A procedure for detection of humans from long distance images

Hrvoje Turić [1], Vladan Papić [1], Hrvoje Dujmić [2]
[1] Faculty of Natural Sciences, Mathematics and Education, University of Split, Teslina 12, 21000 Split, Croatia
[2] Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, R. Boskovića bb, 21000 Split, Croatia
E-mail: *hrvoje.turic@pmfst.hr*

Abstract - *The article presents a new procedure for detection of humans from images of the mainly mountain terrain. Our approach uses the hue and saturation components of the image as the input to the segmentation module that uses the mean shift method. The clusters obtained as the output of this stage have been processed by the decision-making module in order to find the regions of the image with the significant possibility of presenting human or some other suspicious artifact. Real world images are used for the evaluation of the procedure.*

Keywords – *Search and rescue, mean shift, image segmentation*

1. INTRODUCTION

Search and rescue missions of humans are, unfortunately, almost everyday reality. These missions are including search of the area from the close range as well as surveillance of the landscape and target detection from the distance. Large areas of generally unfamiliar terrain are needed to be covered in order to find the lost, hurt or persons that are in some kind of danger. Such missions are demanding large and diverse task forces. Numerous personnel, various technical support and therefore, significant financial resources are needed for a successful accomplishment of the task. Additional support and, sometimes, the only possible option is autonomous inspection of the desired area using the robotic vehicles or approach that includes some kind of artificial intelligence. Different kinds of sensors such as shape, color, motion, IR signals, temperature, voice signals, CO_2 emissions sensors etc. [1] are used for the detection.

Numerous scientific research articles dealing with the on-ground search of the humans and their rescue can be found but their focus is mainly on various image processing and computer vision methods for detection of human parts, robot localisation and sensor fusion [2][3].

Number of available sensors and the resolution is significantly lower for the long distance, primarily aerial surveillance, and it makes the above mentioned approach generally inapplicable. Aerial surveillance systems that are using IR cameras have some significant gains over conventional photographic methods [4] but they also have some limitations. While it has the advantage for the night surveillance, problems occur during the day because of the temperature raise and inability to distinguish humans and animals from other warm objects. Additionally, night missions with helicopters and other flying vehicles are dangerous in an unfamiliar terrain.

As a necessary component of the efficient and complete search and rescue system, a sub-system that could perform successful surveillance and target detection based on real-time image processing of the aerial day-images is needed. Surprisingly, there are no numerous articles or literature about particular area of interest. Most of the research articles are related with the Unmanned Aerial Vehicles (UAV), their control and terrain mapping [5]. Also, an important issue that is dealt with is, tracking of the moving object because many of the applications are oriented towards traffic surveillance. In our case, motion information obtained from the acquired sequences of images is expected to be non relevant because the individuals that are searched for are mainly non-moving. That means that focus of our research fill be on image processing of static images. Manolakis et al presented a tutorial review of the state of the art in target detection algorithms for hyperspectral imaging applications [6], while an image processing system for the detection of the rescue target in the marine casualty is presented by Sumimoto and Kuramoto [7].

In this paper, a feature space for the image segmentation for the observation will be defined. After the segmentation (and image preprocessing), on the basis of chosen rules, particular regions of the image with high susceptibility of representing humans of will be chosen. Further on, the results of the proposed procedure will be given on real-world images and finally, in the concluding part of the article, ideas for the future work will be explained.

2. METHODS

Proposed procedure consists of three main modules: preprocessing, segmentation and decision-making module (Fig.1.).

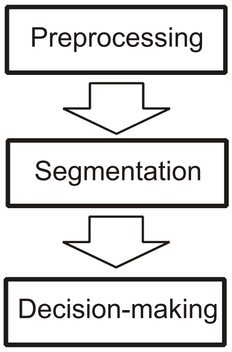

Fig.1. Main modules of the procedure

Preprocessing module has two steps: 1. Translating of the image color format to HSV color format; 2. Filtering of hue (H) and saturation (S) components with median filter. Filtered H and S components are used as the input to the segmentation module. For the segmentation, the mean shift method is used [8].

