Texture based Animal Segmentation in Aerial Videos

By Rishaad Abdoola, Yunfei Fang, Shengzhi Du, Paul Bartels
& Christiaan Oosthuizen

Tshwane University of Technology

Abstract- Animal detection in aerial videos is a challenging problem due to the complex nature of the scenes involved as well as the natural ability of the animals to camouflage their environment. To assist with the detection and classification of animals for the purpose of nature conservation management, texture analysis is applied to aerial videos of wildlife scenes to segment the environment from the animals. To perform automatic wildlife surveying and animal monitoring, it is proposed to use GLCM texture segmentation to reduce the search area for animals in the aerial videos. Using the texture in the scene, the issues of a moving background and unpredictable state of the animal are avoided. The method presented is well suited to implementation on a UAV as it is easily parallelizable.

Keywords: image segmentation, GLCM, texture analysis, animal tracking, animal segmentation.

GJRE-A Classification: UDC: 004.896, 591.5, 621.396

TEXTUREBASEDANIMALSEGMENTATIONINAERIALVIDEOS

Strictly as per the compliance and regulations of:

Texture based Animal Segmentation in Aerial Videos

Rishaad Abdoola [α], Yunfei Fang [σ], Shengzhi Du [ρ], Paul Bartels [ω] & Christiaan Oosthuizen [¥]

Abstract- Animal detection in aerial videos is a challenging problem due to the complex nature of the scenes involved as well as the natural ability of the animals to camouflage their environment. To assist with the detection and classification of animals for the purpose of nature conservation management, texture analysis is applied to aerial videos of wildlife scenes to segment the environment from the animals. To perform automatic wildlife surveying and animal monitoring, it is proposed to use GLCM texture segmentation to reduce the search area for animals in the aerial videos. Using the texture in the scene, the issues of a moving background and unpredictable state of the animal are avoided. The method presented is well suited to implementation on a UAV as it is easily parallelizable.

Keywords: image segmentation, GLCM, texture analysis, animal tracking, animal segmentation.

I. Introduction

Nature conservation and wildlife management depends on accurate and reliable animal census data over time. Such surveys are often conducted on the ground by conservation workers or from the air utilizing rotary or fixed wing aircraft. The former requires a large amount of manpower while the latter is too expensive for the majority of nature reserves and wildlife ranches. Performing census in this manner, means that data cannot be collected regularly, reducing the effectiveness of nature conservation management. An alternative, more efficient and effective method for wildlife data collection is required.

The development and reduction in costs of Unmanned Aerial Vehicles (UAVs) have made them affordable and easy to operate. UAV's have become popular in aerial surveillance and data collection. Applying UAV's to wildlife surveying seems sensible but offers a number of challenges. Since there are currently no well-developed existing systems, the problem lies in the system structure and analysis of methods of the data collected by the UAV.

In a wildlife scenario, a UAV is launched to collect data of regions where animals are active. A programmed FPGA kit is to be mounted on the UAV, serving as an onboard analyzing system, to detect animals quickly and automatically. Frames containing animals will be labelled and recorded for further analysis. Methods with improved accuracy will be applied to locate and track animals in these frames to reveal herd patterns or location information of various species. For each task, methods will be evaluated to minimize error, and to reduce computational load due to the constraints of the UAV.

This research aims to add to a larger project undertaking involving the building of a UAV based wildlife census system to provide an autonomous solution to a series of challenges in wildlife conservation. Tracking animals in the wild is one of the most complex contexts as it deals with tracking multiple objects from a moving camera with a dynamic, cluttered background. The main challenges in animal tracking can be summarized as follows:

- *Aerial perspective:* Appearance of the animals and objects would be totally different from the ground perspective.
- *Moving camera:* Abundant work has been done on object detection in stationary scenes. A moving camera indicates the pixel values of the background changes over time negating most background subtraction methods for detection.
- *Dynamic background:* A dynamic background is similar to a moving background in which the pixel value changes but obeys a regular pattern. Examples of a dynamic background include noise, changing illumination, waves on water, trees waving in the wind, shadows, smoke, fire flames, etc.
- *Unpredictable state of the animal:* Successfully applied tracking systems usually assume that the object is always in motion; some even assume the orientation, velocity, or acceleration of the motion. But the state of the animals cannot be predicted in the wild.
- *Protective coloration:* Animals' skin tends to exhibit similar color and texture to the background, which will reduce the effectiveness of methods that are sensitive to color.

