Human Body Detection and Geolocalization for UAV Search and Rescue Missions Using Color and Thermal Imagery

Piotr Rudol and Patrick Doherty
Department of Computer and Information Science
Linköping University, SE-58183 Linköping, Sweden
{pioru, patdo}@ida.liu.se

Abstract—Recent advances in the field of Unmanned Aerial Vehicles (UAVs) make flying robots suitable platforms for carrying sensors and computer systems capable of performing advanced tasks. This paper presents a technique which allows detecting humans at a high frame rate on standard hardware onboard an autonomous UAV in a real-world outdoor environment using thermal and color imagery. Detected human positions are geolocated and a map of points of interest is built. Such a saliency map can, for example, be used to plan medical supply delivery during a disaster relief effort. The technique has been implemented and tested on-board the UAVTech [1] autonomous unmanned helicopter platform as a part of a complete autonomous mission. The results of flight-tests are presented and performance and limitations of the technique are discussed.

TABLE OF CONTENTS

1. INTRODUCTION

Unmanned Aerial Vehicles have become more and more common and are able to perform missions with increasing levels of complexity. At the same time, they require less human operator involvement due to the increase in autonomous behavior. Flying robots can perform a wide range of tasks which are considered dirty, dull, or dangerous by humans. Missions such as search and rescue or surveillance, where camera coverage of a given area must be guaranteed, are examples of such tiresome tasks. They require high precision and endurance from human pilots. In the case of emergency situations such as natural disasters, finding potential survivors requiring medical attention is of utmost importance. Such missions require high flight precision and long operation times and this is tedious for human pilots. Our UAV systems can autonomously plan and execute complete missions from take-off to landing [1], where, for example, video footage of every square meter of an area of interest must be collected.

In order to further reduce human involvement and speed up the process of finding causalities, the task of analyzing collected video can be delegated to an automated algorithm which analyzes the video footage in real time, on-line. An algorithm which can identify and geographically locate places where human bodies can be found is required to achieve such a task.

The technique presented in this paper takes advantage of two video cameras. One of them delivers thermal video and the second one is a standard color camera. Commercially available, low-cost thermal cameras are not sufficient to classify humans at larger distances (40 meters) because of low image resolution and quality. A human body becomes just a blob and it is hard to distinguish it from any other object of the same size. Human detection algorithms working with color imagery also give best results at low distances and often have to rely on downsizing of images to achieve high rate of detection. The technique presented detects humans at a rate up to 25Hz (sporadically lower for scenes with high numbers of potential bodies) by first analyzing an infrared image to find human-temperature silhouettes and then using the corresponding color image regions to classify human bodies. Thanks to the high processing rate, the certainty of a correct classification can be assessed by collecting statistics over human body positions. The algorithm presented is suitable for real world operation on-board a UAV platform. Details of the technique are presented in section 4. Sections 5 and 6 present an example flight test setup and results for the technique used on-board the UAVTech helicopter as part of a fully autonomous mission.

1-4244-1488-1/08/$25.00 ©2008 IEEE
IEEEAC Paper #1274, Version 1 Updated 13/12/2007.
[1] Autonomous Unmanned Aerial Vehicle Technologies Lab, Linköping University, Sweden, http://www.ida.liu.se/~patdo/auttek/

2. Related work

The task of observing and analyzing human appearance and movement has been of interest to the computer vision community for many years. Techniques can be categorized in many ways. One of them is the need for pre-processing, such as background subtraction, which can be achieved by frame differencing [2]. Other factors include the types of features which are needed for describing human appearance e.g. shape, color, contour. A considerable amount of work is based on the idea of detecting humans by parts. For example humans can be modeled as assemblies of parts which are detected separately and represented by co-occurrences of local features [3]. A cascade-of-rejectors with variable size blocks of histograms of oriented gradients as features can also be used. AdaBoost is used as a feature selection technique to choose appropriate blocks from a large set of possible blocks. The use of integral image representation and a rejection cascade allows for 5 to 30 Hz human detection performance (for images of 320x280 pixels size) [4]. Another approach takes advantage of a classifier which is a cascade of boosted classifiers working with Haar-like features. The classifier is learned using boosting [5] and its details are presented in section 4.

