Human Detection Based on Infrared Images in Forestry Environments

Ahmad Ostovar, Thomas Hellström, Ola Ringdahl

Department of Computing Science
Umeå University, Sweden
{ahmado,thomash,ringdahl}@cs.umu.se

Abstract. It is essential to have a reliable system to detect humans in close range of forestry machines to stop cutting or carrying operations to prohibit any harm to humans. Due to the lighting conditions and high occlusion from the vegetation, human detection using RGB cameras is difficult. This paper introduces two human detection methods in forestry environments using a thermal camera; one shape-dependent and one shape-independent approach. Our segmentation algorithm estimates location of the human by extracting vertical and horizontal borders of regions of interest (ROIs). Based on segmentation results, features such as *ratio of height to width* and *location of the hottest spot* are extracted for the shape-dependent method. For the shape-independent method all extracted ROI are resized to the same size, then the pixel values (temperatures) are used as a set of features. The features from both methods are fed into different classifiers and the results are evaluated using *side-accuracy* and *side-efficiency*. The results show that by using shape-independent features, based on three consecutive frames, we reach a precision rate of 80% and recall of 76%.

Key words: Human detection, Thermal images, Shape-dependent, Shape-independent, Side-accuracy, Side-efficiency

1 Introduction

Detecting presence of humans is crucial for both manned and autonomous forestry machines. Sometimes it is difficult for machine operators to detect humans due heavy occlusion and focus on other tasks, and it is valuable to have a reliable system that alerts the operator (or automatically stops the current operations) if humans are present close to the machine. For an autonomous machine, this functionality is even more important. In this paper, we present methods for detection of humans in forestry environments using a thermal camera. Whereas pixel values in images from an RGB camera represent color, pixel values in thermal images represent temperature. Using thermal images has both advantages and disadvantages compared with RGB images [1, 2]. For example, shadows and lighting conditions do not affect the thermal images. On the other hand, the

resolution of thermal images are often lower, and the texture of objects do not appear in thermal images.

Similar to earlier works, regions of interest (ROIs) are first extracted by applying a segmentation method on thermal images such as thresholding based on pixel intensity, and then subject to human detection approach (shape-dependent or shape-independent) features for each ROI are generated [3–5]. Each ROI is then classified as containing or not containing a human by feeding the features into a classifier. We have developed and evaluated one shape-dependent and one shape-independent method. Shape-dependent methods are based on human characteristics [3, 6] such as shape, height, length and location of the head [7, 8]. Yasuno et al. [7] used the P-tile method for head detection. This method is based on the assumption that the object of interest covers a defined ratio of the image. As demonstrated by Bertozzi et al. [9] using lenient symmetry checks can filter out non-human objects by considering the assumption that humans are symmetrical objects. In our shape-dependent method, ratio of height to width and location of the hottest spot are used as features. Shape-independent methods typically use statistical characteristics of thermal images [1, 5] as features. Fang et al. [5] used inertial similarity among ROIs to construct a set of features. Inertial of ROIs are based on the pixel brightness values and distance of a pixel to the center after size normalization. In our case all pixel values of the extracted ROIs are used as features.

The rest of the paper is organized as follows. After a description of image collection in Section 2, methods for extraction of regions of interest (ROIs) and the two proposed human detection methods are introduced in Section 3. The evaluation method is described in Section 4 and results are presented in Section 5 and discussed in Section 6.

2 Image Collection

For acquisition of data we used the *Optris PI200*[1] camera with spectral range of 7.5 to 13 μm. Images were collected at temperatures ranging from -10 to +27 degrees Celsius, with different amount of occlusion in forestry environments to simulate real working conditions. Furthermore, to consider changes in heat radiation [8] several persons walking and sitting at different distances from the camera and with different types of clothing were included. Totally 1135 thermal images were acquired[2] and then divided into 23 sets based on ambient temperature. Two sets were removed due to the amount of frames containing heavily occluded persons (more than 50% of the human body). The reason for this occlusion was mainly that the snow was knee deep and all forestry objects were covered by snow, preventing human body heat radiation to reach the camera. In the remaining sets, 1042 thermal images were manually labeled such that if

[1] `http://www.optris.com/thermal-imager-pi200`, Optris Infrared Thermometers Products homepage, accessed 2015-11-15

[2] `https://archive.cs.umu.se/papers/2016-ThermalHumanDetection-AOstovar/`, A selection of thermal images

an image included a human, a region of interest was selected to cover the whole human body. 687 images included 754 humans and 355 of them contained only forestry objects and no humans.

