Fast and viewpoint robust human detection for SAR operations

Paul Blondel, Alex Potelle, Claude Pégard, Rogelio Lozano

Fast and viewpoint robust human detection for SAR operations

Paul Blondel
Université Picardie Jules-Vernes
80000 Amiens, France
Email: paul.blondel@u-picardie.fr

Alex Potelle and Claude Pégard
Université Picardie Jules-Vernes
80000 Amiens, France
Email: firstname.lastname@u-picardie.fr

Rogelio Lozano
Université Technologique de Compiègne
60200 Compiègne, France
Email: rogelio.lozano@hds.utc.fr

***Abstract*—There are many advantages in using UAVs for search and rescue operations. However, detecting people from a UAV remains a challenge: the embedded detector has to be fast enough and viewpoint robust to detect people in a flexible manner from aerial views. In this paper we propose a processing pipeline to 1) reduce the search space using infrared images and to 2) detect people whatever the roll and pitch angles of the UAV's acquisition system. We tested our approach on a multimodal aerial view dataset and showed that it outperforms the Integral Channel Features (ICF) detector in this context. Moreover, this approach allows real-time compatible detection.**

I. Introduction

UAVs are interesting tools for finding people in distress in complex and/or in large environments because of their maneuverability, their rapidity of deployment and because they can be intelligently controlled. They can work automatically and in swarm in order to shrink the search time; but for this, each UAV should have its own detection system. Unfortunately, automatically detecting people from a UAV is not an easy task. Numerous constraints have to be taken into account in this context and the detection system has to be designed accordingly.

A. Existing work in human detection

Detection methods based on background substraction are not well suited when the camera undergoes complex movements or when people are not moving. The detection process should only rely on the visual information contained in one frame. Monolithic and part-based detectors fit this condition.

1) Monolithic detection: Monolithic detectors search for monolithic parts of the image looking like people. Gavrila et al proposed to use a hierachy of human contour templates obtained using training [1]. This hierachy, used together with the chamfer matching algorithm, allows the detection of people in images. But more discriminative methods based on powerful descriptors have also been developed. The visual information is locally extracted and collected. Finally the information is compared to a general model of people with a classification algorithm. Papageroriou et al were among the first to propose this pipeline [2]. They used wavelet descriptors, a sliding-window method to exhaustively scan the image and a SVM classifier. Many of current detectors are still based on this approach. Viola et al based their work on the work of Papageoriou et al [2]. They used integral images and a cascade classifier to speed up the computation of the Haar-like wavelet features and reach real-time performance for face detection [3]. The Histogram of Oriented Gradients (HOG) detector of Dalal and Triggs [4] is an efficient human detector using a variant of the very well-known and quite efficient SIFT descriptor [5]. Visual information is extracted using SIFT-like descriptors over a sliding-window. All the information is classified using a linear SVM classifier trained on images of people. The SIFT-like HOG descriptor still remains very competitive for object detection.

Fig. 1. First row: results obtained with our method, second row: results obtained with the ICF. Our method is less sensitive to changes of shape and angle.

Some detectors combine multiple descriptors, image features and/or information sources to increase the detection rate. Wojek et al showed that combining HOG, Haar-like descriptors, shaplets and the shape context outperform the HOG detector alone [6]. Dollar et al proposed a mix between Viola et al's detector and the HOG detector [7]. This detector computes simple rectangular features on integral images of different channels: L,U,V, gradient magnitude and six "HOG channels". The classification is performed using a fast soft-cascade classifier.

2) Multiple parts detection: Instead of considering the human body as one monolithic part, some detectors consider it as a set of parts. Felzsenzwald et al proposed a method to detect people by fragments and re-build human models by using a pictural structure reprensentation [8]. Each part of the human model is separately learned. An incorrect labelling of the fragments could decrease the performance of the detector [9]. That is why Felzenszalb et al introduced a detector using a new classifier: the latent SVM classifier [9]. With this classifier the most discriminative information is selected during the training to produce a more robust detection.