Detecting humans in thermal imagery poses additional challenges such as lower resolution, halos around hot or cold objects and smudging artifacts in case of camera movement. An approach which first performs a fast screening procedure using a template to locate potential person locations, which is then tested using an AdaBoosted ensemble classifier using automatically tuned filters has been proposed [6]. The technique, however, has been tested on footage collected by a stationary thermal camera and therefore the applicability of the technique to a moving camera is unknown.

Techniques using both color and thermal images have been suggested. One example uses color and infrared cameras and a hierarchical scheme to find a correspondence between the preliminary human silhouettes extracted from both cameras using image registration in static scenes. Authors also discuss strategies for probabilistically combining cues from registered color and thermal images [7]. A technique for detecting and tracking moving targets in overlapping electro-optical and infrared sensors by a probabilistic framework for integrating multiple cues from multiple sensors has been proposed [8]. The method has been tested on footage collected by a UAV but its computational requirements are not discussed and its usability onboard a UAV is unknown.

Finding humans from air vehicles in outdoor environments is receiving more and more attention. A summary and a discussion about information flow requirements for Wilderness Search and Rescue which takes advantage of Micro Aerial Vehicles (MAVs) is presented in [9].

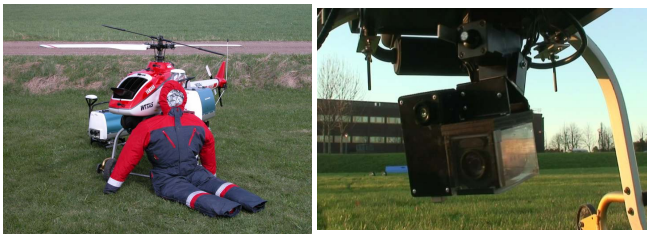

Figure 1. The UAVTech UAV and the on-board camera system mounted on a pan-tilt unit.

3. Hardware platform

The UAVTech UAV platform [10] is a slightly modified Yamaha RMAX helicopter (Fig. 1). It has a total length of 3.6 m (including main rotor) and is powered by a 21hp two-stroke engine with a maximum takeoff weight of 95 kg. The on-board system contains three PC104 embedded computers. The primary flight control (PFC) system includes a Pentium III 700Mhz, a wireless Ethernet bridge, a GPS receiver, and several additional sensors including a barometric altitude sensor. The PFC is connected to the RMAX helicopter through the Yamaha Attitude Sensor (YAS) and Yamaha Attitude Control System (YACS), an image processing computer and a computer responsible for deliberative capabilities. The deliberative/reactive system (DRC) runs on the second PC104 embedded computer (Pentium-M 1.4GHz) and executes all high-end autonomous functionalities such as mission or path planning. Network communication between computers is physically realized with serial lines RS232C and Ethernet.

The image processing system (IPC) runs on the third PC104 embedded Pentium III 700MHz computer. The camera platform suspended under the UAV fuselage is vibration isolated by a system of springs. The platform consists of a Sony CCD block camera FCB-780P and a ThermalEye-3600AS [11] miniature infrared camera mounted rigidly on a Pan-Tilt Unit (PTU) as presented in Fig. 1. The video footage from both cameras is recorded at a full frame rate by two miniDV recorders to allow processing after a flight.

Camera calibration

In order to find corresponding pixels in both images, as well as calculate a geographical location of a classified human body, both cameras have been calibrated to find their intrinsic and extrinsic parameters. The color camera has been calibrated using the Camera Calibration Toolbox for Matlab [12]. The same toolkit alone could not be used for finding optical parameters of the thermal camera because it was infeasible to obtain sharp images of the chessboard calibration pattern. The results of image undistortion gave poor results. To find focal length, principal point and the lens distortion parameters, a custom calibration pattern and an add-on to the toolkit have been used [13]. A specially prepared pattern has been fabricated to allow using the calibration procedure normally used for color images. The calibration setup is schematically

depicted in Fig. 2A. The custom calibration pattern (b) made out of a thin material (e.g. a sheet of overhead plastic) is placed between a warmed up (or cooled down) metal plate (a) and the camera to be calibrated. Black parts of the original calibration pattern are cut out from the plastic to allow the heat to pass through to produce an image similar to one obtained during a standard calibration of a color camera. The resulting image (after inverting colors) is shown in Fig. 2B.

