2017 IEEE 3rd International Conference on Electro-Technology for National Development (NIGERCON)

Human Detection Using Speeded-Up Robust Features and Support Vector Machine from Aerial Images

Buhari U. Umar
Computer Engineering Department
Federal University of Technology
Minna, Nigeria
buhariumar@futminna.edu.ng

James Agajo
Computer Engineering Department
Federal University of Technology
Minna, Nigeria
james.agajo@futminna.edu.ng

Ahmed Aliyu
Computer Engineering Department
Federal University of Technology
Minna, Nigeria
aliyu.ahmed@futminnna.edu.ng

Jonathan G. Kolo
Computer Engineering Department
Federal University of Technology
Minna, Nigeria
jgkolo@futminna.edu.ng

Olayemi M. Olaniyi
Computer Engineering Department
Federal University of Technology
Minna, Nigeria
mikail.olaniyi@futminna.edu.ng

Olakunle S. Owolabi
Computer Engineering Department
Federal University of Technology
Minna, Nigeria
samk4christ@gmail.com

Abstract— **Human detection from an aerial image has attracted wide attention due to its vast area of application such as in surveillance, search and rescue operation, and for visual understanding of the image. Unlike object detection, human detection from an aerial image is a challenging classification problem because of different posture appearance of human in an image. More so, at high altitude human shape appear deformed. Different features selection and different algorithm have been proposed. Although effective, but limited due to, characteristic of human posture in an image. In order to address this problem, this research proposed a Speeded-Up Robust feature selection and SVM for human detection from an aerial image due to computational speed and robustness of the SURF feature. This approach would help in better human detection from aerial images irrespective of position and movement for either rescue or surveillance mission. Aerial images were acquired preprocess and segmented using Otsu segmentation. A database comprises of two hundred images was created; 70 percent (140 images) of it was used in training the classifier and 30 percent (60 images) for testing the classifier. Accuracy of 50%, specificity of 57.1%, sensitivity of 46.2% and precision of 66.7% was achieved. These results can be used for a better human detection from an aerial image irrespective of the position or movement.**

Keywords— Human Detection, SURF Feature, SVM, Aerial Images and UAV

I. INTRODUCTION

Over the years, there have been increased use of unmanned aerial vehicle (UAV) for human detection for rescue operation, surveillance, delivery, forest fire monitoring and better understanding of human behaviour from an aerial image [1]. Human detection from an aerial image using unmanned aerial vehicles has attracted so much attention for visual image understanding. Unlike other object detection, human detection has some characteristic and challenges. Human usually have a different posture, appearance in an image [2, 3]. Human detection from an aerial image is a challenging classification task because of the different appearance of human in an image, the background or variation of illumination, also at high altitude, human appear so tiny and deform. These characteristics of human form influence the choice of feature selection and algorithm for human detection or classification [2]. There have been several recognition algorithm and feature extraction for human recognition, such as [4-9] over the years, but they have been no suitable or better algorithm or feature selection for human detection for an aerial image capture using UAV [10].

In order to address the problem stated above, the this research proposed an algorithm for human detection from an aerial image using Speeded-up Robust Feature (SURF) and Support Vector Machine (SVM) for classification. This approach would help in better human detection from an aerial image irrespective of the position and movement, as the SURF feature is a local feature descriptor which is robust, scale invariant and less computational time compared to global features or other types of local features.

The remaining part of this paper is organized as follows: in section one, introduction human detection from an aerial image using UAV was discussed. In section two, feature extraction, human detection, human detection from an aerial image and SVM was discussed. The Methodology of the experiment was discussed in section three. The result of the research was discussed in section four, the conclusion of the

978-1-5090-6422-9/17/$31.00 ©2017 IEEE

research and future recommendation are discussed in section five.

II. PREVIOUS RELATED WORK

This section describes some related works and their limitations that have been reviewed in relation to this research work. Also, in this section are the description of different image processing techniques, features extraction methods, and other human detection approach used in this research work.

A. Features Extraction and Classification

Feature extraction is one of the important fields of image processing and artificial intelligence. Its major function is to extract relevant features from an image and assign it into a label. Analysing the properties of image features and organizing them into classes of numerical features is the most important step. That it is image are classified according to their content [11]. In human detection process, feature selection and extraction is very key because proper feature selection can improve accuracy of the classification, while the improper feature selection may lead to wrong classification or misrepresentation. Therefore feature selection for human detection is very important.