B. Existing work in human detection from a UAV

Detecting people is difficult and it becomes more difficult in a UAV context. Most human detectors focus on detecting upright people at nearby distances and from a more or less invariant viewpoint. The current two main applications of human detection is the security monitoring and the driving assistance. Until now little work has been done on detecting humans from a UAV. Unlike a pedestrian view, an UAV view is more complex to manage because the UAV undergoes pitching and rolling rotations. People are also on average further from the camera in this context.

Gaszczak et al proposed to use both thermal and visible imagery to better detect people and vehicles [10]. Features extracted on thermal and visible imagery are fused together to boost the confidence level of detection. The thermal camera is used for extracting Haar-like features while the optical camera is used for a contour shape analysis as a secondary confirmation to better confirm the detection. This method permits to detect upright people at a distance of about 160m using a fixed camera pitch rotation of minus 45 degrees and in real-time. This method does not seem flexible enough for detecting people closer to the UAV.

Rudol et al also use thermal and visible imagery but in a pipeline way [11]. They first identify high temperature regions from the thermal image and they reject the regions not fitting a specific ellipse. The corresponding regions are then analyzed in the visible spectrum using a relaxed Haar-like detector. Upright and seated people can be detected with this method. However, the thermal imagery can easily become very tricky to analyze with this method when the UAV is too close and the information becomes too noisy.

Reilly et al have a different approach [12]. They use people's shadows as a key clue to detect and localize people. But strong assumptions on weather conditions have to be made with this technique.

Andriluka et al evaluated various detection methods for detecting victims at nearby distances [13]. They showed part-based detectors are better suited for victim detection from a UAV because they natively take into account the articulation of the human body. The authors propose the use of complementary information using several detectors and inertial sensor data to obtain a better detection rate. However, part-based detection is a slow process [8] and this seems not suitable for detecting people too far from the camera.

C. Project context

This work is part of the French regional project SEARCH whose goal is to develop a system using several UAVs to rapidly find and rescue people in distress in the Somme estuary in northern France. The rising tide often trap walkers, bringing with it a risk of death from drowning. The current surveillance system employs helicopters to perform prevention and rescue missions. Unfortunately, this system makes use heavy equipment and is very expensive. Using a fleet of UAVs would be an interesting alternative for both the tax payers and the people in distress. This work is mainly about the development of a suitable human detection system for the SEARCH project.

The search and rescue context of the Somme estuary is the following: the environment is uncluttered, people may be close or far from the UAV (we consider a distance range of 15 to 50m), people may be walking or standing upright and the illumination may change over time and may not be the same everywhere. The use of UAVs adds some more constraints. The embedded camera of the UAV moves in a 3D world the detection system should thus be robust to the acquisition system's viewpoint. Moreover, a sufficient reactivness is required by the system in order for it to be usable in tortuous flights.

D. Content of the paper

The goal of this work is to design a detection system for addressing all of the following identified constraints: illumination robustness, detection of moving and stationary people, real-time compatibility, distance robustness and viewpoint robustness. The proposed approach is described in details in this paper.

Section II presents our approach for automatically detecting people from a UAV in open and natural environments. Section III concerns the conducted tests and the results obtained in a sequence taken in the Bay of Somme. This section also talks about the hardware. The results are discribed in this section.

II. Proposed approach

A specific processing pipeline was adopted in order to obtain better performance: it allows faster computation and a greater viewpoint robustness (Fig.2). At first visible and thermal images are extracted during the acquisition phase (1), a pre-processing phase follows immediately (2), the thermal images are analyzed to greatly reduce the search space (3), the spotted areas are fully analyzed in the visible light spectrum by a viewpoint robust detector to detect people (4) and to finish, the detected people are looped in the tracker (5).

