# A new method for detection of artificial objects and materials from long distance environmental images

H. Dujmić, V. Papić and H. Turić

**Abstract—**The article presents a new method for detection of artificial objects and materials from images of the environmental (non-urban) terrain. Our approach uses the hue and saturation (or Cb and Cr) 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 representing human. Although this method will detect various non-natural objects, it is primarily intended and optimized for detection of humans; i.e. for search and rescue purposes in non-urban terrain where, in normal circumstances, non-natural objects shouldn't be present. Real world images are used for the evaluation of the method.

***Keywords***— Landscape surveillance, mean shift algorithm, image segmentation, target detection.

## I. 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, helicopters or Unmanned Aerial Vehicles (UAV) 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 detection of humans. But, almost exclusively, scientific research articles are dealing with the on-ground search of the humans and their focus is mainly on various image processing and computer vision methods needed for detection of human parts, robot localization methods 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. Clothes or some other part that could be correlated with the missing person can be represented with only a small number of image pixels which, along with possible occlusion, makes shape-oriented techniques useless. The modern human (and general target) long distance detection is dependent upon several types of images data including photographic (optical) data, infrared data and radar data. Aerial surveillance systems that are using infrared 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 from other warm objects. Additionally, night missions with helicopters and other flying vehicles are dangerous in an unfamiliar terrain. Radar images are used mainly for military application for target detection (mostly non-human and moving objects such as thanks, vehicles, etc.).

As a necessary component of the efficient and complete search and rescue system, a sub-system that could perform successful surveillance and human detection based on real-time optical image processing of the aerial (or generally long distance) day-images is needed. At this moment searches for humans are usually conducted by naked eye, and where closer inspection of a particular area is needed, binocular is used as an aid. Sometimes helicopters or UAV are used for searching. But, combination of flying height and speed results in relatively low probability of detections. In some rarely occasions, pictures are taken using helicopters or UAV in order to be examined later by humans. This approach is extremely demanding in term of man hours needed and can not be done in real time.

Surprisingly, it is hard to find an article or system dealing

H. Dujmić is with the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, R. Boskovića bb, 21000 Split, Croatia; [3] Croatian Mountain Rescue Service, Sibenska 41, Split, Croatia

V. Papić is with Faculty of Science, University of Split, Teslina 12, 21000 Split, Croatia

H. Turić is with the Faculty of Science, University of Split, Teslina 12, 21000 Split, Croatia

with human detection from long-distance images. Most of the research articles related with the Unmanned Aerial Vehicles (UAV) deals with their control and terrain mapping [5]. Long distance images (including satellite images) are also used for soil type detection, road and building detection [10].

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 optical images is expected to be non relevant because the individuals that are searched for are mainly non-moving. That means that the focus of our research should 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 (boats) in the marine accidents is presented by Sumimoto and Kuramoto [7].

In this paper, a color model used as a basis for the image segmentation for the observation will be defined. After the segmentation (and image preprocessing), particular regions of the image with the high susceptibility of representing humans will be chosen.

Rest of the paper is organized as follows. Section II presents proposed method. Results are presented in section III. Section IV gives some ideas regarding future work. Conclusions are made in section V.

## II. Proposed method

Proposed method consists of three main modules: preprocessing, segmentation and recognition module (Fig.1.).

### *A. Preprocessing*

Preprocessing module has two steps: 1. Translating of the image color format to the chosen color model; 2. Filtering of the color components with median filter.

We have tested two color models: HSV and YCbCr. In the case of HSV color model, hue (H) and saturation (S) components were filtered with median filter. Filtered H and S components are then used as the input to the segmentation module. In the case of image presentation with YCbCr model, Cb and Cr (color differences) components were filtered.

### *B. Segmentation*

After segmentation with the k-means method [8] and quite disappointing segmentation results, the mean shift method has been chosen [9]. 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

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

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

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

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

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

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

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

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 $m_h\left(x^t\right)$,
- translation of the window $x^{t+1}=x^t+m_h\left(x^t\right)$

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

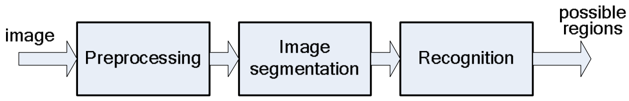

Fig.1. Main modules of the method

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.