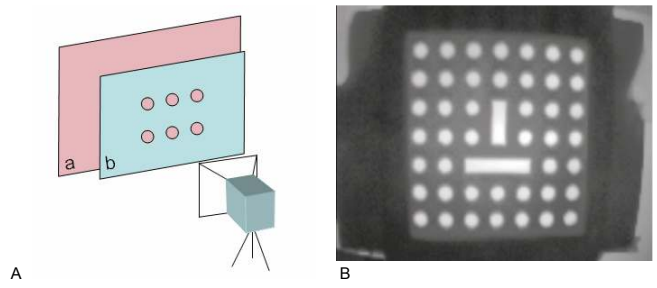

Figure 2. Thermal camera optical parameters calibration. A. Schematic view of the procedure. B. Example image.

4. IMAGE PROCESSING

Video footage collected by a UAV differs substantially from images acquired on the ground and the use of standard techniques is not straight forward. The following aspects and assumptions were taken into account when designing the image processing algorithm. First of all, a typical distance from a camera to an object of interest is larger than in the case of standard "ground" techniques (e.g. office-like environments). Additionally, as a UAV flies over an object of interest, it stays a short time in the camera field of view depending on the vehicle velocity. Additional footage of an object might require a UAV to return to the point of interest. It can, depending on the platform, be time consuming. Both maximum and minimum speeds are determined by an aircraft's properties. Nevertheless, high flight speed is preferred in case of search and rescue applications. Therefore it is essential for the image processing algorithm to perform close to the full frame rate to process all frames of the video. The flight altitude depends on a camera's resolution and field of view. For a standard PAL (768x576) resolution and a 50 degrees field of view, the maximum flight altitude is approximately 50 meters in order to obtain body sizes no less than 30 pixels. Higher resolution cameras would allow flight at higher altitudes. Flying at lower altitudes makes the task of image processing easier since a human body appears larger in a video frame. On the other hand the flight at a lower altitude requires more time to complete since the camera covers a smaller area per video frame.

The algorithm takes as input two images (camera planes are assumed to be close to parallel to the earth plane) and the processing starts by analyzing the thermal image. The image is first thresholded to find regions of human body temperature. Shapes of the regions are analyzed and those which do not resemble a human body (i.e. wrong ratio of minor and major axes of the fitted ellipse and incorrect area) are rejected. Additionally, regions which lie on the image border are rejected as they may belong to a bigger warm object. Once human body candidates are found in the thermal image, corresponding regions in the color image are calculated.

Computation of the corresponding region in the color image could be achieved by performing image registration or feature matching in both images. The former technique is too time consuming and the latter is infeasible because of mostly different appearance of features in color and thermal images. Pixel correspondences could also be encoded in a set of lookup tables depending on the distance to the object of interest. Such a solution would require additional memory for accurately covering a depth of interest. Here, a closed form solution is used which takes into account information about the UAV's state.

Computation of the corresponding region in the color image starts with calculating coordinates of a point T ($\tilde{v_T}$) whose projection is the pixel in the thermal image $\tilde{u_t}$ i.e.

$$\tilde{u_t} = P_t \tilde{v_T} \quad \tilde{u_t} \in \mathcal{P}^2 \quad \tilde{v_T} \in \mathcal{P}^3 \tag{1}$$

where P_t represents extrinsic and intrinsic parameters of the thermal camera. The general scheme of the problem is shown in Figure 3. A line equation with the direction vector $\tilde{v}_{cam}$