In this paper we focus on reducing the search area for the animal in the scene by using both the texture of the environment and the animal. Using the texture in the scene, the issues of a moving background and unpredictable state of the animal are avoided.

Author α σ ρ: Department of Electrical Engineering, French South African Institute of Technology, Tshwane University of Technology, Staatsartillerie Road, Pretoria, South Africa.

Author ω: Department of Nature Conservation, Tshwane University of Technology, Staatsartillerie Road, Pretoria, South Africa.

Author ¥: Department of Mechanical and Mechatronics Engineering, Tshwane University of Technology, Staatsartillerie Road, Pretoria, South Africa. e-mail: oosthuizencc@tut.ac.za

II. RELATED WORK

Sirmacek and Wegmann [1] extracted local features to detect focus regions in aerial images and then applied a mean-shift segmentation algorithm to detect and locate animals in the image. Many aerial images were used to test and evaluate the algorithm. This method was aimed at segmentation in single images and did not address the topic of detection and tracking in videos. Gemert and Verschoor [2] investigated the current detection methods designed for human centered objects and evaluated three lightweight detection and tracking algorithms suitable for onboard implementation. The datasets for experimentation were videos of cows recorded on a farm, which has a much less sophisticated environment and is not representative of the types of scenes we encounter in the wild. Animal counting is based on detection, and KLT (Lucas-Kanade-Tomasi) tracker for salient points was used which is more stable to enumerate. Counting results showed a very low precision.

In conventional outdoor applications like pedestrian surveillance and traffic monitoring, a dynamic background is obviously inevitable. A common idea is to build a background model based on the statistical distribution of the pixel color value and distinguish the pixels that belong to the foreground. In [3] a Gaussian mixture model (GMM) was used to model each pixel and utilize an on-line approximation to update the model. The background pixels were classified based on whether the Gaussian distribution represents it effectively or not. In [4], the background was modelled by an Autoregressive Moving Average Model (ARMA), followed by the application of a Kalman Filter to estimate the state of the object in successive frames. Implementations were done on single objects in outdoor scenes. C. Ridder, O. Munkelt, and H. Kirchner [5] also proposed a Kalman filtering based algorithm to form a real-time tracker that can deal with illumination changes and repetitive motions of the background. The system was successfully applied in human body tracking in a real-world scene. All of these algorithms are focused on scenes with stationary cameras and constant background.

To deal with moving object detection with moving cameras, a typical method is the extension of background subtraction. In [6], the authors used registration methods to model the background in subsequent frames. Backgrounds from different frames were then stitched together into a planar mosaic. The moving objects were then segmented in a similar way to the stationary camera case. The method was implemented in indoor scenes to detect human bodies. In [7] the authors extended such mosaic algorithms to the outdoor scene to monitor vehicles on highways from an aerial perspective. Motion-based approaches [8] were also proposed to detect and track moving objects with moving cameras. In [9], the authors proposed to use a GMM to model the motion changes (optical flow value changes) for background subtraction. Pixel value modeling that is usually used in a stationary camera scene is replaced by motion vectors, since it is assumed that the motion of the background pixels follows some pattern like that of the pixel values in a stationary camera scene. This is also orientated to vehicle detection and tracking.

III. PROPOSED METHOD

The proposed method is to use the GLCM of both a texture of the background and the animal species and then determine the intersection of the two to generate the final result.
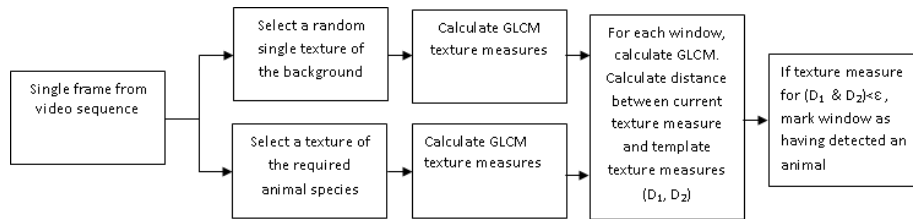

Fig. 1: Flow chart of algorithm

a) Gray Level Co-Occurrence Matrix

The grey Level Co-occurrence Matrix (GLCM) is defined over an image to be the distribution of co-occurring values at a given offset. It measures how often different combinations of pixel brightness values or grey levels occur in an image. GLCM is based on the relationship between two pixels. The distance (in terms of pixel) between a reference pixel and its neighbor defines the offset. The offset will be equal to one for this paper. Larger offsets could be used, resulting in a smaller number of pixel combinations. The neighboring pixel can take different positions with respect to the reference pixel; it can be next to (left, right), above or on the diagonal with respect to the reference pixel. The neighbor pixel will be selected to be to the right of the reference pixel for this paper. To reduce the size of the GLCM and increase the occupancy level of the matrix, the quantization level is set to 16.