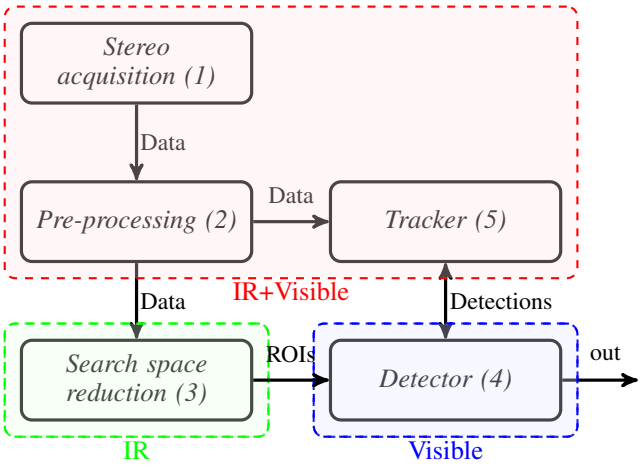

Fig. 2. General pipeline of the onboard detection system. At first visible and thermal images are obtained and registered (1), a pre-processing of the data is following this step (2), a search space reduction is performed in the thermal images (3), the corresponding visible ROIs are analyzed by our viewpoint robust human detector (4) and detections are looped in the tracker (5)

This paper focuses on stage (1), (2), (3) and (4) of the proposed pipeline. The acquisition system is composed of a

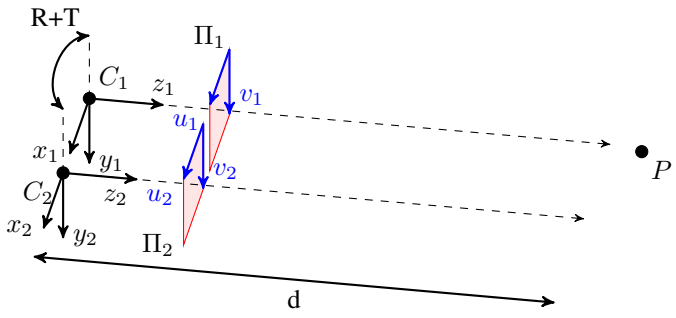

Fig. 3. Layout of the stereo acquisition system. C_1 and C_2 are respectively camera 1 and 2, Π_1 and Π_2 are the image planes of respectively camera 1 and 2. P is a person's location. R: rotation between the two cameras, T: translation between the two cameras. C1 is the thermal camera and C2 the visible light spectrum camera.

visible and a thermal camera. The two cameras are set one beside the other so that the two axis of the cameras are parallel and at the same height.

A. Stereo acquisition (1)

This step of the pipeline is dedicated to the registration of the visible and the thermal images: the pixels of both the cameras are spatially aligned. The goal of this step is to benefit from the discriminatory nature of thermal images and the information richness of visible images. Some knowledge about the intrinsic parameters of the cameras and the geometric relationship of the cameras are required to align the two modalities.

Because the objects of the scene are mostly far from the cameras and because the baseline is short (Fig.3) it has been decided that the impact of the baseline may be neglected (the closest people to detect are at about 15m from the UAV). Thus, for matching the pixels of the two modalities we use the infinite homography [14] (equation 1 and 2). K_1 and K_2 are the matrices of the intrinsic parameters of respectively camera 1 and 2. R is the rotational matrix between the two cameras and T is the translation between the two cameras.

$$H_\infty = K_2 \times R \times K_1^{-1} \tag{1}$$

$$\begin{pmatrix} u_2 \\ v_2 \\ w_2 \end{pmatrix} = H_\infty \times \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} \tag{2}$$

This technique is less time-consuming but also less accurate than the full rectification of the stereo pair with respect to a known distance. Besides, we found this assumption accurate enough for our case. The regions of interest extracted from the infrared images do not have to exactly match people locations but rather match coarse areas around people. Indeed, these coarse areas and their close neighborhood are finely analyzed afterwards. Tab.I shows the average pixel error between the two cameras (along u_1 or u_2) and with respect to the distance to the persons. For a better accuracy, these offsets can be roughly corrected if the distance between the camera and the persons can be known at run-time.