3 Human Detection

Detection of humans in the thermal images is done in two steps. In step one, regions of interest (ROIs) containing possible humans are identified. In step two, the ROIs are classified as human or non-human using two different methods, denoted *Shape-dependent* and *Shape-independent*, as described below.

3.1 Extracting Regions of Interest (ROIs)

To extract ROIs as rectangles in a thermal image, adaptive thresholds are first computed. Then vertical and horizontal borders are extracted. Given a thermal image tm, an initial threshold $t = \mu_{tm} + 2\sigma_{tm}$ is defined, where μ and σ are mean and standard deviation respectively, for all pixels in the image. Then an adaptive threshold $t_A = \mu_{tm>t} + \sigma_{tm>t}$ is computed, where μ and σ are computed for all pixels larger than t.

To define vertical borders of the ROIs, a vector $\mathbf{T}^{col}_{max}$ of max values for each column is created. Then, a binary vector $\mathbf{T}^{col}_{bin}$ is defined, with 1 for columns for which $\mathbf{T}^{col}_{max} > t_A$. In the next step, we detect crossings in $\mathbf{T}^{col}_{bin}$ from 0 to 1 and 1 to 0 to determine left and right borders of the ROIs.

The performed thresholding and also occlusion caused by objects or clothing, may result in regions of interest being split into several smaller segments (this problem has also been recognized in [8]). Therefore, segments are merged if the distance between them is smaller than half the size of the largest involved segment. Moreover, if within 80 pixel distance of a border, any value in $\mathbf{T}^{col}_{max}$ is greater than the average of $\mathbf{T}^{col}_{max}$, the border is extended to that point. The number of pixels which are considered for any extension possibility is not critical and varying this number would present similar results. Small segments are less likely to contain a nearby human and due to the low resolution of the thermal camera cannot be accurately detected. Therefore, segments smaller than five pixels are ignored. Horizontal borders are extracted in a similar way as the vertical borders, but only the part of the image that falls inside the generated vertical borders is considered in this operation. Figure 1 illustrates extracted ROIs in a thermal image including two humans.

The extracted ROIs are primarily used as input to the detection methods, but can in itself also be seen as a simple detection algorithm. Performance for this algorithm is presented in Section 5 and in Table 1.

3.2 Shape-dependent Method

In this method, human characteristics including the ratio of height to width [3, 6], and the location of the hottest spot (head) [7, 8] in the ROI are extracted

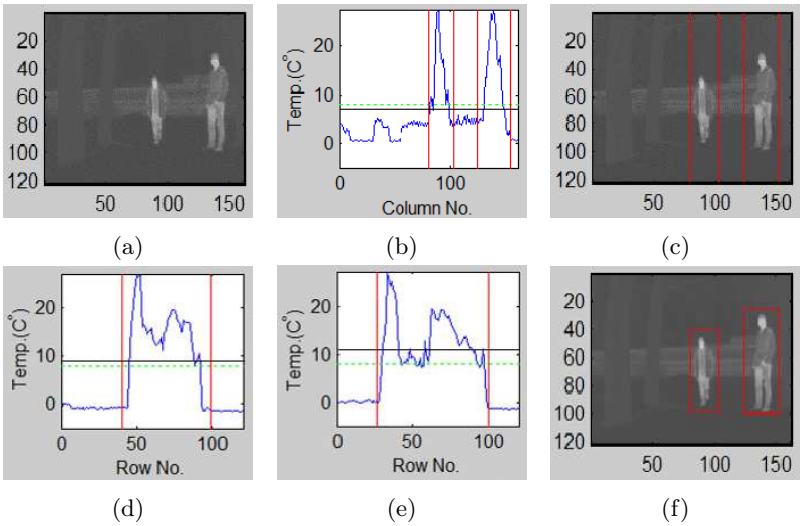

Fig. 1: Extracting ROIs. (a): A thermal image including two humans. (b): Vertical borders based on histogram of $\mathbf{T}^{col}_{max}$. (c): Vertical borders in thermal image. (d): Horizontal borders within the first vertical borders. (e): Horizontal borders within the second vertical borders. (f): Extracted ROIs. Red, green dashed and black lines are borders, adaptive threshold (t_A) and average of $\mathbf{T}^{col}_{max}$.