b) Texture Measures from the GLCM

Texture calculations performed on the GLCM are of "second order". First order texture measures are

calculated from the original image values and they do not consider spatial relationships. Second order measures consider the relationship between groups of two neighbor pixels in the original image. Texture calculations are weighted averages of the normalized GLCM cell contents. Once a texture calculation is performed for one specific GLCM corresponding to a particular position of the window, the window moves to the next position and the same procedure is repeated. Texture measures can be classified into different categories depending on the type of information they provide. Certain measures are based on the contrast information while others are based on the orderliness and the descriptive statistics of the GLCM texture measures. The following describes only the texture measures used in this paper.

i. *Contrast*

$$Contrast = \sum_i \sum_j |i - j|^2 P(i,j) \qquad (1)$$

where *P(i, j)* is the probability at row *i* and column *j* of the GLCM.

ii. *Energy*

$$Energy = \sum_i \sum_j P(i,j)^2 \qquad (2)$$

where *P(i, j)* is the probability at row *i* and column *j* of the GLCM. Energy will be equal to 1 for a constant image.

iii. *Correlation*

$$Correlation = \sum_i \sum_j \frac{(i - \mu_i)(j - \mu_j)P(i,j)}{\sigma_i \sigma_j} \qquad (3)$$

where $\mu_i, \mu_j, \sigma_i, \sigma_j$ are the means and standard deviations respectively. Correlation is a measure of the gray level linear dependence between the pixels at the specified positions relative to each other. A value of 0 implies that the pattern is uncorrelated, 1 implies perfect correlation and -1 implies that the spatial set exhibits a dissimilar, deterministic structure.

iv. *Homogeneity*

$$Correlation = \sum_i \sum_j \frac{P(i,j)}{1 + |i - j|} \qquad (4)$$

Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. In the homogeneity measure, the weight values are the inverses of the contrast weight values. As we move further away from the diagonal, the weights decreases quadratically.

c) *Background Detection using GLCM*

The background is segmented by selecting a single texture feature of 60x60 pixels of each scene tested. The background texture feature was manually selected to demonstrate the method. A single texture was selected to show the viability that using a single texture can still provide good results. In the future work, a number of textures of background scenes will be selected and trained offline and used to classify any scene or multiple random textures will be selected during the UAV flight and the average of the textures will be used as the texture template.

Fig. 2: Two typical scenes (Zebra [Equus quagga] scene on the left and Blesbok [Damaliscus pygargus phillipsi] scene on the right) of the environments of the datasets and the respective textures selected

A number of texture measures were tested but the ones that appeared to give the best results for the background were energy, contrast, correlation and homogeneity. The distance measure for the comparison to the image window was calculated as

$$(\mid Energy_{template} - Energy_{window} \mid < \varepsilon_1) \wedge (\mid Corr_{template} - Corr_{window} \mid <$$
$$\varepsilon_2) \wedge (\mid Hom_{template} - Hom_{window} \mid < \varepsilon_3) \wedge (\mid Cont_{template} - Cont_{window} \mid < \varepsilon_4) \qquad (5)$$

Figure 3 shows an example of the original output frame of the zebra video sequence and the frame after background estimation. While some artefacts or false detections exist, these will be reduced in the following steps. It can also be seen that parts of the scene in the top of the image were not segmented as well as some of the trees. This will be solved once more scene features are added and trained but it will also be

seen in the following steps that this undetected background scenes will not severely affect the final outcome. None of the zebras in the scene were selected as being part of the background. Figure 4 shows similar results but for the blesbok output sequence.
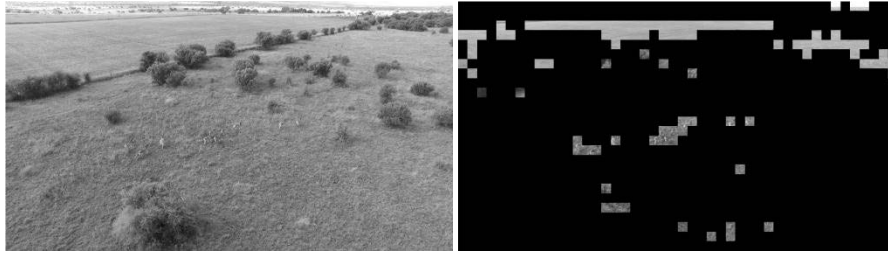

