Radial greed algorithm with rectified chromaticity for anchorless region proposal applied in aerial surveillance

Anton Louise Pernez De Ocampo [a,b,1,*], Elmer Dadios [b,2]

[a] Batangas State University, Rizal Ave. Extension, Batangas City 4200, Philippines
[b] De La Salle University, 2401 Taft Ave., Malate, Manila 1004 , Philippines
[1] anton_louise_deocampo@dlsu.edu.ph; [2] elmer.dadios@dlsu.edu.ph
* corresponding author

ARTICLE INFO

Article history
Received July 8, 2019
Revised August 13, 2019
Accepted November 15, 2019
Available online November 16, 2019

Keywords
Radial greed algorithm
Anchorless region proposal
Human detection
Aerial surveillance
UAV-based monitoring

ABSTRACT

In aerial images, human figures are often rendered at low resolution and in relatively small sizes compared to other objects in the scene, or resemble likelihood to other non-human objects. The localization of trust regions for possible containment of the human figure becomes difficult and computationally exhaustive. The objective of this work is to develop an anchorless region proposal which can emphasize potential persons from other objects and the vegetative background in aerial images. Samples are taken from different angles, altitudes and environmental factors such as illumination. The original image is rendered in rectified color space to create a pseudo-segmented version where objects of close chromaticity are combined. The geometric features of segments formed are then calculated and subjected to Radial-Greed Algorithm where segments resembling human figures are selected as the proposed regions for classification. The proposed method achieved 96.76% less computational cost against brute sliding window method and hit rate of 95.96%. In addition, the proposed method achieved 98.32 % confidence level that it can hit target proposals at least 92% every time.

1. Introduction

Human detection is one of the core subjects in machine vision applications, such as surveillance and monitoring systems. In applications where the video or images are captured using unmanned aerial vehicles (UAV) [1], the difficulty of automatic human detection is amplified by the fact that the Region-of-Interest (ROI) is most likely just a portion of the captured image rendering the human figure at low resolution. Sometimes, the person's figure resembles other non-person objects when looked overhead. Another challenge in the automatic region proposal for human detection is the variation in the human pose.

Detection proposals or region proposals methods are intended to narrow down the areas in the image to examine and classify for human detection. Traditional paradigm such as a sliding window or "brute force" windowing is improved by notable approaches described in a unified list authored by Hosang et al. [2]. Detection proposal methods listed were tested in Pascal VOC [3] and ImageNet [4] datasets. The review was further elaborated in Zhu et al. [5] by identifying the effects of object-level characteristics to the existing detection proposal methods. Size, aspect ratio, color distinctiveness, shape regularity, and texture are just some of the object characteristics that affect the performance of detection proposal methods. The variations on object characteristics were addressed intuitively upon the introduction of

Region Proposal Networks. RPNs are neural networks that are slide through the feature map generated by conventional Convolutional Neural Networks (CNN).

Such architecture that utilizes RPN for detection proposal was first presented as Faster R-CNN in the works of Ren et al. [6] and Salvador et al. [7]. The term RPN or region proposal network is used to refer to a small neural network, composed of a regressor and classifier modules, within a larger deep neural network model. The classifier determines the probability that the object of interest is in the proposed region, while the regressor identifies the coordinates of the proposed regions. Technically, RPNs still perform sliding window method in making a proposal, only, the coverage in RPN is not the whole image but the location of anchors. Anchors are predefined bounding boxes that are distributed across the image at different scales and aspect ratio [8], [9]. Another difference with the traditional sliding window, the kernel used in RPN is slide across the feature map of the output of the last convolutional layer.

CNN-based region proposal methods use high-level semantics while conventional detection proposals use low-level object features. Li et al. described a CNN-based region proposal method which is based on utilizing feature maps at different anchor sizes to identify "objectness" in an image. The Zoom Out-and-In Network with Map Attention Decision or ZIP uses a "MAP" unit to determine the feature map using a selective vector that allows neurons from low- and high-level layers to decide where to focus the attention [10]. This is quite different from the approach used by Wang et al. [11]. The problem of using a dense number of anchors for region proposal is addressed by using Guided Anchoring (GA) which leverages the semantic features to guide where the anchors should be placed. An anchor generation module is used to predict anchor location and shape for each feature map in the feature pyramid. Although guided anchors improved RPNs, the presence of variations in scales and position still affects RPN's performance. This is addressed by Scale-Invariance and Position-Sensitive RPN that uses 4 interdependent modules, namely: translation-invariant objectness classification, translation-invariant bounding box regression, large effective receptive fields and scale-invariant region proposal [12]. An RPN can also be cascaded with another to improve localization. Cascade region proposal is technically putting another RPN after an RPN in a classifier network. Since RPN2 is based on the output of RPN1, the score and location of the proposed regions are refined [13]. Another CNN-based RPN is the AttractionNet which is centered on the use of pre-defined boxes that are initially distributed across the image uniformly, and then, subsequently, move as iteration continues. Each ARN is composed of an objectness module and a location refinement module [14]. Similar to AttractionNet, which uses a small network just before classification, HyperNet inserts an additional feature map called Hyper Features. This Hyper Feature is a concatenation of multiple feature maps into a cube. The objective of HyperNet is to reduce the number of proposals per image [15]. In addition to the list of CNN-based RPNs is SPOP-net which is a deep neural network that has two localization modules intended to define the positions of small objects using a divide-and-conquer approach [16].