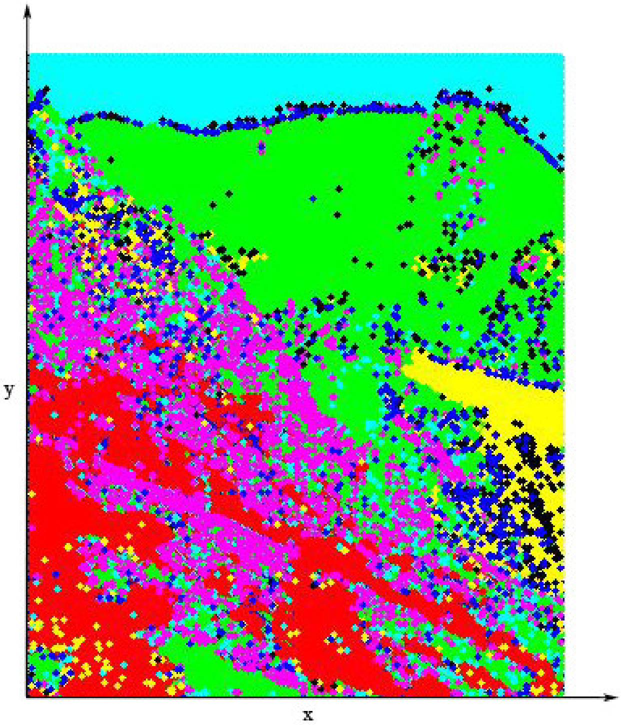

Fig.2. Typical output after segmentation. 18 clusters are found in the input image (Fig. 4) after mean shift segmentation using hue and saturation.

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, recognition module.

### C. Recognition

Recognition module has six phases (Fig. 3). Phases 1 to 4 reject all the regions that are not likely to present humans. The first phase erases clusters 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 eliminates clusters containing too few pixels, i.e. less than $X_2$. $X_2$ is calculated in the same way as $X_1$ (in practice $X_2$ is set between 3 and 8). That way the noise presented by some scattered pixels left after median filtering is being eliminated. The third phase merges spatially close pixels belonging to the same cluster. This is done using the dilatation with the 4x4 mask. The fourth phase rejects clusters having more then designated number of spatially separated areas. Assumption is that there won't be more than few persons wearing the clothes with the same characteristics in chosen color space. The maximal number of spatially separated regions is set to 3 but it can be adjusted according to the specific situation.

All the remaining regions that were not eliminated by the previous four phases are forwarded to the next two phases that are optional. The fifth phase assigns scores to the remaining region in the image according to the available data and situation knowledge. Possible information for the region scoring can be colors of the missing person clothing, usual terrain colors, etc. It reduces the number of wrongly detected regions. To be as general as possible, the results that are presented in Section III don't use any additional knowledge such as cloth colors. It means that candidate regions and regions with detected persons (Fig. 3) are identical.

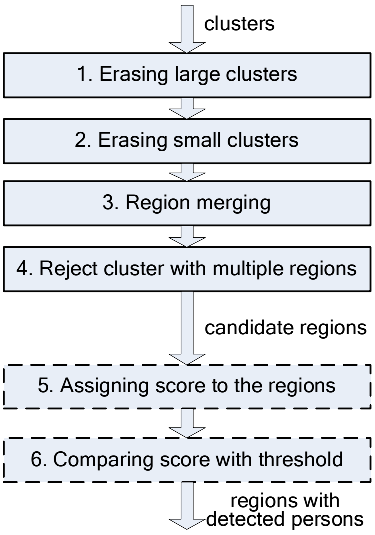
clusters
1. Erasing large clusters
2. Erasing small clusters
3. Region merging
4. Reject cluster with multiple regions
candidate regions
5. Assigning score to the regions
6. Comparing score with threshold
regions with detected persons

Fig.3. Recognition module

## III. Results

The system has been evaluated with 50 real-world images. Images were taken from the various terrains, different times of year and with varying distances from the observed surface. On 36 images, there were one or more persons (altogether 49 persons). Also, 14 images contained no person. Although the original resolution was higher, in order to speed up the processing, all the images were transformed to the lower resolution (320 x 240).