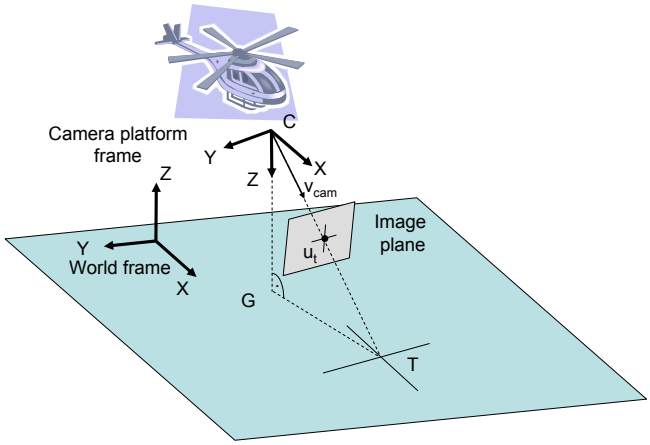

Figure 3. Calculation of a target coordinates.

which goes through camera center through pixel $\tilde{u_t}$ and inter-

sects the ground plane in point T is:

$$\tilde{v}_T - \tilde{v}_C = t \cdot \tilde{v}_{cam} \quad t \in \mathbb{R} \tag{2}$$

The ground plane is defined by the point G($\tilde{v}_G$) and the normal vector $\tilde{n}$ which is the down component of the NED (North, East, Down) frame:

$$(\tilde{v}_T - \tilde{v}_G) \cdot \tilde{n} = 0 \tag{3}$$

Finally, the vector $\tilde{v}_T$ which describes the point of intersection of a ray of light going through the camera center and the pixel of the target can be calculated according to:

$$\tilde{v}_T = \tilde{v}_C + \frac{(\tilde{v}_G - \tilde{v}_C) \cdot \tilde{n}}{\tilde{v}_{cam} \cdot \tilde{n}} \cdot \tilde{v}_{cam} \tag{4}$$

In order to calculate $\tilde{v}_{cam}$ the vector along the X axis of the camera frame must be expressed in the world coordinate frame. This transformation can be expressed as:

$$^{w}\tilde{v}_{cam} = P_{heli} P_{ptu} P_p \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T \tag{5}$$

where P_p describes the transformation depending on the undistorted pixel position $\tilde{u_t}$. Matrix P_{ptu} is built to represent a transformation introduced by the pan-tilt unit. P_{heli} represents the attitude of the UAV and is built up from roll, pitch and yaw angles delivered by the YAS system.

The method presented can be extended to relax the flat world assumption. The point T can be found by performing ray-tracing along the line described by equation Eq. 2 to find the intersection with the ground elevation map.

Calculated world position can additionally be checked against the on-board geographic information database to verify whether the calculated point is valid. Depending on the situation, certain positions can be excluded from the map. If the world position is accepted, its projection is calculated for the color camera using the following formula:

$$\tilde{u}_c = P_c \tilde{v}_T \quad \tilde{u}_c \in \mathcal{P}^2 \quad \tilde{v} \in \mathcal{P}^3 \tag{6}$$

where P_c constitutes the matrix encoding intrinsic and extrinsic parameters of the color camera.

The classifier

Once the corresponding pixel in the color image is identified, a sub-window with the pixel P_c in the center is selected and it is subjected to an object detector first suggested in [5]. The work was used as a basis for several improvements, one of which was presented in [14]. One of these included extending the original feature set which is presented in Fig. 4.