and used in a heuristically designed decision rule. To be classified as containing a human, the ratio of height to width must fall within the range of 0.8 to 1.8. These values are chosen as they are typical ratio for human in sitting and walking poses. A second requirement is that the hottest spot, presumable corresponding to the human head [7, 10], is located in the upper third of the bounding box. Performance for this method is described in Section 5 and in Table 1.

3.3 Shape-independent Method

In an initial step, the input ROI is resized to a fixed size of 25*35 pixels. These 875 pixels are input as features [1, 5] to kNN, SVM and Naive Bayes classifiers. For kNN, k=5 is used. To further improve classification performance, the outcome of the three classifiers are fused by using majority voting. For each classifier, we combine classifications of three consecutive frames (frames 1 to 6 are considered) with majority voting. The assumption is that the location and poses of humans change slightly within consecutive frames, but the object of interest will still remain within the same extracted ROI. Images in the data set are collected with a rate of 10 Hz, thus, the maximum time difference between three consecutive frames is about 0.3 second. Performance for this method is presented in Section 5 and in Table 1.

4 Evaluation Method

To evaluate the developed algorithms, 21-fold cross validation has been used. Hence, classifiers are built using 20 data sets, and evaluated using the remaining set. The final performance values are the mean values from 21 such evaluations. Each extracted ROIs is compared with the ROIs that have been manually labeled as consisting a human. For a ROI to be regarded as correct it must overlap with a labeled ROI in the same image. Computation of overlap is based on the two concepts *side-accuracy* and *side-efficiency* as illustrated in Figure 2, similar to performance indices introduced by Fang et al. [5]. Side-accuracy quantifies how much of the labeled ROI is covered by the extracted ROI, and is computed as the square root of the overlap area $S_{Overlap}$ divided by the area of the labeled ROI (S_{Human}). Side-efficiency quantifies how much of an extracted ROI is covered by the labeled ROI, and is defined as the square root of $S_{Overlap}$ divided by the area of extracted ROI (S_{ROI}). Both side-accuracy and efficiency are expressed as a percentage between 0% and 100%. In [11] it is suggested that both side-accuracy and side-efficiency should be greater than 50% in order to avoid accepting ROIs that barely overlap a labeled ROI, and maybe do not even contain the object of interest in the overlap area.

Hence, an extracted ROI is regarded as a true positive (TP) if it overlaps a labeled ROI with both side-accuracy and efficiency greater than 50%. Otherwise it is regarded as a false positive (FP). Also, an extracted ROI is considered as false positive if it is classified as human while there is no labeled ROI in the same image. If both side-accuracy and efficiency are greater than 50% but the extracted ROI is classified as non-human then it is counted as false negative (FN). Performance is computed using precision and recall as defined in Equations 1 and 2. Precision represents the fraction of all extracted ROIs that are correct. Recall represents the fraction of all labeled ROIs that are extracted by the system.

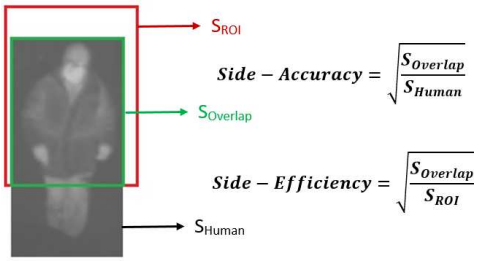

Fig. 2: Side-Accuracy and Efficiency definitions (adapted from [5]).