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

TABLE 1
RECALL AND PRECISION RATES OF THE PROPOSED METHOD

| | Correct | Missed | False Positive | Recall | Precision |
|---|---|---|---|---|---|
| HSV | 38 | 11 | 30 | 77.6% | 55.9% |
| YCbCr | 28 | 21 | 14 | 57.1% | 66.7% |

Results obtained after processing of the images are presented in Table1. Highly useful way to measure and compare effectiveness of different algorithms is to compute their Recall (R) and Precision (P):

$$R = \frac{Correct}{Correct + Missed} \times 100 \qquad (6)$$

$$P = \frac{Correct}{Correct + FalsePosit\ ive} \times 100 \qquad (7)$$

In Table 1 Recall and Precision rates are given for HSV and YCbCr color model. Correct is the number of correctly recognized persons. Missed is the number of persons not recognized by our method. FalsePositive is the number of false alarms, i.e. the number of wrongly detected regions.

The very important observation regarding false positive regions is that almost all false positive regions are on the images that already contain person. Only one image without person (in HSV as well as in YCbCr color model) contains false positive region. It means that, regarding target application (i.e. human recognition), false alarms are not problematic in practice. In very rare occasions, operator will be in situation to have an alarm in the images without person. In this context, it also means that Recall rate is a more important measure than Precision rate.

Also, there is only one image without person (out of 14) that gives false alarm (i.e. method has detected person where there is no one). It means that this method (especially using HSV model) can be efficiently used in real situations.

To our best knowledge we couldn't find any similar application in the literature so we cannot provide any comparison with other results.

Considering this, we can say that implementation using HSV color model (R=77.6%) is better than using YCbCr model (R=57.1%) although YCbCr model has better performance regarding Precision.

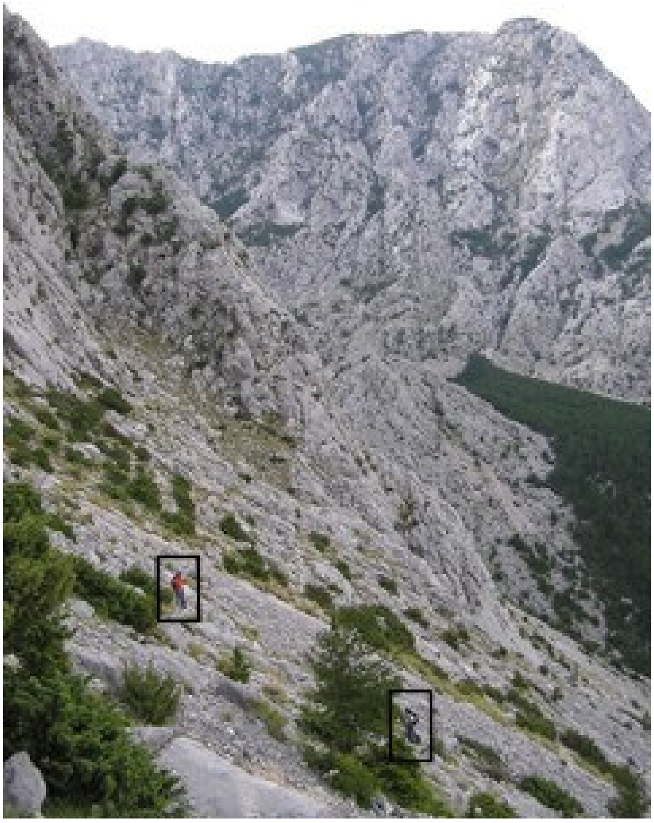

Fig 4. Correct detection example (2/2 persons)

To our best knowledge we couldn't find any similar application in the literature so we cannot provide any comparison with other results.