Another approach to the region proposal is to use an algorithm based on traditional image processing techniques such as Superpixel and Markov Random Field. Superpixel is a segment generated by over-segmentation. The purpose of super-pixelation is to reduce the computations in many computer vision applications [17]–[19]. By adding boundary constraints in superpixel segmentation and allowing the positions to update every iteration, the superpixel segmentation can improve in terms of effectiveness and accuracy as used in region proposal systems. Initially, pre-defined seed points are scattered across the image. The seeds' positions shift as the superpixel refinement is activated. This is accomplished by continuously adjusting the distance function [20]–[23]. Aside from putting boundary constrains to create super-pixelation, segmentation can be achieved by forcing spatial coherence between adjacent regions in an image. In region-based MRF, pivot regions are used to adopt the MRF model for unsupervised segmentation. It can also adapt to varying image complexity by making the superpixels finer at first. Then local energies per segment are calculated to assign constraints for each segment [24]–[26].

Because of the nature of RPN's architecture, training is therefore required. This becomes a challenging task when an adequate dataset is not available. Unlike machine learning classifiers that can

be trained with a relatively small dataset, CNN classifiers need a large size training dataset [27]–[29]. Another problem encountered in the use of RPNs for region proposal is its inability to localize small objects accurately due to the low-resolution feature map representing small objects [30]. In aerial surveillance and monitoring, it is common that the target objects appear relatively small and only occupy a very small portion of the entire image. In such applications, a region proposal method that can identify trust regions for smaller target objects is needed. In a more specific application of aerial surveillance such as in large smart farm monitoring, persons or livestock are often rendered small.

In this work, the following are the identified contributions: A dataset for persons working in a vegetative environment; and an anchorless, unsupervised region proposal algorithm that doesn't require training and can emphasize objects of interest from other objects and vegetative background. The rest of the paper is arranged as follows: Section 2 discusses the conduct of experimentation carried out; Section 3 discusses the results of the experiment and test; and, lastly, Section 4 summarizes the relevant points.

2. Method

2.1. Experiment set-up

Surveillance videos of people working in the field are captured using unmanned aerial vehicle equipped with a full HD camera with 4K resolution paired with a 3-axis gimbal system for mechanical image stabilization. The camera is configured to three different azimuths, 30-, 45- and 90-degrees, which is used one per flight mission. The UAV is programmed to cover the exact same location in the field which lasts on an average of 15 minutes per flight. The video recording is configured to take 3840 x 2160 frames at 23.98 fps from different altitudes: 5m, 7m, 15m, and 30m. Each recording is converted to frames and each frame is annotated with the number of persons present as well as the coordinates of the bounding boxes containing these persons. In Fig. 1, samples of these images are presented.

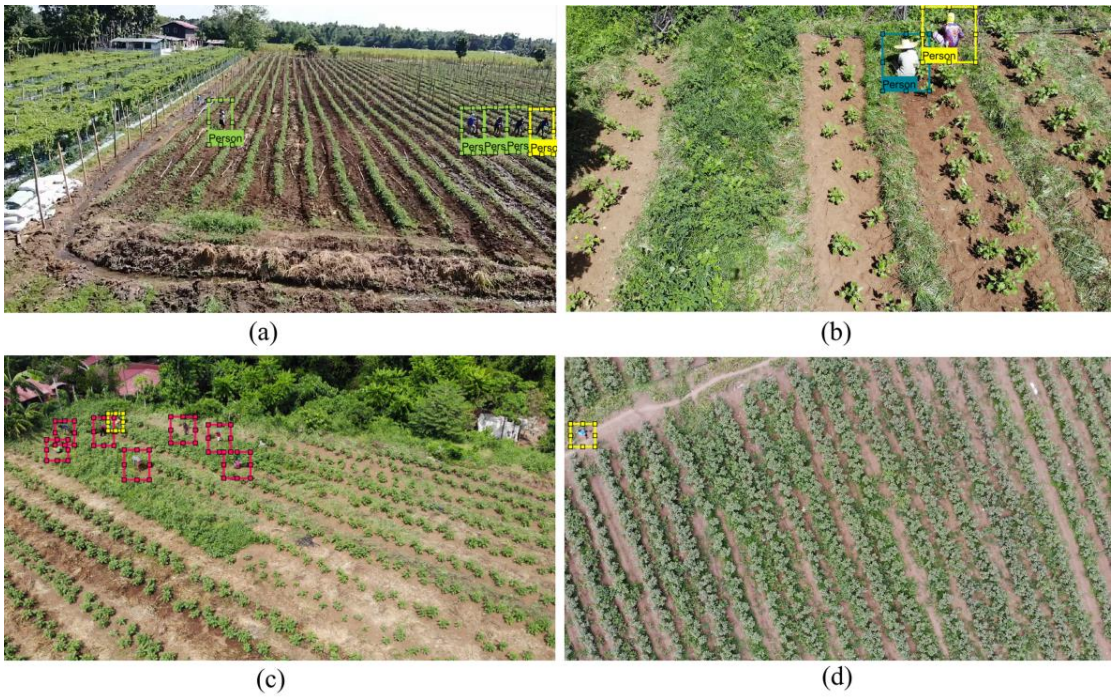

Fig. 1. Samples of frames extracted from aerial recordings. Camera azimuth: 30-deg (a), 45-deg (b, c), 90-deg (d). Altitude: 7m (a,b), 15m (c), 30m (d)

The dataset on which the proposed method is characterized by the huge variation in the angle, height, bearings, illumination, scales, and pose. With such a dataset, the robustness of the proposed method over such variations can be estimated. The proposed method is evaluated by counting the number of

proposed regions whose centroids are within the ground truth bounding box of the annotated image, and otherwise. In addition, miss and hit rates are also measured. To compare the proposed method with the standard sliding window method for region proposal, the time it takes to perform person detection from a set of images while using the proposed method and conventional sliding window as region proposal mechanisms is measured.