Feature extraction can be classified into two: global feature and local feature. Global features comprise of shape, colour and texture, while local feature comprises of SURF, SIFT, HoG, BRISK FREAK MSER etc. Global feature have been used by researchers for human detection. However, global features are not flexible as compared to a local feature, since human can have different forms or shape [12]. This research makes use of local feature called SURF feature because of it high computational speed and robustness.

- Speeded-Up Robust Feature

SURF is local second order Hessian matrix approximation feature descriptor. It is extracted at the region of interest in an image. It is usually used for feature point detection. The descriptor has been used for to detect the extent of matching between two images or for the detection of a particular objects in an image, in which many object are present. For a particular pattern or object in image, the feature extracted corresponding to them should be similar even if they are extracted at different illumination, scale and noise. SURF is a well-known feature descriptor due to it invariant to scale, illumination and rotation. It has been widely used for image recognition and retrieval, due to its high computational speed and robustness as compared to other types of local feature [13-17].

B. Human Detection

Human detection is a challenging classification problem which has many wild range of applications such as in surveillance, search and rescue operation and security monitoring. Various approaches have been proposed to solve this problem ranging from image matching, human motion detection, and facial recognition to the use of artificial intelligence classifiers and specific feature extraction.

- Types of Human Detection Techniques

With the advancement of image processing different techniques for human detection has evolved. Some these techniques are image matching, motion based human detection, sound based human detection, facial recognition techniques and many others.

- Image Matching

This technique involves acquiring a database of different human posture and runs a matching algorithm of each posture to the input image to get a posture match. When a match is detected, its assumed human is detected. This technique is not very effective when measured in time because if the database comprises of fifty images, each of the fifty images would go through the matching processes. And it is shaped, scale and texture variant that is the shape size, color and texture of each of the image needs to be considered.

- Motion Based Human Detection

This approach could use the image matching or image classification using an intelligent classifier. It involves extracting different human motion postures (walking, jumping, crawling, swimming etc.) stores them in a database and when an input image has given it, extracts the motion of the image and either uses the matching technique or classification technique to determine which human motion is detected in the image. This technique is limited in time and in efficiency, if someone is in a wheel chair, motion would be made, but in a static posture there by limiting this technique.

- Facial Recognition

  The facial recognition algorithm identifies the feature by extracting features or landmark from an image of the object. The algorithm may analyse the shape or size of the eye or nose, relative position, jaw and cheekbones. The feature is used to search for other images with similar features [18].

- Human Detection Using Unmanned Aerial Vehicles

With the growing application range of unmanned aerial vehicles, the need to enhance its result has evolved over time. Different approach to make the output of aerial vehicles useful to man has been developed over time. Image processing and different artificial intelligence techniques have been used to interpret the result images from an unmanned aerial vehicle.

For applications such as security surveillance or search and rescue mission, different image processing approach has been used to extract humans from the imagery of the unmanned vehicle. Different human detection techniques from unmanned aerial vehicles has been developed and is still evolving to attain the perfect result.

C. Support Vector Machine (SVM)

SVM has evolved as a very important tool for solving the problem of pattern classification in a contest to the neural network. SVM do not minimize any artificial error metric, but maximize the margin of linear decision bounding to achieve maximum separation between the object classes. SVM classifier has been used in combination with various features for human detection [3]. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicit mapping their inputs into high-dimensional feature spaces [19].

- Classification of Support Vector Machine

SVM Type 1: For this type of SVM, training involves the minimization of the error function:

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \tag{1}$$

Subject to the constraints:

$$y_i \left(w^T \phi(x_i) + b \right) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, i = 1, \ldots, N \tag{2}$$

Where C is the capacity constant, w is the vector of coefficients, b is a constant, and ξ_i represents parameters for handling non-separable data (inputs). The index I labels the N training cases. Note that $y \in \pm 1$ represents the class labels and xi represents the independent variables. The kernel ϕ is used to transform data from the input (independent) to the feature space. It should be noted that the larger the C, the more the error is penalized. Thus, C should be chosen with care to avoid over fitting.