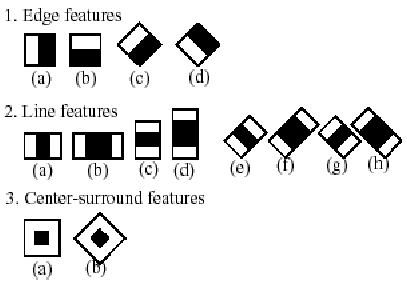

Figure 4. Leinhart's extended set of available features

The classifier which is in fact a cascade of boosted classifiers working with Haar-like features requires training with a few hundred positive and negative examples. During learning the structure of a classifier is learned using boosting. The use of a cascade of classifiers allows for dramatic speed up of computations by skipping negative instances and only computing features with high probability for positive classification. The speed up comes from the fact that the classifier, as it slides a window at all scales, works in stages and is applied to a region of interest until at some stage the candidate is rejected or all the stages are passed. This way, the classifier quickly rejects subregions which most probably do not include features needed for positive classification (i.e. background processing is quickly terminated). The classifier works with features which can be quickly extracted using intermediate image representations - integral images. The reason for working with features instead of pixel intensities is that features encode knowledge about the domain, which is difficult to learn from raw input data. The features encode the existence of oriented contrasts between regions of an image. The Haar-like features used here can be calculated at any position and any scale in constant time using only eight look-ups in the integral image.

The classifier parameters have been adjusted to minimize false negative cases. In case of rescue operations it is better to find more false positives then missing potential victims. The number of neighboring rectangles needed for successful identification has been set to 1 which makes the classifier accept very weak classifications. The factor by which the search window is scaled between the subsequent scans has been set to 1.2 meaning that the search window is increased by 20%.

The classifier used in this work is a part of the Open Source Computer Vision Library [15] and the trained classifier for upper-, lower- and full human body is a result of [16]. The trained classifier is best suited for pedestrian detection in frontal and backside views which is exactly the type of views

a UAV has when flying above the bodies lying on the ground.

Since the body classifier is configured to be "relaxed" it delivers sporadic false positive classifications. To counter for most of them the following method is used to prune the results. Every salient point in the map has two parameters which are used to calculate certainty of a location being a human body: T_{frame} which describes the amount of time a certain location was in the camera view and T_{body} which describes the amount of time a certain location was classified as a human body. The certainty factor is calculated as follows:

$$p_{body}(loc_i) = \frac{T_{body}}{T_{frame}} \tag{7}$$

A location is considered a body if $p_{body}(loc_i)$ is larger than a certain threshold (e.g. 0.5 during the flight tests) and T_{frame} is larger than a desired minimal observation time. Locations are considered equal if geographical distance between them is smaller then a certain threshold (depending on the geolocation accuracy) and the final value of a geolocalized position is an average of the observations (c.f. Section 6).

5. Experimental setup

Several flight tests were performed in a test field of the Swedish Rescue Services Agency which is used by rescue services, such as fire-fighters, police and medical personnel to train for rescue routines. The flight presented took place over several types of terrain such as asphalt and gravel roads, grass, trees, water and building roof tops which resulted in a variety of textures in the images. Generation of the saliency map was performed as part of a fully autonomous mission carried out by two UAVs over a search area of 290x185 meters. The results of one of the UAVs are presented here. The total number of eleven bodies (both human and dummies with close to human temperature) were placed in the area. The mission general plan is presented in Fig. 5. Before take-off one of the UAVs, given a scan area (dashed line polygon), planed scanning flight paths over the search area for both helicopters. The mission started with a simultaneous autonomous take-off at positions H_1 and H_2 and the UAVs flew to scanning starting positions S_1 and S_2. Throughout the flight the saliency map was being built until the UAVs reached ending positions E_1 and E_2. The mission finished by returning to the take-off position for a simultaneous landing. The mission took approximately ten minutes to complete and each UAV travelled a distance of around 1km.

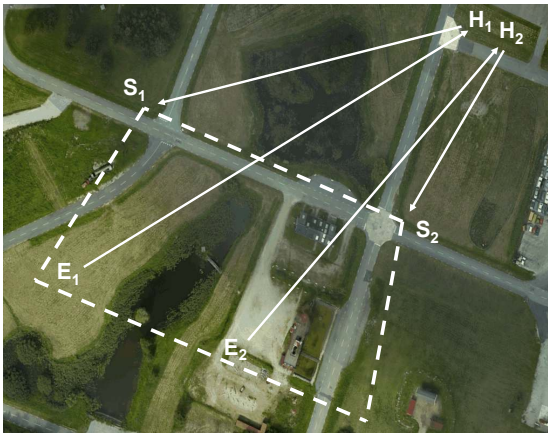

Figure 5. Mission overview.