There are 13 videos with an average length of 14 minutes each, which are used as test dataset for evaluation of the proposed method. The footages are taken around eight to ten o'clock in the morning of different dates on the 11-hectare farm in Ibaan, Batangas, Philippines. The implementation of the proposed method is on a Windows 10 workstation running with an Intel i7 processor at 3.20 GHz, equipped with 8Gb internal ram and another 8Gb memory for GPU.

2.2. Radial-Greed Algorithm with Rectified Chromaticity (proposed algorithm)

Chromaticity is the quality of color independent of luminance. By considering chromaticity only, the proposed method becomes more robust to varying illumination. In the proposed method, the original aerial image in RGB is translated to CIELAB color space and is rectified to levels where the gradient in luminosity, L- channel, is reduced to zero while a*b*- channels are normalized to binary levels. The rectification of chromaticity in CIELAB color space allows the merging of neighboring pixels included in an object. This process can be likened to crude image segmentation.

Region proposal methods for human (or object) detection have the common objective, that is, to localize the trust regions where potential human (or object) figures can be found. The proposed method is presented as a flowchart shown in Fig. 2. Consider the j^{th} frame on a captured video of farmers working in the field. The RGB image, I_j, is converted to La*b* color space of which only the chromatic channels are subjected for rectification. The L-channel, luminosity, is normalized to unity to significantly reduce the variability between pixels of the same object in terms of lightness. Rectification of a*- and b*-channels are described in (1) and (2) which translates all colors to the nearest corner in CIELAB color-space. This process creates a pseudo-segmented four-color version of the image.

$$R_{ja} = I_{ja}(x,y) = \begin{cases} 1, & a^* \geq 0 \\ 0, & otherwise \end{cases} \quad (1)$$

$$R_{jb} = I_{jb}(x,y) = \begin{cases} 1, & b^* > 0 \\ 0, & otherwise \end{cases} \quad (2)$$

The pixel values of the rectified image are referred to as R which is equal to the index image, I, in CIELab colorspace. In Table 1, the four-color corners of CIELAB at constant L are presented. All other colors are reduced cyan, green, magenta and yellow. Objects commonly found in natural sceneries such as vegetation and earth can easily be removed using this new color scheme.

Table 1. Rectified values of corner colors of CIELAB

Rectified Channels			Corner Colors
L	*a**	*b**	
1	0	0	Cyan
1	0	1	Green
1	1	0	Magenta
1	1	1	Yellow

By rectifying chromaticity, color segments are formed of which geometric features are extracted. Geometric features include the area of blobs created AN, centroids CN, distances between centroids D, bearings of one centroid from another θ, width w and height h of the bounding box around the extents of these blobs. These features are used for the Radial-Greed algorithm which combines blobs that are

potentially part of a human silhouette in an image. Radial-Greed Algorithm is described in Table 2. In *Sequence 1*, all the generated geometric features of the blobs are initiated. The number of blobs is trimmed down using the conditions presented in *Sequence* 3. Aside from the area, height and width, the self-bearing and distance from the nearest blobs are then used to combine segments.

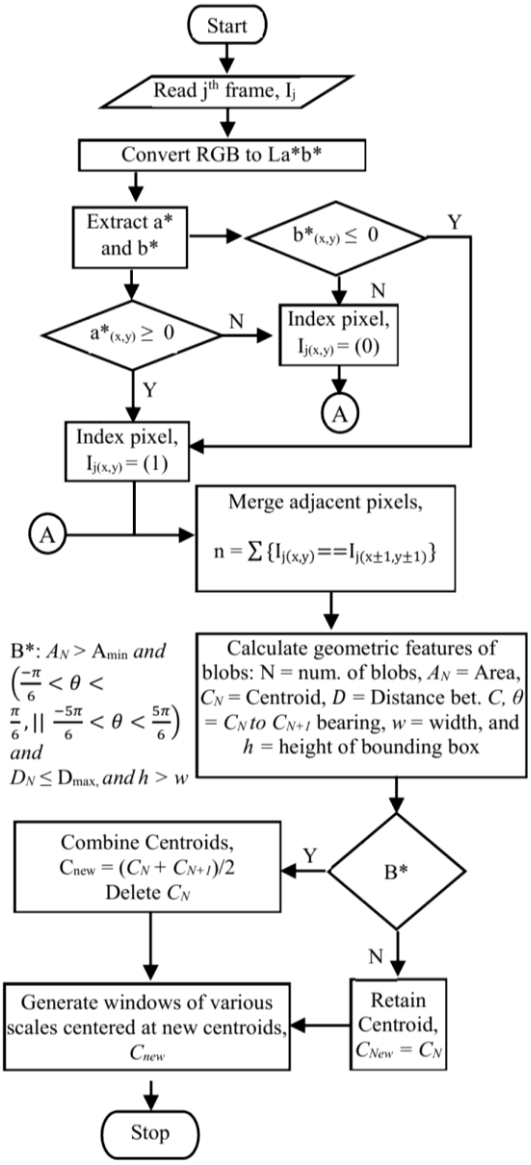

Fig. 2. Overview of the anchorless region proposal via Radial-Greed Algorithm with rectified chromaticity