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

5 Experimental Results

In Table 1, cross-validated performance for four methods are presented and Table 2 demonstrates performance of shape-independent method based on different classifiers for one and three consecutive frames. The method for extraction of ROIs is used to generate input for the detection methods, but has in itself a recall rate of 84%, but only 59% precision. This means that ROIs for 84% of all humans are extracted, but also a lot of non-human ROIs are extracted. The shape-dependent method increases precision to 79%, but the recall rate is reduced to 38%. For the shape-independent method, precision is 80% and recall 74%. As presented in Table 2, the recall rate of the classifiers increased using three consecutive frames. Performance for the shape-independent method using three consecutive frames maintains the precision of 80% and increases the recall to 76%.

Performance is highly influenced by the ambient temperature. In Figure 3, performance for the shape-independent method is plotted versus ambient temperature. Recall is reduced in summer scenarios with higher temperatures than 7 degrees Celsius, while precision is roughly independent of temperature.

Table 1: Performance for algorithms for extraction of regions of interest (ROIs), the shape-dependent and the shape-independent method. Numbers in parentheses in the shape-independent method define the number of consecutive frames.

Performance	Precision	Recall
Extracted ROIs	59%	84%
Shape-dependent	79%	38%
Shape-independent (1)	80%	74%
Shape-independent (3)	80%	76%

Table 2: Performance of shape-independent method using SVM, kNN and Naive Bayes classifiers for one and three consecutive frames.

-	Shape-independent (1)		Shape-independent (3)	
Performance	Precision	Recall	Precision	Recall
SVM	80%	75%	82%	79%
kNN	80%	70%	80%	73%
Naive Bayes	75%	65%	75%	66%

6 Summary and Discussion

The best results, precision 80% and recall 76% were achieved by the shape-independent method using three consecutive frames. This method uses majority

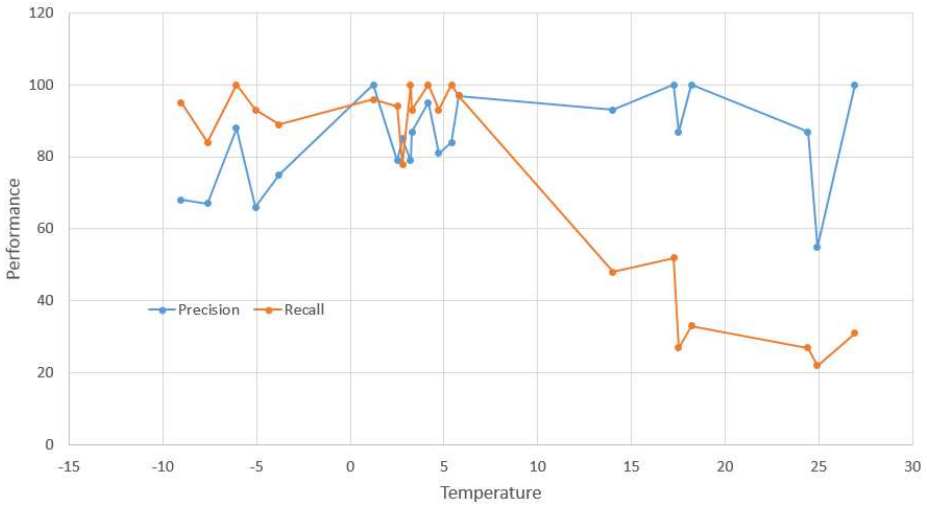

Fig. 3: Performance of shape-independent method based on ambient temperature. Recall is reduced in summer scenarios (higher temperatures than 10 degrees Celsius) while precision shows more stability against higher temperatures.

voting combining a Naive Bayes classifier, kNN, and SVM. The performance for shape-dependent method was lower than for the shape-independent method. One reason for this could be that it is impossible to set parameters that cover all human poses. This was also noted by Fang et al. [5]. Considering three consecutive frames improved performance. It supports the hypothesis that small movements of a human between consecutive frames can be utilized to reduce the effect of occlusion and increase performance.