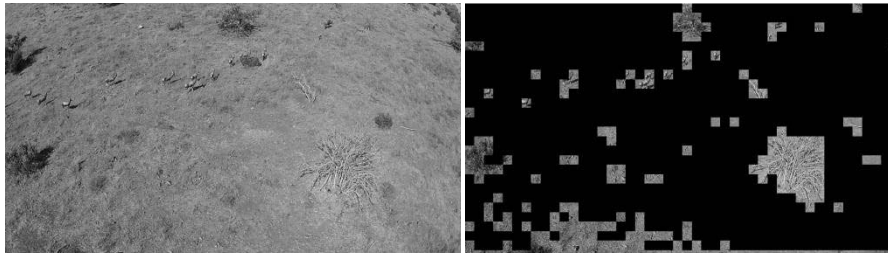*Fig. 3:* (Left) Zebra scene frame and (Right) Background area segmented out

Fig. 4: (Left) Blesbok scene frame and (Right) Background area segmented out

d) Animal Detection using GLCM

The animals are detected by using the 60x60 texture pattern selected for the zebra and 60x40 for the blesbok. The distance measure is calculated in the same manner as (5). The texture patterns selected for the zebra and blesbok are shown in figure 5.
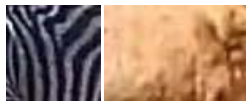

Fig. 5: (Left) Zebra texture and (Right) Blesbok texture

Figure 6 shows the foreground texture segmentation applied to frames from the data-set. It can be seen that all the zebras have been selected as being part of the fore-ground but there remains a lot of artefacts. In the left image it can be seen that a single blesbok has not been correctly segmented. The artefacts could be reduced by adjusting the parameters but this causes some of the zebras to not be selected. Since the aim is to reduce the search area, it is imperative that as many of the zebras and blesbok as possible are selected at this stage.
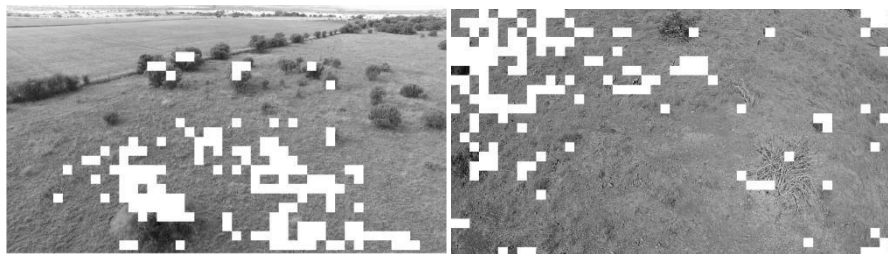

Fig. 6: (Left) the zebras are detected using the zebra texture and (right) Blesbok detected using blesbok texture

e) Intersection of the Background and Animal Detection to Improve Results

The intersection of the two results, the foreground detection of the animals and the background detection of the environment, is now performed to reduce the number of artefacts present in the images. Figure 8 and 10 shows the final result in which the inverse of the background segmentation is intersected with the foreground segmentation to provide the final output. While a few artefacts still remain, the animals have been largely segmented from the background.

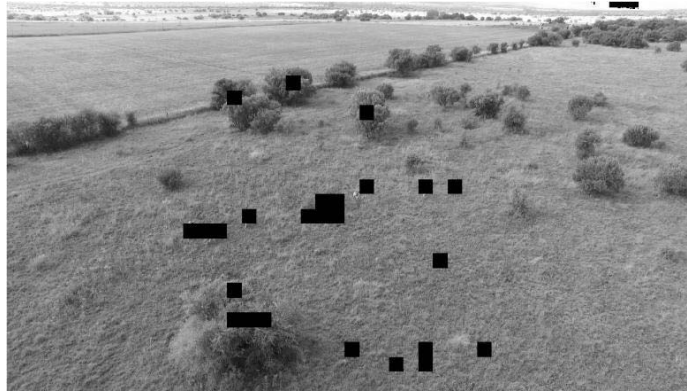*Fig. 7:* Single frame from zebra scene video