- SVM Type 2: In contrast to Classification SVM Type 1, the Classification SVM Type 2 model minimizes the error function:

$$\frac{1}{2} w^T w - \nu\rho + \frac{1}{N} \sum_{i=1}^{N} \xi_i \tag{3}$$

Subject to the constraints:

$$y_i \left(w^T \phi(x_i) + b \right) \geq \rho - \xi_i, \xi_i \geq 0, i = 1, \ldots, N \text{ and } \rho \geq 0 \tag{4}$$

D. Review of Some Related works

Research in [20], studied the question of feature sets for robust visual object recognition, adopting linear SVM based human detection as a test case in the paper Histograms of Oriented Gradients for Human Detection. After reviewing existing edge and gradient based descriptors, they proposed experiments that grids of Histograms of Oriented Gradient (HOG) descriptors significantly outperform some existing feature sets for human detection. The paper only focuses on feature selection.

Approach [21], Histogram of Oriented Gradient was used for Person Detection from Aerial Images trained the Support Vector Machine (SVM) using 120 images gotten from MF Media house aerial videos based on their HOG features and tested with 100 images, 50 false and 50 true. 50% accuracy was achieved.

Research in [4], presented improvements to shape and heat flow-based technique of detection and classification of humans in unrestricted poses with the addition of image fusion. He focuses on both rural and urban environments and demonstrate the effectiveness of using image fusion as a preprocessing procedure for improved human detection and classification. Extensive simulations using MWIR images were conducted and results were obtained. Little improvement was achieved.

In [22], a new human detection scheme for thermal images by using Census Transform histogram (CENTRIST) features and Support Vector Machines (SVMs) was presented. Due to low image resolution, thermal noising, lack of colour, and poor texture information human detection in thermal images, the result was not significant.

Studied in [23], Counted people in the crowd using a generic head detector by investigating the estimate of people in the crowd scene using the head region since it is the most visible part of the body in a crowded scene. The head detector was based on the state-of-art cascade of boosted integral features. To prune the search region, they proposed a novel interest point detector based on gradient orientation feature to locate regions similar to the top of head region from gray level images. Two different background subtraction methods were evaluated to further reduce the search region. This approach was evaluated on PETS 2012 and Turin metro station databases. Experiments on these databases showed good performance of the method for crowd counting. The result was only significant for crowd monitoring.

Also [24], developed a real-time human detection system by integrating the research in [25] and Histogram of Oriented

Gradients (HOG) features. By substituting the Haar features with the HOG features, the system keeps the speed advantage of Viola and Jones' object detection framework, as well as the discriminating power of HOG features on human detection. Experiments demonstrated that the system achieved a better accuracy at nearly the same speed as the original Haar features for human detection.

From the review, it shows that using SURF features with a good classifier can detect the presence of human from an aerial image.

III. Proposed Methodology

To detect human from aerial images requires a range of precise steps, from image acquisition to finally displaying the detected image. The entire process is divided into the following major functions: Image acquisition; Image pre-processing; Image segmentation; Feature extraction; SVM training; and SVM classification. This section describes the approach adopted towards the design and development of human detection based for aerial images. Figure 3.1 gives the system overview.

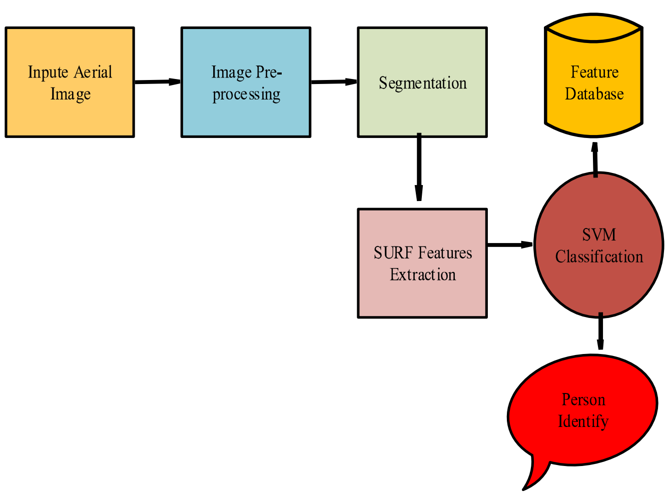

Fig. 1. Overall System Block Diagram

A. Image Pre-processing

All images used in this research were pre-processed, that is enhanced to balance up the intensity of each pixel and to remove blurs and noise. The average filter was used to enhance the intensity of each pixel. The average filter is computed by obtaining a mask matrix that computes the average of each pixels based on each neighbor pixel. The mask matrix determines the number of neighbors to be considered. The histogram equalization technique was used for image sharpening (blur removal).