Table 2. Radial-Greed Algorithm for merging segments of a human figure

Sequence	Descriptions
1	initialize calculated features of selected segments: A_N, C_N, D, θ, w & h, $Pool = \{\}$
2	**for** $N = 1$ to *num. of blobs* **do**
3	**if** $A_N > A_{min}$ && $h > w$ **then,** $Pool\{end+1\} = index\ (C_N)$; **end if**
4	**sort** rows in ascending $C_N(x)$; **compute** $\theta_N = arctan\left(\frac{C_N(y) - C_{N+1}(y)}{C_N(x) - C_{N+1}(x)}\right)$
5	**if** $-\frac{\pi}{6} < \theta_N < \frac{\pi}{6} \| -\frac{5\pi}{6} < \theta_N < \frac{5\pi}{6}$, **compute** $D = \sqrt{(C_{N+1}(x) - C_N(x))^2 + (C_{N+1}(y) - C_N(y))^2}$
6	**if** $D < D_{max}$, **compute** $C_{N(new)} = (C_{N+1} + C_N)/2$ remove C_N and C_{N+1} from Pool{}
7	**else,** $C_{N(new)} = C_N$; **end if, end if, end for**

In Fig. 3, the iterations for centroid merging are presented with the selected segments having the potential to be included in a human silhouette. Suppose that the current centroid under observation is C1. Centroids C5 and C7 fall within the range D_{max} set for merging. However, the bearings of these centroids with respect to C1 do not confine with the conditions for the bearing angle θ. Therefore, no centroid shall be combined with C1 for the conditions that were not fully met. The next iteration is in reference to the nearest centroid along the x-axis which is C2. As seen in Fig. 3b and 3c, no adjacent centroids comply with the merging conditions. Iterations continue centroid C4 where it is combined with centroid C5 forming a new centroid C45 which will be used instead of centroid C5 as the reference in the next iteration.

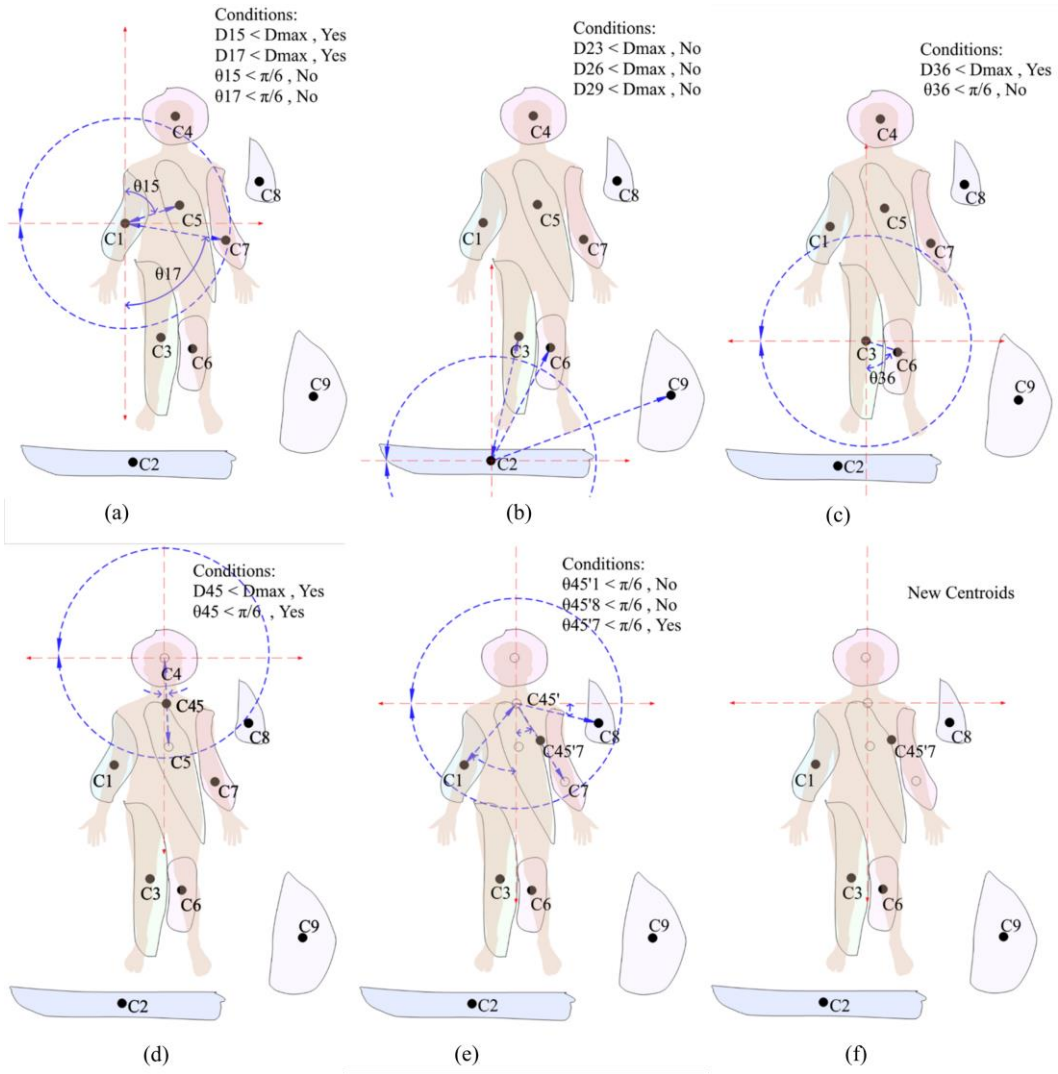

Fig. 3. Iterations of Radial-Greed Algorithm where centroids are combined.

