010). Survey of pedestrian detection for advanced driver assistance systems. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(7):1239 –1258.

Goubet, E. (2006). Pedestrian tracking using thermal infrared imaging. *Proceedings of SPIE*.

Han, J. and Bhanu, B. (2007). Fusion of color and infrared video for moving human detection. *Pattern Recognition*, 40(6):1771 – 1784.

Ikemura, S. and Fujiyoshi, H. (2011). Real-time human detection using relational depth similarity features. In Kimmel, R., Klette, R., and Sugimoto, A., editors, *Computer Vision ACCV 2010*, volume 6495 of *Lecture Notes in Computer Science*, pages 25–38. Springer Berlin Heidelberg.

Königs, A. and Schulz, D. (2012). Evaluation of thermal imaging for people detection in outdoor scenarios. In *IEEE International Workshop on Safety, Security & Rescue Robotics (SSRR)*.

Lampert, C. H., Blaschko, M. B., and Hofmann, T. (2008). Beyond sliding windows: Object localization by efficient subwindow search. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

Lehmann, A., Leibe, B., and van Gool, L. (2009). Feature-centric efficient subwindow search. In *IEEE Conf. on Computer Vision (ICCV)*.

Lehmann, A., Leibe, B., and Van Gool, L. (2010). Fast PRISM: Branch and bound hough transform for object class detection. *Int. Journal of Computer Vision (IJCV)*, pages 1–23.

Leibe, B., Leonardis, A., and Schiele, B. (2008). Robust object detection with interleaved categorization and segmentation. *Int. Journal of Computer Vision (IJCV)*, Vol. 77:pp. 259–289.

Leibe, B., Seemann, E., and Schiele, B. (2005). Pedestrian detection in crowded scenes. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

Li, J., Gong, W., Li, W., and Liu, X. (2010). Robust pedestrian detection in thermal infrared imagery using the wavelet transform. *Infrared Physics and Technology*, 53.

Lietz, H., Thomanek, J., Fardi, B., and Wanielik, G. (2009). Improvement of the classifier performance of a pedestrian detection system by pixel-based data fusion. In Serra, R. and Cucchiara, R., editors, *AI*IA 2009: Emergent Perspectives in Artificial Intelligence*, volume 5883 of *Lecture Notes in Computer Science*, pages 122–130. Springer Berlin Heidelberg.

Miezianko, R. and Pokrajac, D. (2008). People detection in low resolution infrared videos. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*.

San-Biagio, M., Crocco, M., Cristani, M., Martelli, S., and Murino, V. (2012). Low-level multimodal integration on riemannian manifolds for automatic pedestrian detection. In *Information Fusion (FUSION), 2012 15th International Conference on*, pages 2223 –2229.

Schiele, B., Andriluka, M., Majer, N., Roth, S., and Wojek, C. (2009). Visual people detection: Different models, comparison and discussion. In *ICRA Workshop People Detection and Tracking*.

Serrano-Cuerda, J., Lopez, M., and Fernandez-Caballero, A. (2011). Robust human detection and tracking in intelligent environments by information fusion of color and infrared video. In *Int. Conf. on Intelligent Environments*.

Spinello, L. and Arras, K. (2012). Leveraging rgb-d data: Adaptive fusion and domain adaptation for object detection. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 4469 –4474.

Suzuki, S. and Abe, K. (1985). Topological structural analysis of digitized binary images by border following. *CVGIP*.

Zhang, L., Wu, B., and Nevatia, R. (2007). Pedestrian detection in infrared images based on local shape features. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*.