Two-stage segmentation for detection of suspicious objects in aerial and long-range surveillance applications

VLADAN PAPIĆ[1], HRVOJE TURIĆ[1], HRVOJE DUJMIĆ[2]
Department of polytechnics[1], Department of electronics[2]
University of Split
Teslina 12
CROATIA
vpapic@pmfst.hr http://www.pmfst.hr/~vpapic

Abstract: - This paper presents a novel two-stage data segmentation approach for detection of artificial materials and objects in non-urban terrain. High resolution image of the unknown terrain taken with the digital camera from relatively large distance is divided into smaller sub-images for further processing. Each sub-image is segmented and information about obtained cluster centers is transferred to the next stage. Second segmentation stage uses information about cluster centers from all sub-images and applies the same clustering method to this data set. Finally, decision-making module evaluates all potential candidate segments and eventually proposes segments that have high possibility of presenting artificial material or object in the input image.

Key-Words: - Image segmentation, mean shift, object detection, aerial images

1 Introduction

Surveillance of various terrains in order to find some object of interest is a task that can be associated with various civil and military activities. Possible applications encompass so called search and rescue missions that are aimed towards finding lost, hurt or persons that are in some kind of danger. Large areas of generally unfamiliar terrain must be thoroughly inspected. Such missions can last for several days and they are demanding large and diverse task forces, various technical support. Therefore, significant financial resources are needed.

As an additional support and, sometimes, the only possible option is introduction of autonomous inspection of the area of interest using the robotic vehicles or approach that includes some kind of artificial intelligence. Articles presenting algorithms and methods for rescue missions are mainly focused on on-ground rescue that includes processing of images taken from the short range, sensor fusion and specificity of this kind of applications such as detection of shapes corresponding to human parts etc. [1], [2]. Object shapes are not likely to be helpful in detection of objects from the images taken from the long distances (due to occlusions, resolution limitations,…) such as aerial photos taken from the helicopter or airplane. Aerial surveillance systems that are using IR cameras have some significant gains over conventional photographic methods [3] but they have limitations, also. While this type of systems has the advantage for the night surveillance, problems occur during the daylight 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.

It can be deduced that an image processing sub-system is a necessary component of an efficient and complete search and rescue system. 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 [4]. Long distance images (including satellite images) are also used for soil type detection, road and building detection [9]. 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 will 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 [5], while Sumimoto and Kuramoto depicted an image processing system for the detection of the rescue target in the marine casualty [6].

In the preceding research, various image segmentation algorithms were investigated [10] [12] and the mean shift algorithm has been chosen as the algorithm that demonstrated the best results regarding quality of segmentation, speed and stability regarding segmentation parameters. Also, segmentation in HSI (hue-saturation-intensity) color space was investigated [11] because of expected benefits related to high saturation values of artificial objects and materials.

However, comparison of segmentation results obtained for HSI and YCbCr (luminance-blue difference chroma-blue difference chroma) color space showed that processing in YCbCr color space produces slightly better results regarding precision [11].

The biggest issue that was left unanswered is the resolution quality and processing time. Before further discussion, the processing modes for this type of applications should be explained. Possible processing modes are real time image processing of video stream acquired by video camera and off-line processing of digital camera images. It should be mentioned that, for this type of applications, standard video cameras working in interlaced mode are not applicable because of the blurring effect in images due to high speed of an airplane or helicopter with the camera relative to the ground. Higher resolution of the images acquired by digital camera provides significant benefits for positive detection rate of the relatively small objects and allows larger area (terrain) to be captured in single image (photos taken from higher altitude). However, larger images are requiring longer processing time so our idea was to propose a procedure that could be implemented for parallel processing of several CPU-s or exploit the capacities of modern graphical processor units (GPU-s) that have parallel "many-core" architecture, each core capable of running thousands of threads simultaneously.

Rest of the paper is organized as follows. Section 2 presents used algorithms and methods as well as the proposed two-stage segmentation approach. Results are presented in section 3. Conclusions are made in section 4 and acknowledgements are given in section 5.

2 Methods

2.1 Image preprocessing

Choice of the color space used in a particular computer vision application is not a trivial task because different color spaces have different advantages and disadvantages [7]. It can be stated that traditional RGB color space is not convenient for this kind of applications due to the high correlation between color components. For computer vision applications, HSI (HSV) and YCbCr color spaces are, generally, the better choice.

Preprocessing module has two steps. In the first step, translating of the image color format to YCbCr color format is done. After that, blue-difference chroma component (Cb) and red-difference chroma component (Cr) components are filtered using the median filter. Filtered Cb and Cr components are used as the input to the following modules.

2.2 Mean shift algorithm

The mean shift algorithm [8] 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

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

where $g(s) = -k'(s)$. The first term is proportional to the density estimate at x computed with kernel $G(x) = c_{g,d} \cdot g\left(\|x\|^2\right)$ 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 $m_h\left(x^t\right)$ and 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.

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.