TABLE I. AVERAGE PIXEL ERROR ALONG THE u AXES WITH RESPECT TO THE DISTANCE FROM THE CAMERA

d (m)	10	20	30	40	50
Average pixel error (pixels)	5.998	3.04	2.03	1.53	1.22

B. Pre-processing (2)

We found it better to treat the thermal images using a saliency algorithm in order to enhance the visibility of the warmest areas. We noticed that it improves considerably the job of the search space reduction: thanks to that the reduction is cleaner and more robust. It has been decided to use the Achanta's saliency algorithm [15] which is fast and compute outstanding saliency maps. The algorithm has been slightly modified for the thermal case: only the L component of the Lab color space is considered.

C. Search space reduction (3)

The infrared map of a scene permits to better distinguish people from other objects of the surrounding environment. Indeed, human beings are natural Long Wave Infrared sources. But, it is not appropriate to only use this information to detect people because many objects of the scene can emit infrared on certain conditions (warm wall, stones, leafs of a tree, etc.). However, the search space can be greatly reduced for further treatments by simply analyzing these thermal images for extracting some regions of interest.

The special segmentation algorithm of San-Biagio and al [16] was implemented to reduce the search space. This algorithm is fast and spot relatively well warm areas of the scene that might correspond to human beings. It works recursively by extracting ROIs and refining them. It stopes when all ROIs do not change size after two iterations. The first ROI is the entire image. For each recursion the ROI is analyzed in the following way: the number of pixels with a brightness bigger than Thr_{step} is stored for each row. The same process is repeated for each column. The boundaries of the new ROIs within this ROI are found by thresholding the previously stored numbers for the rows and the columns. The thresholds for the rows ($Thr_{pixel,row}$) and for the columns ($Thr_{pixel,col}$) depend on the size of the image and on some constants [16]. Thr_{step} is refined after each recursion as showed in equation 3 and 4. $w1$, $w2$, $w3$ and Thr_{Sk} are parameters.

$$Thr_{step} = w1 \times Thr_{step-1} + (1 - w1) \times Thr_{ROI} \tag{3}$$

$$\begin{aligned} Thr_{ROI} = & w2 \times maxGrayLevel(ROI) + \\ & w3 \times meanGrayLevel(ROI) + \\ & (1 - w2 - w3) \times Thr_{Sk} \end{aligned} \tag{4}$$

D. Detection (4)

This is the most important step of the detection pipeline. The regions of interest extracted on the thermal images are analyzed in the visible light spectrum. This step is divided into two sub-steps: 1) random generation of candidate windows in and around the ROIs (Fig.4.4) and 2) analysis of each of the candidate windows using a supervised detector (Fig.4.5).

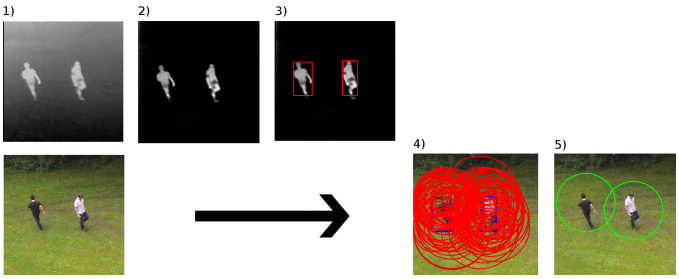

Fig. 4. 1) stereo-acquisition: input visible en infrared images, 2) pre-processing: computation of the saliency map of the infrared image, 3) search space reduction: segmentation of the "warmest areas" of the saliency map, 3) search space reduction and detector: transposition of the ROIs in the visible image and random generation of candidate windows around the ROIs, 4) detector: treatment of the candidate windows by our PRD and Non-Maximum Suppression treatment.