6. Experimental results

The algorithm found all eleven bodies placed in the area. The images of identified objects are presented in Fig. 7. Several positions were rejected as they were not observed long enough (i.e. 5 seconds). Images 7, 9, and 14 present three falsely identified objects. Erroneous classifications were caused by configuring the human body classifier to accept weak classifications. A more restrictive setup could result in missing potential victims. Both human bodies and dummies were detected despite the lower temperature of the latter.

The accuracy of the body geolocation calculation was performed by measuring GPS (without differential correction) positions of bodies after an experimental flight. Figure 6 presents the error measurement for seven geolocated objects.

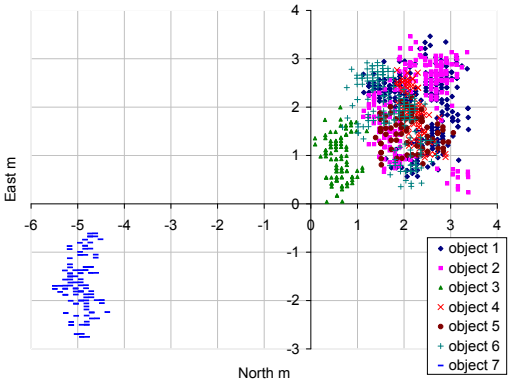

Figure 6. Geolocation error for multiple objects.

Figure 7. Images of classified bodies. Corresponding thermal images are placed under color images.

The measurement has a bias of approximately two meters in both east and north directions. It is a sum of errors in GPS measurement, accuracy of the camera platform mounting, PTU measurement, and camera calibration inaccuracies. The spread of measurement samples of approximately 2.5 meters in both east and north directions is a sum of errors of the UAV's attitude measurement, the system of springs in the camera platform, the flat ground assumption, and time differences between UAV state estimate, PTU angle measurement and image processing result acquisition. A detailed analysis is required to accurately measure error contributing factors and improve the precision. Nevertheless, the current accuracy of the system is sufficient for assessing a victim's position within 3 meters radius. A large geolocation error of object 7 is caused by the erroneous GPS measurement. Object 7 was located on a metal foot-bridge and the GPS antenna during static measurement was additionally partially occluded by metal railings. The noise on the measurement however is consistent with the rest of the objects.

7. Conclusions and Future Work

The algorithm presented solves a task of finding humans lying or sitting on the ground in video sequences collected onboard an unmanned aerial vehicle. The technique presented uses two video sources (thermal and color) and allows for high rate human detection at larger distances then in the case of using the video sources separately with standard techniques. The high processing rate is essential in case of video collected onboard a UAV in order not to miss potential objects as a UAV flies over it. A thermal image is analyzed first to find human body sized silhouettes. Corresponding regions in a color image are subjected to a human body classifier which is configured to allow weak classifications. This focus of attention allows for maintaining a body classification at a rate up to 25Hz. This high processing rate allows for collecting statistics about classified humans and pruning false classifications of the "weak" human body classifier. Detected human bodies are geolocalized on a map which can be used to plan supply delivery. The technique presented has been tested onboard the UAVTech helicopter platform and is a part of an au-

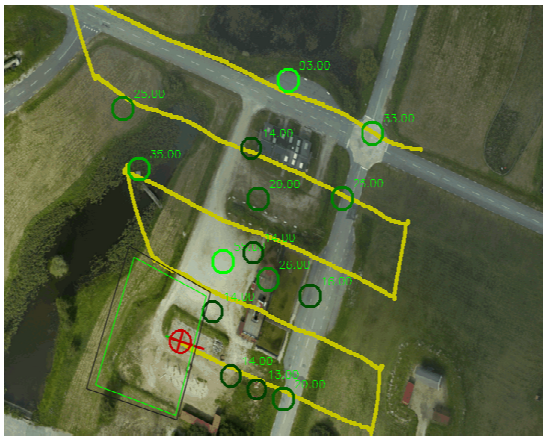

Figure 8. Flight path and geolocated body positions.

tonomous search and rescue mission. Details of the complete mission (supply delivery planning etc.) can be found in [17]. The ongoing work includes integration of a winch system on the UAV platform for delivering packages to victims detected and geolocalized by the technique presented.