2.3 Decision-making module

Decision making module has several steps. First, large clusters are erased (larger than defined threshold value). Threshold value depends on the estimated distance from camera to the observed surface. Presupposition is that if the candidate region that could present a suspicious object (person) has more than estimated number of 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 step of the decision-making module erases areas inside the cluster containing too few pixels (less than estimated threshold). That way the noise presented by some scattered pixels left after median filtering is being eliminated. After this step, obtained image is dilated in order to merge similar nearby segments.

Next step excludes as candidate segments those segments that belong to the cluster with more than three spatially separated areas. Assumption is that there won't be more than three suspicious objects in the image having the same color.

2.4 Procedure overview

The whole procedure can be presented as a modular, two-stage approach (Fig.1).

First stage comprises initialization, division of original image acquired from the camera into sub-images. It also includes preprocessing and segmentation of each sub-image. It should be pointed out that the output of this stage is matrix (K) containing information of all the clusters in original image. Complete cluster matrix K is obtained by merging resulting cluster matrix of sub-image segmentation.

Second stage comprises segmentation of cluster matrix K as well as decision-making module. Instead of segmentation of sub-image pixel values in the first stage, segmentation in this stage uses cluster centers (mean Cb and mean Cr for each cluster) as input points.

This approach assures that number of points for segmentation stays reasonably low because number of pixels in sub-images is, of course, N times smaller than number of pixels in the original image and number of cluster centers in the second stage is even smaller than the first one.

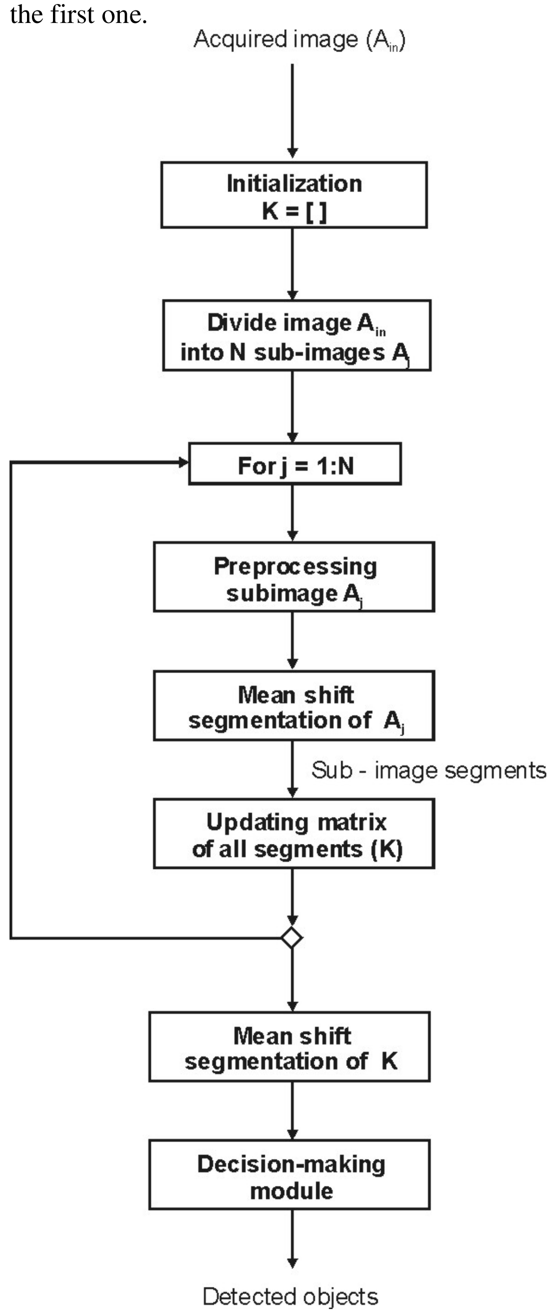

Fig.1. Procedure overview

3 Results

Programming was done using the Matlab 2008a software package with Image Processing Toolbox. Computer on which the processing was done is Pentium Dual Core on 1.86 GHz and 2 Gb of RAM. Installed graphic card was GeForce 8800GT 512 DDR3.

Test images were taken by the digital camera allocated in MI-8 helicopter that was flying at the flight speed of around 100 km/h.

In order to evaluate proposed procedure, we have compared processing speed and output for three different approaches. First approach (A) segments original image as a whole (without subdivision), second approach (B) is based on segmentation of sub-images without segmentation of second stage. Third approach (C) corresponds to the approach described in the previous section. All the approaches are using the same decision-making module as the final filtering step of possible objects.

Fig.2. Example image #1 (original).