For the first sub-step, the candidate windows are generated according to the three following rules: the center of the window is randomly chosen within the ROI, the size of the window is randomly chosen between a min and a max scale (0.5 and 1.5 respectively) and the number of windows to generate for each ROI depends on the surface of the ROI. For the second sub-step, the candidate windows are analyzed by our viewpoint robust detector: the Pitch and Roll-trained Detector (PRD). The PRD is based on the Integral Channel Features (ICF) detector of Dollar et al [7]. Our detector requires a specific training phase with correctly labeled training data. The detection phase and the training phase are more detailed in the following sections.

1) Detection phase: During the detection phase the PRD analyzes the visual content of the circular window. Local visual features are extracted at different places of the circular window using integral images of ten different channels, which are: L, U, V, gradient magnitude and six "HOG channels". A visual feature is simply the sum of the pixels contained within a rectangle and associated to one of the ten channels. These visual features are extracted from a resized version of the circular window (resized with a radius of 64 pixels). A person is detected in a window if the visual features match the human model previously learned during the training phase.

A coarse-to-fine approach is adopted to speed the classification: the soft-cascade. For greater performance Dollar et al [17] propose to approximate the features between image scales to speed up the detection, they named this detector: The Fastest Pedestrian Detector in the West (FPDW) [17]. This same technique can also be used with our PRD for even greater speed performance.

2) Training phase: The training has been thought to deal with the change of human appearance occuring when the UAV's camera looks at people on the ground. The impact of the camera angles on the human appearance are the followings: the roll angle tends to rotate the shape of people and the pitch angle tends to change the shape of people (Fig.5 and 6). When the visual changes are too important, they cannot be managed by a detector with a classic design whose aim is to detect people in a pedestrian view. In order to overcome this problem a more general human model is trained by taking into account both the effect of the roll and the pitch during the training.

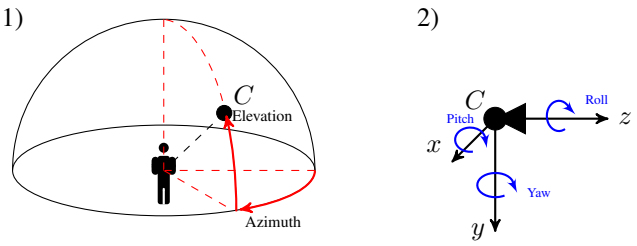

Fig. 5. C: optical center of the camera, 1) elevation and azimuth: position angles. 2) camera's angles.

Fig. 6. Examples of people images with different roll and pitch angles.

A specific training dataset was built: the GMVRT-v2 training dataset[1]. This dataset contains 3846 images of people taken at different pitch angles of the camera (Fig.7). This dataset is loaded in memory five times. Subsets of all the training images are rotated every 5deg so that all the images are spread from -90deg to 90deg. This step is necessary to minimize the effet of the roll on human apperance at training. A modified version of the Cluster Boosting Tree (CBT) algorithm [18] is used to learn the multiple aspects of the human class in relation to the viewpoint. This algorithm is based on AdaBoost, which is a very famous learning method in pattern recognition. AdaBoost [19] selects the best discriminant combination of visual features on positive and negative training images.

Fig. 7. Examples of GMVRT-v2 images.

Alg.1 presents our modified version of the CBT, see [18] for the original implementation. Our version is lighter and proved to be more efficient for our needs. Line 1 is the angular spreading of the data. Lines 7 and 8 (Alg.1) are typical AdaBoost procedures, except that S_+^k is a subset of S_+^0 for k>0. Line 9 is the condition to trigger the clusterization: the classification power $(h(t,k).Z)$ of the three latest trained weak-classifiers are compared to θ_Z. Lines 12 and 13 are re-training procedures of the previously trained weak-classifiers. We found it better to set θ_Z to 0.98 and to authorize as many clusterizations as possible.

Dollar and al recommend to use 30.000 random candidate features at training to build the weak-classifiers [7]. 45.000 candidate features were generated in order to keep a relatively similar density of candidates because the surface of the circular window of the PRD is 1.5 bigger than the surface of the classic detection window (64x128).