The proposed regions are the patches of different scales that are centered at each resulting centroid generated after the merging process. The patch sizes are determined by measuring the smallest bounding box encapsulating the merged segments and multiplying it to the number of scales defined.

3. Results and Discussion

The initial step in the proposed method is to rectify the chromaticity of the test image. The objective of this process is to emphasize the differences in chromaticity of objects within an image. Consider the set of sample images presented in Fig. 4. Two persons, in white clothes, working in the field are presented in Fig. 4a, while two trashed papers, also in white, are shown in Fig. 4b. Upon rectification of

the images' chromaticity, the parts that belong to the persons and the paper trash are both emphasized. These results are shown in Fig. 4c and 4d. Although the trashed papers and the persons have the same colors, it can be noticed that the proposed method for selecting ROIs successfully captured the regions where the two persons are located (Fig. 4e) but didn't pick the regions containing the trash (Fig. 4f). This is the purpose of the second step in the proposed method. Radial Greed Algorithm selects segments that may contain parts of the human figure and combines them into one larger segment. If the combined segment meets the conditions required by RGA (as described in Section 2.2), then that segment will be marked as a proposed ROI or trust region. The proposed method is also robust against variation of scale which is very common on UAV aerial images.

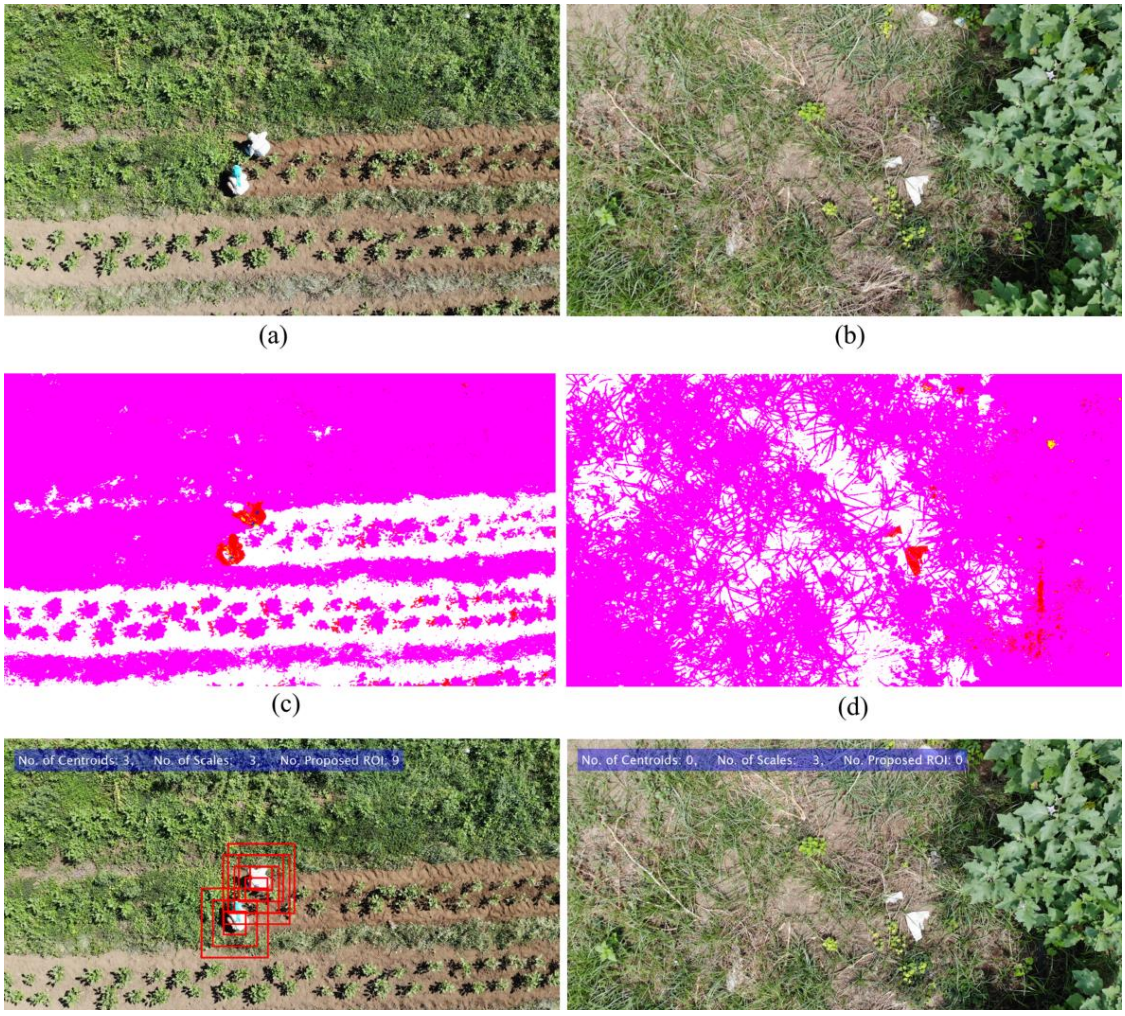

Fig. 4. Original image of two persons (a) and trash paper (b); rectified chromaticity (c,d); and, regions proposed (e,f). No regions are proposed for (f) since no candidate for a human figure is present.