The ambient temperature and degree of occlusion directly affect performance of human detection using thermal images [2, 12]. For the 14 data sets with temperature below 7 degrees Celsius, recall is above 80% and precision above 60% (see Figure 3). In higher ambient temperature, objects other than humans produce additional bright areas in thermal images. These additional source of heat reduce the human detection performance based on pixel values of thermal images (shape-independent approach). Research by Nanda and Davis [3] report that to get 90% recall rate in summer, the false discovery rate has to be raised to 100%. Xu and Fujimura [12] detect humans in a winter scenarios and report a precision of 97.37%, but with a recall rate no more than 35%. The introduced shape-independent method showed satisfactory performance in temperatures below 7 degrees Celsius. However, to be used in a safety system further improvements are necessary. We have considered to present performance for the shape-independent method using receiver operator characteristic (ROC) curve, but there are no suitable parameters to construct basis for the analysis. To enhance performance, additional features such as other morphological characteristics of humans [13], histograms of oriented gradient (HOG) [14] and inertial features [5] should be

evaluated. Reconstruction of occluded humans using morphological approaches might also improve performance and should be investigated as possible extensions to the proposed methods.

Since the extracted ROIs are used as input to the other detection methods, their recall rate cannot exceed the 84% recall rate of the extraction algorithm. To further improve performance, further development of the extraction algorithm should therefore also be considered.

References

1. Fang, Y., Yamada, K., Ninomiya, Y., Horn, B., Masaki, I.: Comparison between infrared-image-based and visible-image-based approaches for pedestrian detection. In: Intelligent Vehicles Symposium, 2003. Proceedings. IEEE. (2003) 505–510
2. Gavrila, D., Giebel, J.: Shape-based pedestrian detection and tracking. In: Intelligent Vehicle Symposium, 2002. IEEE. Volume 1. (June 2002) 8–14 vol.1
3. Nanda, H., Davis, L.: Probabilistic template based pedestrian detection in infrared videos. In: Intelligent Vehicle Symposium, 2002. IEEE. Volume 1. (2002) 15–20
4. Bertozzi, M., Broggi, A., Ghidoni, S., Meinecke, M.: A night vision module for the detection of distant pedestrians. In: Intelligent Vehicles Symposium, 2007 IEEE. (June 2007) 25–30
5. Fang, Y., Yamada, K., Ninomiya, Y., Horn, B., Masaki, I.: A shape-independent method for pedestrian detection with far-infrared images. Vehicular Technology, IEEE Transactions on **53**(6) (2004) 1679–1697
6. Kancharla, T., Kharade, P., Gindi, S., Kutty, K., Vaidya, V.: Edge based segmentation for pedestrian detection using nir camera. In: Image Information Processing (ICIIP), 2011 International Conference on. (2011) 1–6
7. Yasuno, M., Ryousuke, S., Yasuda, N., Aoki, M.: Pedestrian detection and tracking in far infrared images. In: Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE. (Sept 2005) 182–187
8. Xu, F., Liu, X., Fujimura, K.: Pedestrian detection and tracking with night vision. Intelligent Transportation Systems, IEEE Transactions on **6**(1) (March 2005) 63–71
9. Bertozzi, M., Broggi, A., Fascioli, A., Graf, T., Meinecke, M.: Pedestrian detection for driver assistance using multiresolution infrared vision. Vehicular Technology, IEEE Transactions on **53**(6) (Nov 2004) 1666–1678
10. Meis, U., Oberlander, M., Ritter, W.: Reinforcing the reliability of pedestrian detection in far-infrared sensing. In: Intelligent Vehicles Symposium, 2004 IEEE. (June 2004) 779–783
11. Wang, D., Posner, I., Newman, P.: What could move? finding cars, pedestrians and bicyclists in 3d laser data. In: Robotics and Automation (ICRA), 2012 IEEE International Conference on. (May 2012) 4038–4044
12. Xu, F., Fujimura, K.: Pedestrian detection and tracking with night vision. In: Intelligent Vehicle Symposium, 2002. IEEE. Volume 1. (June 2002) 21–30 vol.1
13. Bertozzi, M., Broggi, A., Grisleri, P., Graf, T., Meinecke, M.: Pedestrian detection and tracking with night vision. In: Intelligent Vehicles Symposium, 2003. Proceedings. IEEE. (2003) 662–667
14. Suard, F., Rakotomamonjy, A., Bensrhair, A., Broggi, A.: Pedestrian detection using infrared images and histograms of oriented gradients. In: Intelligent Vehicles Symposium, 2006 IEEE. (June 2006) 206–212