[1] http://mis.u-picardie.fr/~p-blondel/papers/data

input : GMVRT-v2 training dataset[1]
output: viewpoint robust classifier

several angular spreading of the data;
extracting all candidate features for all the data;
$c \leftarrow 1$;
for $k \leftarrow 0$ **to** c **do**
    reset default weights of S_+^k *and* S_-;
    for $t \leftarrow tinit(k)$ **to** T **do**
        build best weak-classifier h(k,t);
        update weights of S_+^k *and* S_-;
        if *h(k,t).Z* $> \theta_Z$ *and h(k,t-1).Z* $> \theta_Z$ *and h(k,t-2).Z* $> \theta_Z$ **then**
            split S_+^k *into* S_+^k *and* S_+^{c+1} ;
            h(c+1,t') = h(k,t'), $\forall t' \in [0, t]$;
            retrain weak-classifiers h(k,t'), $\forall t' \in [0, t]$ *with* S_+^k *and* S_-;
            retrain weak-classifiers h(c+1,t'), $\forall t' \in [0, t]$ *with* S_+^{c+1} *and* S_-;
            tinit(c+1) = t;
            c++;
        end
    end
end
$\forall k \in [0, c]$ *compute the soft-cascade for channel k*;

Algorithm 1: Our CBT implementation. c: number of clusterizations, k: index of the cluster (or channel), T: maximum number of weak-classifiers, h(k,t): weak-classifier number t of channel k, tinit(k): starting index for cluster k, θ_Z: clustering critera, S_+^k: cluster k of positive image, S_-: all the negative images.

III. Tests and Results

A. Hardware

We built our own stereoscopic system in order to conduct the tests. This system is composed of an infrared and a classic visible camera. The optical axes of the two cameras are at the same height and the baseline between the two cameras is exactly 5cm. The infrared camera is a Flir Tau2 running at 7 fps in VGA mode. The visible camera is a GoPro 3 HD (we took into account the distortion effect of the camera). Each camera is connected to a grabber. The two grabbers have the same specifications. The whole stereoscopic system weights about 470gr. The stereoscopic system is embedded in a Pelican quadrotor UAV (Fig.8).

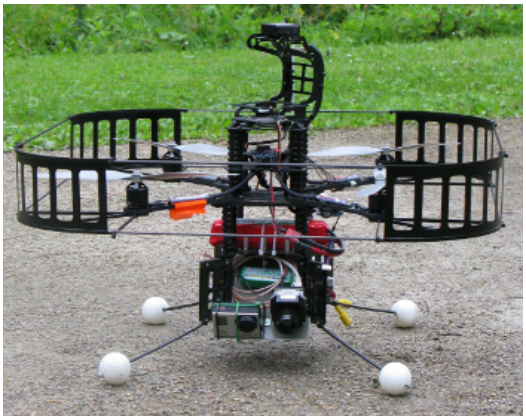

Fig. 8. The Pelican UAV with the embedded acquisition system.

Fig. 9. Examples of images from AerialTest1.

Fig. 10. Examples of images from AerialTest2.

B. Tests

In order to build the test datasets we flied the UAV in natural and open environments. Images of people were taken for different angular configurations of the acquisition system: the UAV flied over and around the persons. Two datasets were built: a first one containing only visible images taken from very challenging viewpoints of the camera (AerialTest1[1]), and a second dataset containing spatially and temporally syncronized visible and infrared images with some challenging viewpoints (AerialTest2[1]). In the second case, we flied the UAV at a lower altitude. AerialTest1 is used to demonstrate the robustness of our pipeline's core detector: the PRD. AerialTest2 is used to demonstrate the feasibility of our approach and to compare the performance of it to the performance of the classic approach (ICF).