The proposed method is also tested on images that contain both persons and non-person objects. Fig. 5 shows an image that contains a person standing (in blue clothes) and other objects (white straw sacks). The proposed method successfully picked the person and ignored the other objects. In addition, the proposed method is able to capture even small human figures. Another property of the proposed method is its robustness to variation in color. In previous samples, the clothes worn by the persons are of a single color. Consider the image in Fig. 6. Due to variations in the colors between clothes used by each person, color-based segmentation cannot effectively highlight the segments that belong to non-vegetative materials. By rectifying chromaticity, the original image is translated into a four-color version where the non-vegetative materials are highly emphasized. In Fig. 7, the translated image shows that by rectifying chromaticity, adjacent materials that are not vegetative can be combined.

Fig. 5. A person in blue clothes is captured in the proposed regions (3 scales of red bounding boxes) while other objects are ignored.

Fig. 6. Original subject image for persons in various colors of clothes.

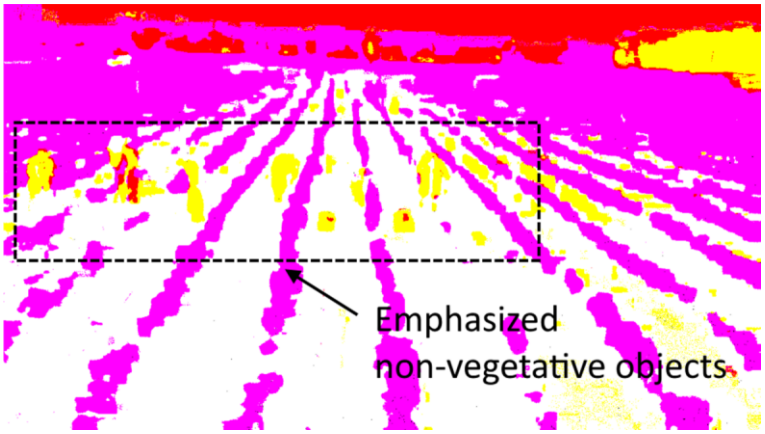

Fig. 7. Result of rectifying chromaticity of the image. The four-color segmented image is produced where non-vegetative materials are emphasized.

Then, the geometrical features of these segments such as centroid, area, orientations, minor and major axes are calculated. In Fig. 8, the centroids of each blob are presented using blue markers. Here it can be noticed that although the adjacent colors are already combined, a multitude of small blobs can still appear anywhere in the image. For this reason, the Radial-Greed algorithm is used to combines centroids of blobs that satisfy the configured conditions. Fig. 9 shows the result of RGA.

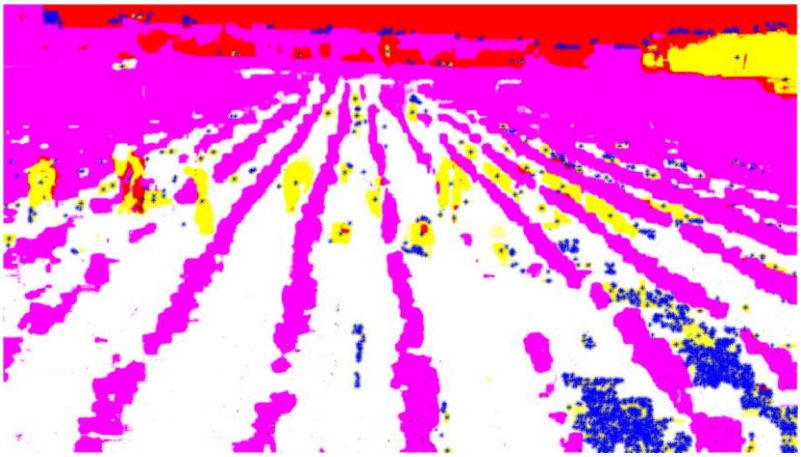

Fig. 8. Example of a figure caption.

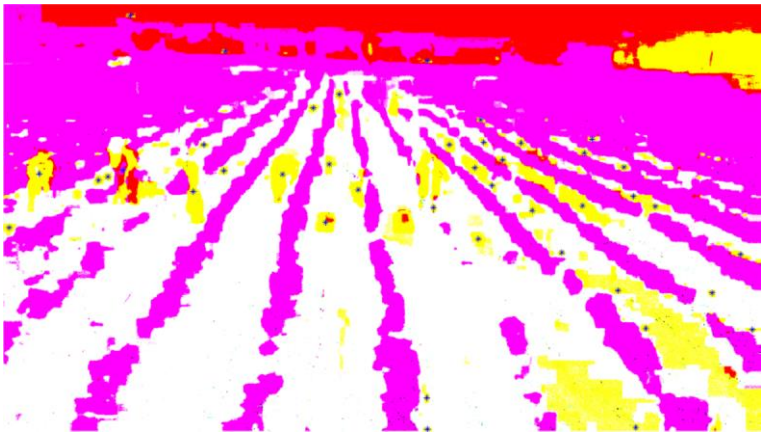

Fig. 9. Centroids of regions that may contain human figures are retained after RGA.