B. Image Segmentation

Image segmentation means extracting given interest points of an image. There are different techniques or methods in achieving image segmentation. The technique approached in this research is the image thresholding technique. The OTSU segmenting algorithms were employed.

C. Feature Extraction

Extracting the features of images is one of the core and compulsory process in this work. Without the features, the classifier cannot be trained and the aim of this project would not be achieved. The features detected from each image both training data set and testing data set where the Speeded-Up Robust Features. This was chosen due to its high computational speed and robustness. For each image a matrix of values was generated as the detected features. These mean of each feature matrix was obtained and stored to train and to test the classifier.

- Algorithm for Feature Detection and Extraction

These features were detected and extracted using MATLAB defined syntax.
Input: Pre-processed Image.
Output: Mean of the SURF features
Start:
Step-1: Read images from a file.
Step-2: detect the SUR-Features.
Step-3: Calculate the mean of SURF matrix
Stop

D. Support Vector Machine Training and Classification.

The training process consists of inputs, target images and the result of the classifier. The mean of the SUR-Features was used to train the Support Vector Machine. Each feature matrix represents different possible postures a human can be in an image. A total of 140 images was used for to train the classifier. 70 different postures of humans and 70 different non-humans' images, including trees, animals, cars and buildings. An image is represented as a set of matrixes, it could range from 2 by 2 matrix of 1000 by 1000 matrices depending on the quality of the image acquisition device or camera. The extracted features of each image give an output of 106 by 64 matrix, which is large for a single image, so they mean of these features was calculated which gave a single column vector output. The mean matrix saved and it represents a single image, this process was repeatedly carried out for the 140 training images.

- SVM Training

The supervised learning method was used in training the classifier, the process involved obtaining the mean of the human images and labelled human and mean of the non-human images labelled as non-human. The svmStruct in MATLAB was used to input the mean data and their classes, while the svmTrain command was used to initialize the SVM training protocol. The mean of the features of each image was documented in an excel sheet and their classes. The xlsread command was used to read in the features and their classes into the svmStruct function, these data serve as the basics for the classification.

- Performance Evaluation

The proposed research performance was evaluated using accuracy, sensitivity, precision, and specificity. Sensitivity, measured how well a classifier can recognize positive cases. That is a human image of non-human, while specificity measure non-human image. That the negative cases. The accuracy measured the sensitivity and specificity. The precision measured the similarity as compared to the training information. They are represented mathematically using equation 5, 6, 7 and 8.

$$\text{Sensitivity} = TP \div (TP + TN) \quad (5)$$

$$\text{Specificity} = TN \div (TN + FP) \quad (6)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

Where:

TP (True Positive): means correctly classified as positive cases

TN (True Negative): means correctly classified negative cases

FP (False Positive): means incorrectly classified negative cases

FN (False Negative): means incorrectly classified positive cases

E. System Implementation

Human detection in aerial image algorithm is implemented in MATLAB using the following procedures.

Start

Step-1: Input image

Step-2: Enhance image

Step-3: Segment image

Step-4: detect SUR-Features

Step-5: compute mean of features detected

Step-6: repeat step-1 to 4 for all training data set.

Step-7: Train SVM

Step-8: Test SVM

Step-9: compute system performance using equation 5, 6, 7 and 8.

IV RESULTS AND DISCUSSION

This section presents the results obtained from the various stages of detecting humans from aerial images using SUR-Features and SVM classifier. The images were acquired and pre-processed, the features were extracted and their mean values calculated. The feature extracted after resizing an image was 86 by 64 matrix for each image. The mean of the extracted features was used in training the SVM. The table 2 presents the mean of the SUR-Features for human postures and non-human postures. The values in table 2 were used as the input training data for the SVM classifier. From the table 2, the mean values of human are clearly different from that of non-human. This shows that, using the SURF features and SVM classifier can detect the presence of human from an aerial image.