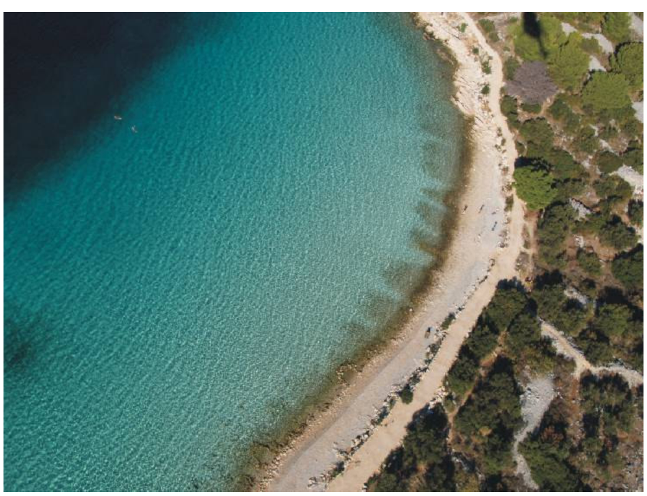

Fig.3. Example image #2 (original).

Approach	A	B	C
Average processing time (in %)	100%	35.5%	37.8 %

Table 1. Comparison of average processing time for different approaches applied to the set of input images. Original images size: 2560 x 1920 pixels. A – no subdivision; B – subdivision without additional segmentation; C – subdivision with additional segmentation.

Fig.4. Example image #1 (result for approach A & C). Detected person in the shadow of the tree.

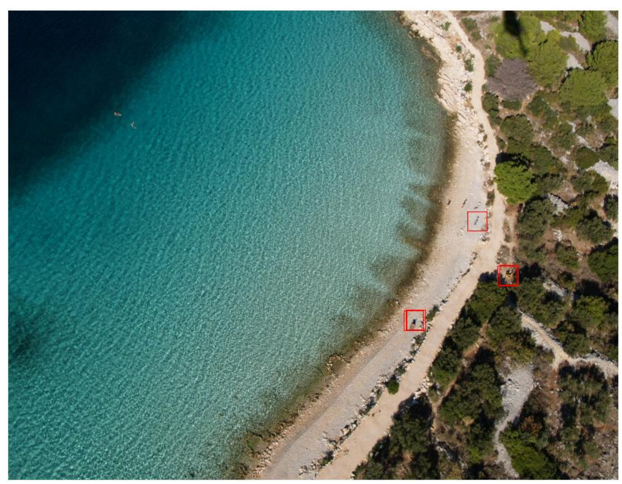

Fig.5. Example image #2 (result for approach A & C).

As it can be seen from Table 1, new approach is almost three times faster than the old one and the obtained results of the suspicious object detection are the same (Fig. 4 and 5). Processing time of the whole image without segmentation on test configuration was 1230 s (100%). In Table 1, results for the approach B (without second segmentation phase) are also given but this approach has been discarded due to poor elimination of potential segments (high false alarm percentage).

4 Conclusion

In this paper we have presented a new two-stage segmentation approach for detection of artificial materials and objects in non-urban terrain. High resolution image of the unknown terrain taken with the digital camera from relatively large distance (> 200 m) is divided into smaller sub-images for further processing. First stage of the segmentation procedure includes sub-image preprocessing and transformation of image in YCbCr color model. Blue-difference chroma component and red-difference croma component is used for the segmentation. Each sub-image is segmented and information about obtained cluster centers is transferred to the next stage. Second segmentation stage uses information about cluster centers from all sub-images and applies the same clustering method to this data set. Finally, decision-making module evaluates all potential candidate segments and eventually proposes segments that have high possibility of presenting artificial material or object in the input image.

This approach is compared to the standard approach that segments the whole image and the results showed that the performance is boosted by factor 3 without degradation of detection results. It should be stressed that this novel approach is adopted for the parallel processing on modern graphical processing units. It means that the next step is the code translation so it could be executed on GPU using CUDA compute engine. Speeding of the overall processing time for at least one order of magnitude is expected so our long term objective is to develop a complete, real or almost 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 localization) seams to be achievable.

Even at the present processing speeds, worth of the application is very high because it allows the rescuers to obtain processed images of the very large area after the flight of helicopter within few hours. Having the corresponding global satellite position of each image helps them in precisely locating the missing person and acting in time.

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

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

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

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

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

[6] 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, 2000, pp. 590-595.

[7] H.D.Cheng, X.H.Jiang, Y.Sun, Jing Li Wang, Color Image Segmentation: Advances & Prospects, *Pattern Recognition*, 34(12), 2001, p.p. 2259-2281.

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

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

[10] B.Barišić, M.Bonković and V.Papić, Evaluation of fuzzy clustering methods for segmentation of environmental images, SoftCOM 2008, *Proceedings of 2008 International Conference on Software, Telecommunications and Computer Networks*, Split-Dubrovnik, Croatia, September, 2008.

[11] A.Gaši, H.Dujmić, V.Papić, H.Turić, On hue and saturation of natural and non-natural materials, SoftCOM 2008, *Proceedings of 2008 International Conference on Software, Telecommunications and Computer Networks*, Split-Dubrovnik, Croatia, September, 2008.

[12] H.Turić, V.Papić, H.Dujmić, A procedure for detection of humans from long distance images, ELMAR '08, *Proceedings ELMAR-2008*, Zadar, Croatia, September, 2008, pp. 109-112.