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

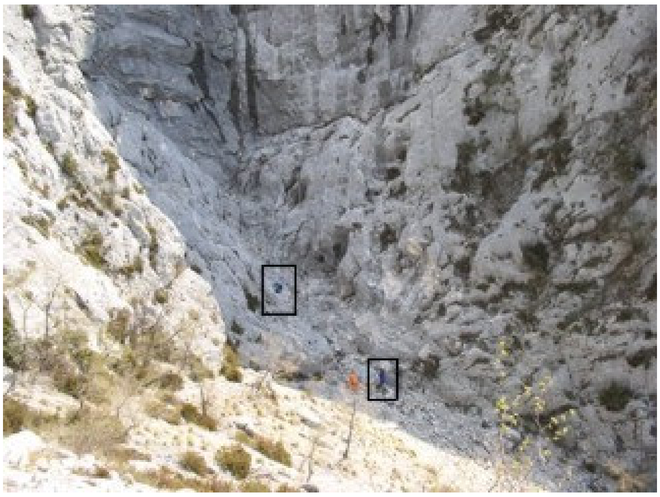

Fig 5. Example of the partial detection (2/3 persons)
One person is undetected because of the merging of its cluster with the clusters of the neighbourhood.

## IV. Future work

The method proposed in this paper showed promising results and inspires us for the future work and introduction of possible improvements regarding robustness, speed and scope of the application.

We will continue with the research of different clustering methods and possible improvements of the best one so far, the mean shift. Also, analysis of the segmentation results for two different color models presented in this paper (HSV and YCbCr) indicates that their combination could increase number of detected regions.

Our long term objective is to develop 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). Real-time acquisition and processing of the images is expected to be available after the code optimisation and translation of code and processing to the graphical processors instead of using the CPU. Calculated critical time available for processing of one image in the case of aerial surveillance using the MI-8 helicopter (minimal safe flight speed is around 28 m/s) is 2 seconds which is achievable using the present hardware and improvements regarding GPU processing power introduction.

This system could provide crucial help in search and rescue missions when the speed is critical factor and available forces are dealing with the unfamiliar and harsh terrain.

## V. Conclusion

In this paper we have presented a new method for detection of artificial objects and materials, especially humans from images of the environmental (non-urban) terrain. Our method segments images according to the two components of the color model (hue/saturation or Cb/Cr). The clusters obtained as the results of mean shift segmentation have been processed by the decision-making module in order to find the regions of the image with the significant possibility of presenting human.

The primarily intention of the proposed method is for the search and rescue missions in the non-urban terrain. Proposed method has been tested on the 50 real-word images. The implementation using HSV color model archives Recall rates of 77.6% and for the implementation using YCbCr model recall rate is 57.1%.

Also, there is only one image without person (out of 14) that gives false alarm (i.e. method has detected person where there is no one). It means that this method (especially using HSV model) can be efficiently used in real situations.

To our best knowledge we couldn't find any similar application in the literature so we cannot provide any comparison with other results.

## Acknowledgment

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 and L. Iocchi, “Human body detection in the RoboCup rescue scenario rescue” *Workshop in RoboCup competitions*, Padua, Italy, 2003.

[2] I. R. Nourbakhsh, K. Scara, M. Koes and M. Yong, “Human-robot teaming for search and rescue”, *Pervasive Computing*, pp. 72-78, 2005.

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

[4] H. Yalcin, R. Collins and 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, pp. 148-161, 2007.

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

[6] D. Manolakis, D. Marden and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications", *Lincoln Laboratory Journal*, Volume 14, Number1, pp. 79-116, 2003.

[7] T. Sumimoto *et al.*, “Detection of a particular object from environmental images under various conditions”, *Proceedings of the International Symposium on Industrial Electronics*, ISIE, IEEE, vol. 2., pp. 590-595, 2000.

[8] J. Peña, J. Lozano and P. Larrsñaga, “An empirical comparison of four Initialization methods for the k-means algorithm,” *Pattern Recognition Letters*, vol. 20, pp. 1027-1040, 1999.

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

[10] C. Ünsalan and K. L. Boyer, “A system to detect houses and residental street in multispectral satellite images”, *Computer Vision and Image Understanding*, 98, 423-461, 2005.