Using the generated centroids from RGA, the proposed regions that will be subjected to human detection are shown in Fig. 10. It can be noticed that all the ROIs are successfully selected by the region proposal method.

Fig. 10. Proposed regions of interest (ROI) captured all regions containing the persons.

The first evaluation used to estimate the performance of the proposed method is to count the correctly proposed regions that may contain a human figure. In Table 3, the number of actual ROIs containing a human figure is presented together with the proposed ROIs whose centroids are found inside the actual ROI. It may happen that multiple centroids of proposed regions fall inside the actual

ROI. Actual regions of interests, which did not contain any centroid after the proposed algorithm is performed, are considered "missed" proposals. Miss and hit rates are calculated as described in (3) and (4).

$$MissRate = \frac{\#MissedROI}{\#ActualROI} x100\% \quad (3)$$

$$HitRate = 100\% - MissRate \quad (4)$$

Table 3. Hit-rate of the proposed algorithm

Video No.	A	B	C	D	% Miss	% Hit
1	272	615	1115	0	0 %	100.00%
2	159	562	2268	0	0 %	100.00%
3	143	145	0	27	18.88%	81.12%
4	141	571	766	0	0.00%	100.00%
5	292	332	378	14	4.79%	95.21%
6	159	285	1033	0	0.00%	100.00%
7	160	199	0	20	12.50%	87.50%
8	176	258	371	7	3.98%	96.02%
9	177	510	2334	0	0.00%	100.00%
10	174	364	11	17	9.77%	90.23%
11	302	979	1108	0	0.00%	100.00%
12	128	296	14	5	3.91%	96.09%
13	292	332	378	14	4.79%	95.21%
Std.Dev	0.06			Avg.	4.04%	95.96%

Note: A = Number of actual ROIs (ground truth)
B = Number of proposed regions whose centroids are within the actual ROIs
C = Number of proposed regions whose centroids are beyond the actual ROIs
D = Number of missed ROIs.

The proposed method achieved a 95.95 % hit rate using the test videos available. The test videos contain a total of 2575 occurrences of persons within all the frames. In Table 4, a one-tail z-test is performed on the results of the proposed method on the processing of the 13 videos. The proposed method can capture ROIs with at least a 92% hit rate with 98.32% confidence level.

Table 4. Confidence Levels for Achieving Various Hit Rates

Hit Rate	p-value	Confidence level
100 %	0.99697835	0.30%
98 %	0.93672033	6.33%
95 %	0.38262064	61.74%
92 %	0.01679239	98.32%
90 % [a]	0.00041493	99.96%

[a] It is 99.96 % sure that the proposed method can achieve a hit rate of 90 %

Table 5 shows the cost savings of implementing the proposed method instead of the conventional sliding window. The basis of comparison is the time it takes to extract the GeHOG features of a 1280 x 720 x 3 image.

Table 5. Time Cost of Extracting Features [a] from a 1280 x 720 RGB Image

	Time (s)	No. of Proposals
Sliding Window	147.42	2801
Proposed Method	4.78	60
Cost savings	**96.76%**	**97.86%**

Note: [a] Feature descriptor used in extraction is GeHOG

4. Conclusion

The presented method for region proposal has effectively addressed the problem of generating proposed regions for small objects and low-resolution images. By rectifying chromaticity, the objects that are non-vegetative and unnatural are easily emphasized without the need of training the region proposal method. Radial-Greed algorithm successfully combined rectified segments that may belong to a human figure. The proposed method achieved a 95.96% hit rate for the test data used while obtaining a 98.32 % confidence that the hit rate would be greater than 92% and a 99.96% confidence for achieving a 90% hit rate at any given sample. Moreover, the proposed method can be used in conjunction with different classifier types used in human detection.

Acknowledgment

The authors want to thank the Gokongwei College of Engineering of De La Salle University, Engineering Research and Development for Technology (ERDT) of Department of Science and Technology (DOST), and Batangas State University for the support on this research.

References

[1] H. Il Son, "The contribution of force feedback to human performance in the teleoperation of multiple unmanned aerial vehicles," *J. Multimodal User Interfaces*, vol. 13, no. 4, pp. 335–342, Dec. 2019, doi: 10.1007/s12193-019-00292-0.

[2] J. Hosang, R. Benenson, P. Dollar, and B. Schiele, "What Makes for Effective Detection Proposals?," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 4, pp. 814–830, Apr. 2016, doi: 10.1109/TPAMI.2015.2465908.

[3] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes Challenge: A Retrospective," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, Jan. 2015, doi: 10.1007/s11263-014-0733-5.

[4] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.

[5] H. Zhu, S. Lu, J. Cai, and G. Lee, "Diagnosing state-of-the-art object proposal methods," in *Procedings of the British Machine Vision Conference 2015*, 2015, p. 11.1-11.12, doi: 10.5244/C.29.11.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[7] A. Salvador, X. Giro-i-Nieto, F. Marques, and S. Satoh, "Faster R-CNN Features for Instance Search," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016, pp. 394–401, doi: 10.1109/CVPRW.2016.56.