The mean shift algorithm is a nonparametric clustering technique which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. Given *n* data points x_i, *i* = 1, ..., *n* on a *d*-dimensional space R^d, the multivariate kernel density estimate obtained with kernel *K(x)* and window radius *h* (bandwidth) is

$$f(x)=\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right). \qquad (1)$$

For radially symmetric kernels, it suffices to define the profile of the kernel *k(x)* satisfying

$$K(x)=c_{k,d}\,k\left(\|x\|^2\right) \qquad (2)$$

where $c_{k,d}$ is a normalization constant which assures *K(x)* integrates to 1. The modes of the density function are located at the zeros of the gradient function $\nabla f(x)=0$. The gradient of the density estimator (1) is

$$\nabla f(x)=\frac{2c_{k,d}}{nh^{d+2}}\sum_{i=1}^{n}(x_i-x)\,g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)$$
$$=\frac{2c_{k,d}}{nh^{d+2}}\left[\sum_{i=1}^{n}g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)\right]\left[\frac{\sum_{i=1}^{n}x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n}g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}-x\right]. \qquad (3)$$

where $g(s)=-k'(s)$. The first term is proportional to the density estimate at *x* computed with kernel $G(x)=c_{g,d}\,g\left(\|x\|^2\right)$ and the second term

$$m_h(x)=\frac{\sum_{i=1}^{n}x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n}g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}-x \qquad (4)$$

is the mean shift. The mean shift vector always points toward the direction of the maximum increase in the density. The mean shift procedure, obtained by successive

- computation of the mean shift vector $m_h(x^t)$,
- translation of the window $x^{t+1}=x^t+m_h(x^t)$

is guaranteed to converge to a point where the gradient of density function is zero.

The mean shift clustering algorithm is a practical application of the mode finding procedure:

- starting on the data points, run mean shift procedure to find the stationary points of the density function,
- prune these points by retaining only the local maxima.

The set of all locations that converge to the same mode defines the basin of attraction of that mode. The points which are in the same basin of attraction is associated with the same cluster.

Segmentation module provides certain number of clusters that have to be processed by the last, decision-making module. Decision making module has five phases as presented in Fig. 2.

The first phase erases any cluster which has more than X_1 pixels.

$$X_1=\frac{image\_width}{a}\times\frac{image\_height}{b} \qquad (5)$$

where *a* and *b* are variables that are depending on the estimated distance from camera to the observed surface. Presupposition is that if the candidate region that could present a person has more than X_1 pixels, it means that the actual person stands too close to the camera and the search is trivial. This presumption efficiently eliminates the big areas from the image.

Second phase erases areas inside the cluster containing too few pixels (less than X_2). That way the noise presented by some scattered pixels left after median filtering is being eliminated.

The third phase increases the area inside the cluster with mask 4x4. This way nearby areas are being merged.

The fourth phase ignores the whole cluster if the cluster has more than three spatially separated areas. Assumption is that there won't be more than three persons wearing the same color on the image.

All the remaining regions that were not eliminated by the previous four phases are proposed in the fifth and the last level as possible solutions.

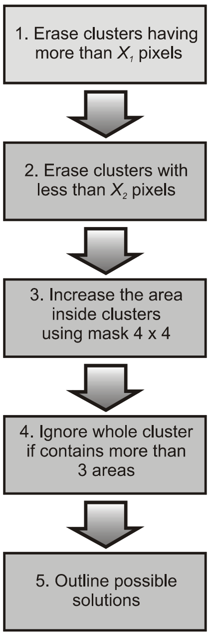

Fig.2. Five phases of the decision-making module

3. RESULTS

The system has been evaluated with 34 real-world images. Images were taken from the various terrains, different times of year and with varying distances from the observed surface. On 28 images, there were one or more persons. Also, 6 images contained no person. Although the original resolution was higher, in order to simulate resolution from the standard video camera, and also to speed up the processing, all the images were transformed to the VGA resolution (640 x 480).

Procedure was coded using the Matlab programming language. Also, functions of the Matlab Image Processing Toolbox were used.

Table 1. Detection results

	Correct detection	Partially correct detection	Incorrect detection
Number of images	20	10	4
Percentage	59%	29%	12%

Results obtained after processing of the images are presented in Table 1. Correct detection means that all the persons on the image are detected and, also, no person is detected when the original image contains no humans. If the procedure detects some, but not all persons in the image with multiple persons, result is presented as partially correct detection. Incorrect detection includes two possible cases: 1. person(s) detected even if there is actually no persons on the image; 2. person(s) are not detected although they are present on the original image.