1) Test 1: general performance of our core detector in a challenging dataset: The ICF detector fails to succeed in most cases (Fig.11): the miss-rate is constant and very high (pratically 1). Better performance are obtained with the Pitch and Roll-trained Detector (PRD). The PRD is much less sensitive to the challenging viewpoints of AerialTest1. The shape of the PRD's ROC curve is very similar to ROC curves of pedestrian detectors tested in pedestrian scenarios [20].

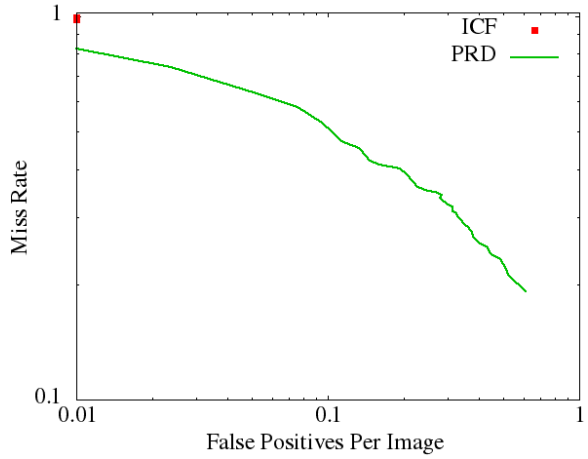

Fig. 11. ROC curves obtained by testing the ICF and the PRD detectors on AerialTest1. The ICF detector clearly fails on this dataset. On the contrary, the PRD is not affected by the challenging viewpoints of AerialTest1.

TABLE II. Reduction efficiency

Reduction Technique	Min (%)	Max (%)	Mean (%)	Sdev (%)
IR+Segmentation	0	0.6185	0.05246	0.0811
IR+Saliency+Segmentation	0.001	0.1009	0.0266	0.0203

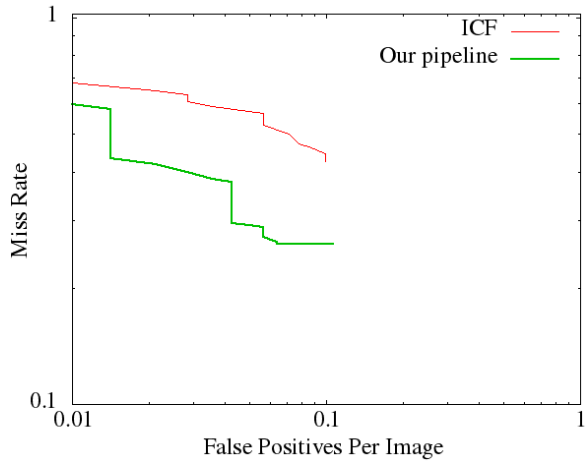

Fig. 12. ROC curves obtained by testing the ICF and our detection pipeline on AerialTest2. Better detection performance are obtained with our detection pipeline. However, the ICF still manages to detect the easiest cases of AerialTest2.

TABLE III. COMPUTATION TIME.

Detection Method	ICF	PRD	Our Pipeline
Computation time	T	1.75×T	1.05×T

2) Test 2: space reduction efficiency using infrared images: Infrared images are sometimes too bright, and therefore difficult to analyze: this leads to a non-optimal space reduction. We improved our space reduction by taking into account the saliency (Tab.II): the mean is about twice smaller. The reduction is more robust to changes, as well: the standard deviation (Sdev) is about four times smaller.

3) Test 3: detection performance of our detection pipeline: Our pipeline has better general performance than the ICF detector on AerialTest2 (Fig.12). However, the ICF detector manages to detect the easiest cases of AerialTest2, when view is close to the pedestrian view (contrary to the cases showed in Fig.1).

4) Test 4: computation time: The PRD alone is 1.75 times slower than the ICF (Tab.III). The computation time of our pipeline is equivalent to the computation time of the ICF. This can be further improved by approximating features between octaves [17] and/or by performing the low-level pixel treatments (such as saliency map computation) using FPGAs.