Fig. 8: Final segmented image of figure 7

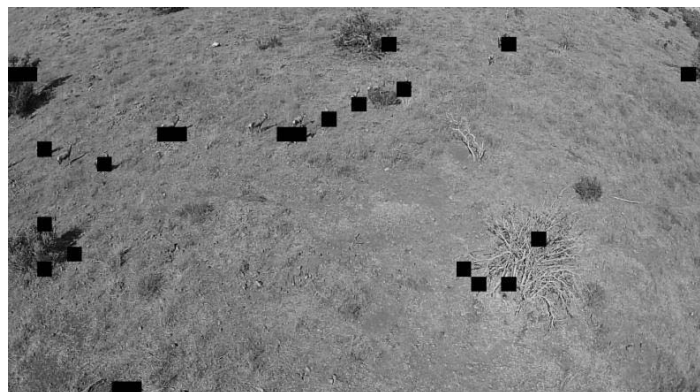*Fig. 9:* Single frame from blesbok scene video

Fig. 10: Final segmented image of figure 9

IV. RESULTS

The results were collected from three video scenes. The first video scene contains zebras and the second and third video scenes contain blesbok (one from a high altitude and another from a lower altitude).
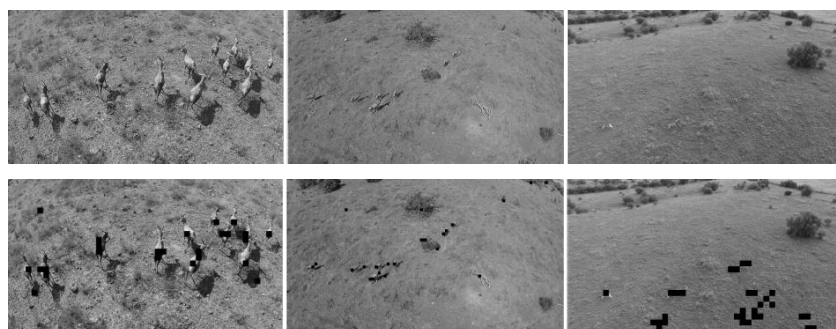
The false negative is more important for the project as the false positives will be reduced during further detection and identification as well as the training of the textures. The objective was to segment areas of the image containing animals and not to detect animals.

Fig. 11: (Top) Various frames from the datasets and (bottom) the final results

Table 1: Results of segmentation

Species	Total Frames	Ground Truth	Positive detection of animals	False negative
Zebra	400	3100	76%	23%
Blesbok(near)	200	2600	84%	16%
Blesbok (Far)	400	5200	81%	19%

V. CONCLUSION AND FURTHER WORK

A GLCM texture segmentation method was presented to segment animals from the background in a natural environment to reduce the search area for animals in the aerial videos. The method presented has shown promising initial results when tested on datasets collected in real South African environment. The method presented is well suited to implementation on a UAV as it is easily parallelizable algorithmically. Future work will involve using Bayesian classifier to train multiple scenes of background textures such as rocks, tress and various types of surfaces encountered. Various animal textures will also be trained and video sequences will be classified into two categories, animal or background. Once an animal is detected it will be tracked to eliminate the need for further detections in frames. Real-time capabilities will be investigated.

REFERENCES RÉFÉRENCES REFERENCIAS

1. B. Sirmacek, M. Wegmann, A. Cross, J. Hopcraft, P. Reinartz and S. Dech: Automatic population counts for improved wildlife management using aerial photography. *iEMSs*, 2012.
2. J. Gemert, C. Verschoor, P. Mettes, K. Epema, L. Koh and S. Wich: Nature conservation drones for automatic localization and counting of animals. ECCV, 2014.
3. C. Stauffer and W.E.L. Grimson. "Adaptive background mixture models for real-time tracking," *Proc. CVPR*. 1999.
4. J. Zhong and S, Sclaroff, "Segmenting Foreground Objects from a Dynamic Textured Background Via a Robust Kalman Filter", Ninth IEEE International Conference on Computer Vision, 2003.
5. C. Ridder, O. Munkelt, and H. Kirchner. "Adaptive background estimation and foreground detection using Kalman filtering," Pro. Int. Conf. on Recent Advances in Mechatronics, 1995.
6. Bhat, K. S., Saptharishi, M., Khosla, P. K, "Motion detection and segmentation using image mosics," *IEEE ICME*, 2000.
7. V. Reilly, H, Idrees and M. Shah, "Detection and tracking of large number of targets in wide area surveillance," Computer Vision ECCV, pp.186-199, 2010.
8. Thakoor, N., Gao, J., "Automatic object detection in video sequences with camera in motion," 2004.
9. Y. Wang, Z. Zhang and Y. Wang, Moving Object Detection in Aerial Video, 11th International Conference on Machine Learning and Applications, 2012, pp: 446-450.