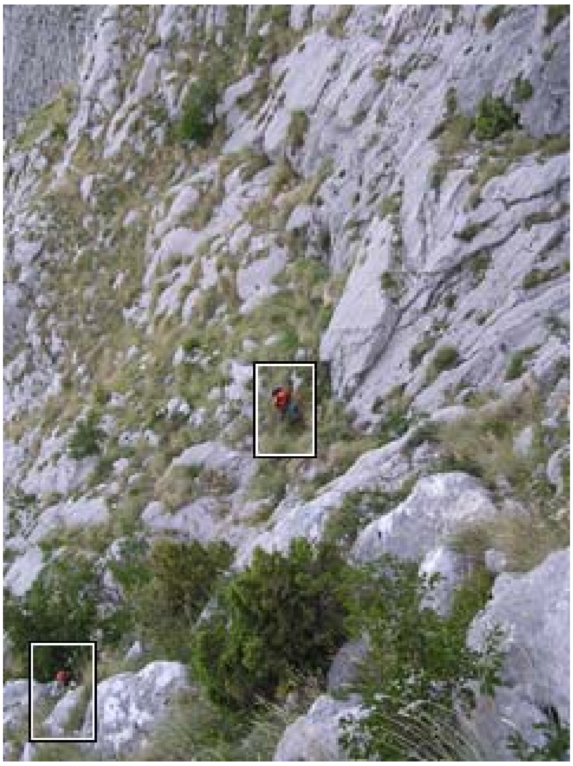

Fig 3. Correct detection example (2/2 persons)

Fig 4. Segments of the image presented in Fig. 3. Segmentation is done using H and S component of the image (for presentation in this figure, V = 0.5).

Fig 5. Example of the partial detection (3/5 persons) Two persons are undetected because of the merging of their clusters with the clusters of the neighbourhood and low saturation clothes

Examples of the output images obtained after processing of the test images are presented in Fig 3, Fig. 4 and Fig. 5.

Based on the results presented in the table, we may say that the final result was successful in 88% (59% + 29%) of the test images because our procedure correctly found persons in the image (although for some cases all the persons were not detected).

4. CONCLUSION

Procedure proposed in this paper showed promising results although it is quite simple. This inspires for the future work and introduction of possible improvements regarding robustness, speed and scope of the application.

Although mean shift method is known as fast and reliable segmentation method, we plan to include some new segmentation methods into the procedure. Selection of the active segmentation method for the procedure should be problem-specific and should be done according to the image analysis in the preprocessing stage. Also, limited set of decision-making rules would be expanded with new knowledge and rules. Our goal is to make this procedure adaptive and that should result with higher correct detection rate and minimisation of the false alarms.

A complete, real - time system for detection of humans and other targets which includes both hardware (unmaned aerial vehicle with camera and computer) and software (detection and accurate localisation) should be the final goal of this research. This system could provide cruical help in search and rescue missions when the speed is critical factor and available forces are dealing with the unfamiliar and harsh terrain.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Science and Technology of the Republic Croatia under projects: Computer Vision in Identification of Sport Activities Kinematics (177-0232006-1662) and ICT systems and services based on integration of information (023-0231924-1661).

REFERENCES

[1] S. Bahadori, L. Iocchi. “Human Body Detection in the RoboCup Rescue Scenario Rescue” Workshop. *RoboCup competitions* Padua, Italy, 2003.

[2] I.R. Nourbakhsh, K. Scara, M. Koes, M. Yong. “Human-Robot Teaming for Search and Rescue”, *Pervasive Computing*, 2005, pp. 72-78.

[3] A. Birk, S. Carpin. "Rescue robotics: a crucial milestone on the road to autonomous systems". *Advanced Robotics,* 20(5), 2006, pp. 595-605.

[4] H. Yalcin, R.Collins, M. Hebert. “Background estimation under rapid gain change in thermal imagery”, *Computer Vision and Image Understanding,* Volume 106, Issues 2-3, Special issue on Advances in Vision Algorithms and Systems beyond the Visible Spectrum, 2007, pp. 148-161,

[5] Ollero, L. Merino. "Control and perception techniques for aerial robotics". *Annual Reviews in Control,* 28, Elsevier, 2004, pp. 167-178.

[6] D. Manolakis, D. Marden, G A. Shaw, "Hyperspectral Image Processing for Automatic Target Detection Applications", *Lincoln Laboratory Journal*, Volume 14, Number1, 2003, pp. 79-116.

[7] T. Sumimoto,K. Kuramoto, S. Okada, H. Miyauchi, M. Imade, H. Yamamoto, Y. Arvelyna, “Detection of a particular object from environmental images under various conditions”, *Proceedings of the International Symposium on Industrial Electronics*, ISIE, IEEE, 2000., vol.2., pp. 590-595.

[8] D. Comaniciu, P. Meer. “Mean shift: A robust approach toward feature space analysis”, *IEEE Trans. Pattern Anal. Machine Intell*, 24, 2002, pp.603–619.