IV. CONCLUSION

In this paper, we proposed a new detection pipeline for viewpoint robust and fast human detection in search and rescue operations. Our approach combines the use of a viewpoint robust detector (PRD) and an accurate search space reduction technique. We showed that our detection pipeline outperforms traditional human detection approaches such as the Integral Channel Features (ICF) detector: one of the top-performing human detectors in pedestrian contexts.

The main contributions of this work are: 1) a search space reduction technique using syncronized infrared images optimized to lighten the work of the core detector which treats the corresponding areas but in the visible light spectrum, and 2) a viewpoint robust detector (PRD) to detect people whatever the roll and pitch angles of the acquisition system in a UAV context.

The next objective is to improve the general performance of the core detector by also taking into account the shape information available in IR images. It is also wished to improve the speed of our detection pipeline in order to improve its reactiveness even further. Test are planned to conduct tests with different types of UAVs in order to show the flexibility of our approach.

REFERENCES

[1] D. Gavrila and J. Giebel, "Shape-based pedestrian detection and tracking," *Intelligent Vehicle Symposium*, 2002.

[2] C. Papageoriou and T. Poggio, "A Trainable System for Object Detection," *International Journal of Computer Vision*, 2000.

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

[4] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Conference on Computer Vision and Pattern Recognition*, 2005.

[5] D. Lowe, "Object recognition from local scale-invariant features," in *International Conference on Computer Vision - Volume 2*, 1999.

[6] C. Wojek and B. Schiele, "A Performance Evaluation of Single and Multi-feature People Detection," *Pattern Recognition, 30th DAGM Symposium*, 2008.

[7] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral Channel Features," in *Proceedings of the British Machine Vision Conference*, 2009.

[8] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial Structures for Object Recognition," *International Journal of Computer Vision*, 2005.

[9] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models." *Transactions on pattern analysis and machine intelligence*, 2010.

[10] A. Gszczak, T. P. Breckon, and J. Han, "Real-time People and Vehicle Detection from UAV Imagery," in *Intelligent Robots and Computer Vision*, 2011.

[11] P. Rudol and P. Doherty, "Human Body Detection and Geolocalization for UAV Search and Rescue Missions Using Color and Thermal Imagery." in *Aerospace Conference*, 2008.

[12] V. Reilly, B. Solmaz, and M. Shah, "Geometric constraints for human detection in aerial imagery," in *European conference on Computer vision: Part VI*, 2010.

[13] M. Andriluka, P. Schnitzspan, J. Meyer, S. Kohlbrecher, K. Petersen, O. von Stryk, S. Roth, and B. Schiele, "Vision based victim detection from unmanned aerial vehicles," in *Conference on Intelligent Robots and Systems (IROS)*, 2010.

[14] S. J. Krotosky and M. M. Trivedi, "Mutual information based registration of multimodal stereo videos for person tracking," *Computer Vision and Image Understanding*, pp. 270–287, 2007.

[15] R. Achanta and S. Sabine, "Saliency Detection Using Maximum Symmetric Surround," in *Internation Conference on Image Processing (ICIP)*, 2010.

[16] M. San-Biagio, M. Crocco, and M. Cristani, "Recursive segmentation based on higher order statistics in thermal imaging pedestrian detection," in *5th International Symposium on Communications, Control and Signal Processing*, 2012, pp. 2–4.

[17] P. Dollár, B. S., and P. Perona, "The Fastest Pedestrian Detector in the West," in *British Machine Vision Conference*, 2010.

[18] B. Wu and R. Nevatia, "Cluster Boosted Tree Classifier for Multi-View, Multi-Pose Object Detection," in *Computer Vision and Pattern Recognition*, 2007.

[19] R. E.Schapire and S. Yoram, "Improved Boosting Algorithms Using Confidence-rated Predictions," *Machine Learning*, 1999.

[20] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *Conference on Computer Vision and Pattern Recognition*, 2009.