[8] P. Tang *et al.*, "Weakly Supervised Region Proposal Network and Object Detection," in *The European Conference on Computer Vision (ECCV)*, 2018, available at : Google Scholar.

[9] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High Performance Visual Tracking with Siamese Region Proposal Network," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8971–8980, doi: 10.1109/CVPR.2018.00935.

[10] H. Li, Y. Liu, W. Ouyang, and X. Wang, "Zoom Out-and-In Network with Map Attention Decision for Region Proposal and Object Detection," *Int. J. Comput. Vis.*, vol. 127, no. 3, pp. 225–238, Mar. 2019, doi: 10.1007/s11263-018-1101-7.

[11] J. Wang, K. Chen, S. Yang, C. C. Loy, and D. Lin, "Region Proposal by Guided Anchoring," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, available at : Google Scholar.

[12] H.-F. Lu, X. Du, and P.-L. Chang, "Toward Scale-Invariance and Position-Sensitive Region Proposal Networks," in *Computer Vision -- ECCV 2018*, 2018, pp. 175–190, doi: 10.1007/978-3-030-01237-3_11.

[13] Q. Zhong, C. Li, Y. Zhang, D. Xie, S. Yang, and S. Pu, "Cascade region proposal and global context for deep object detection," *Neurocomputing*, 2019, doi: 10.1016/j.neucom.2017.12.070.

[14] S. Gidaris and N. Komodakis, "Attend Refine Repeat: Active Box Proposal Generation via In-Out Localization," in *Procedings of the British Machine Vision Conference 2016*, 2016, p. 90.1-90.13, doi: 10.5244/C.30.90.

[15] T. Kong, A. Yao, Y. Chen, and F. Sun, "HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 845–853, doi: 10.1109/CVPR.2016.98.

[16] Z. Jie, X. Liang, J. Feng, W. F. Lu, E. H. F. Tay, and S. Yan, "Scale-Aware Pixelwise Object Proposal Networks," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4525–4539, Oct. 2016, doi: 10.1109/TIP.2016.2593342.

[17] V. Jampani, D. Sun, M.-Y. Liu, M.-H. Yang, and J. Kautz, "Superpixel Sampling Networks," 2018, pp. 363–380, doi: 10.1007/978-3-030-01234-2_22.

[18] W. Tu *et al.*, "Learning Superpixels with Segmentation-Aware Affinity Loss," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 568–576, doi: 10.1109/CVPR.2018.00066.

[19] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art," *Comput. Vis. Image Underst.*, vol. 166, pp. 1–27, Jan. 2018, doi: 10.1016/j.cviu.2017.03.007.

[20] N. Sun, F. Jiang, H. Yan, J. Liu, and G. Han, "Proposal generation method for object detection in infrared image," *Infrared Phys. Technol.*, vol. 81, pp. 117–127, Mar. 2017, doi: 10.1016/j.infrared.2016.12.021.

[21] Y. Zhang, X. Li, X. Gao, and C. Zhang, "A Simple Algorithm of Superpixel Segmentation with Boundary Constraint," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2016, doi: 10.1109/TCSVT.2016.2539839.

[22] X. Chen, H. Ma, C. Zhu, X. Wang, and Z. Zhao, "Boundary-aware box refinement for object proposal generation," *Neurocomputing*, vol. 219, pp. 323–332, Jan. 2017, doi: 10.1016/j.neucom.2016.09.045.

[23] J. Yan, Y. Yu, X. Zhu, Z. Lei, and S. Z. Li, "Object detection by labeling superpixels," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5107–5116, doi: 10.1109/CVPR.2015.7299146.

[24] Y. Yang, L. Guo, and Y. Ye, "Robust natural image segmentation by using spatially constrained multivariate mixed Student's t -distribution and TV flow edge," *J. Vis. Commun. Image Represent.*, vol. 40, pp. 178–196, Oct. 2016, doi: 10.1016/j.jvcir.2016.06.022.

[25] N. K. O and C. Kim, "Unsupervised Texture Segmentation of Natural Scene Images Using Region-based Markov Random Field," *J. Signal Process. Syst.*, vol. 83, no. 3, pp. 423–436, Jun. 2016, doi: 10.1007/s11265-015-1030-4.

[26] D. Yeo, J. Son, B. Han, and J. H. Han, "Superpixel-Based Tracking-by-Segmentation Using Markov Chains," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 511–520, doi: 10.1109/CVPR.2017.62.

[27] A. L. P. de Ocampo and E. P. Dadios, "Mobile Platform Implementation of Lightweight Neural Network Model for Plant Disease Detection and Recognition," in *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*, 2018, pp. 1–4, doi: 10.1109/HNICEM.2018.8666365.

[28] R. G. de Luna *et al.*, "Identification of philippine herbal medicine plant leaf using artificial neural network," in *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology,*

Communication and Control, Environment and Management (HNICEM), 2017, pp. 1–8, doi: 10.1109/HNICEM.2017.8269470.

[29] I. C. Valenzuela *et al.*, "Quality assessment of lettuce using artificial neural network," in *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2017, pp. 1–5, doi: 10.1109/HNICEM.2017.8269506.

[30] C. Eggert, S. Brehm, A. Winschel, D. Zecha, and R. Lienhart, "A closer look: Small object detection in faster R-CNN," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 421–426, doi: 10.1109/ICME.2017.8019550.