ACKNOWLEDGMENTS

This work is supported in part by the National Aeronautics Research Program NFFP04 S4202 and the Swedish Foundation for Strategic Research (SSF) Strategic Research Center MOVIII.

REFERENCES

[1] M. Wzorek, G. Conte, P. Rudol, T. Merz, S. Duranti, and P. Doherty, "From Motion Planning to Control - A Navigation Framework for an Autonomous Unmanned Aerial Vehicle," in *Proc. of the 21th Bristol International UAV Systems Conference*, 2006.

[2] D. J. Lee, P. Zhan, A. Thomas, and R. Schoenberger, "Shape-based human intrusion detection," in *SPIE International Symposium on Defense and Security, Visual Information Processing XIII*, 2004.

[3] K. Mikolajczyk, C. Schmid, and A. Zisserman, "Human Detection Based on a Probabilistic Assembly of Robust Part Detectors," in *European Conference on Computer Vision*, 2004.

[4] Q. Zhu, M. C. Yeh, K. T. Cheng, and S. Avidan, "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients," in *Computer Vision and Pattern Recognition*, 2006.

[5] P. Viola and M. J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," in *Proc. of Conference on Computer Vision and Pattern Recognition*, 2001.

[6] J. W. Davis and M. A. Keck, "A Two-Stage Template Approach to Person Detection in Thermal Imagery," in *Workshop on Applications of Computer Vision*, 2005.

[7] J. Han and B. Bhanu, "Detecting moving humans using color and infrared video," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2003.

[8] J. Kang, K. Gajera, I. Cohen, and G. Medioni, "Detection and Tracking of Moving Objects from Overlapping EO and IR Sensors," in *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, 2004.

[9] M. A. Goodrich, J. L. Cooper, J. A. Adams, C. Humphrey, R. Zeeman, and B. G. Buss, "Using a Mini-UAV to Support Wilderness Search and Rescue Practices for Human-Robot Teaming." in *To appear in Proceedings of the IEEE International Conference on Safety, Security and Rescue Robotics*, 2007.

[10] P. Doherty, "Advanced Research with Autonomous Unmanned Aerial Vehicles," in *Proc. of the Int. Conf. on the Principles of Knowledge Representation and Reasoning*, 2004, pp. 731–732.

[11] L-3 Communications, http://www.l-3com.com/.

[12] J.-Y. Bouguet, "Camera calibration toolbox for matlab." http://www.vision.caltech.edu/bouguetj/calib_doc.

[13] C. Wengert, "A fully automatic camera and hand eye calibration," http://www.vision.ee.ethz.ch/~cwengert/calibration_toolbox.php.

[14] R. Lienhart and J. Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection," in *Proc. of International Conference on Image Processing*, 2002, pp. 900–903.

[15] "Opencv," http://www.intel.com/technology/computing/opencv/.

[16] H. Kruppa, M. Castrillon-Santana, and B. Schiele, "Fast and robust face finding via local context." in *Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, October 2003.

[17] P. Doherty and P. Rudol, "A uav search and rescue scenario with human body detection and geolocalization," in *Australian Conference on Artificial Intelligence*, 2007, pp. 1–13.

BIOGRAPHY

Patrick Doherty is a Professor at the Department of Computer and Information Science (IDA), Linköping University (LiU), Sweden. He is director of the Artificial Intelligence and Integrated Computer Systems Division at IDA and his research interests are in the area of knowledge representation, automated planning, autonomous systems, approximate reasoning and UAV technologies.

Piotr Rudol is a graduate student at Linköping University. His research interests include non-GPS navigation techniques for indoor MAVs and mapping and obstacle avoidance methods for outdoor UAVs. Among other things, he works with sensor integration for environment sensing allowing UAV operation in unknown environments.