A. Result of Image Segmentation

Fig. 2 presents the result of some of the images segmented. The first top three images are human images after segmentation and the three last images are images of non-humans after segmentation. The non-human images comprise of a tree, a building and a car.

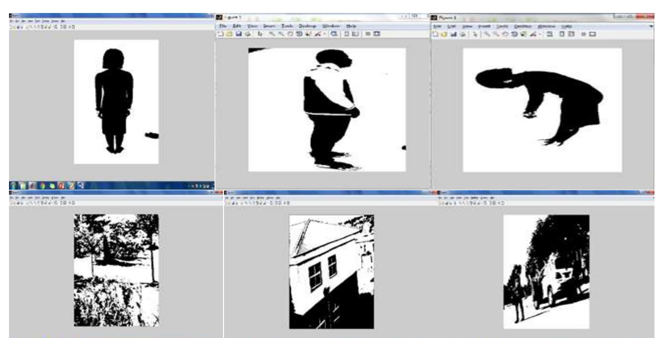

Fig. 2. Results of Image Segmentation

B. SVM Training and Testing Results

The method applied in training the SVM is supervised learning. In supervised learning both training data and classes are used in training the classifier given by the user. Table 2 presents samples of the training inputs; the data extracted from the human postures were classified as human, while those from the non-human images were classified as non-person. The entire testing data was 60 images.

C. Graphical User Interface

In order to make the developed algorithm user friendly, a graphical user interface was developed using MATLAB. Fig. 3 and Fig.4 presents the graphical user interface of the developed algorithm when a human is detected and when no human is detected.

Fig. 3. When Human is Detected

Fig. 4. When Non-Human is Detected

The result obtained from the classifier was evaluated using TP, FN, TN, and FP. True positive is the correctly classified human case of an image as human, while true negative is when non-human is classified as non-human. A false positive is when non-human is classified as human, while false negative is when human is classified as human. The accuracy, sensitivity, precision and specificity were computed using the TP, FN, TN and TP. Table 1, present the summary of the result.

Table 1 Evaluation Table.

Performance Metrics	**SVM Classifier**
True positive (TP)	12
False negative (FN)	14
True negative (TN)	8
False positive (FP)	6

Table 2 Mean of SUR-Features of Human and Non-Human Image

S/N	MEAN		S/N	MEAN		S/N	MEAN		S/N	MEAN	
	HUMAN	NON-HUMAN		HUMAN	NON-HUMAN		HUMAN	NON-HUMAN		HUMAN	NON-HUMAN
1	0.044988	0.005088	21	0.005088	0.003475	41	0.003475	0.025473	61	0.005088	0.022173
2	0.047977	0.005688	22	0.005688	0.003704	42	0.003704	0.010874	62	0.005688	0.026598
3	0.046801	0.004752	23	0.004752	0.004283	43	0.004283	0.026953	63	0.004752	0.031677
4	0.04357	0.006214	24	0.006214	0.004184	44	0.004184	0.021126	64	0.006214	0.027447
5	0.046335	0.005198	25	0.005198	0.003167	45	0.003167	0.013994	65	0.005198	0.021951
6	0.046198	0.006614	26	0.006614	0.00343	46	0.00343	0.012645	66	0.006614	0.025239
7	0.046198	0.008526	27	0.008526	0.005121	47	0.005121	0.014717	67	0.008526	0.029773
8	0.046904	0.005926	28	0.005926	0.003737	48	0.003737	0.011021	68	0.005926	0.025354
9	0.045326	0.008526	29	0.008526	0.005121	49	0.005121	0.014717	69	0.008526	0.029773
10	0.046664	0.002046	30	0.002046	0.002195	50	0.002195	0.033698	70	0.002046	0.013857
11	0.046664	0.00241	31	0.00241	0.002291	51	0.002291	0.026391	71	0.00241	0.01797
12	0.046735	0.00241	32	0.00241	0.002291	52	0.002291	0.026391	72	0.00241	0.01797
13	0.047426	0.001477	33	0.001477	0.001459	53	0.001459	0.030624	73	0.001477	0.018355
14	0.048449	0.002947	34	0.002947	0.001552	54	0.001552	0.021229	74	0.002947	0.014569
15	0.046579	0.002939	35	0.002939	0.002715	55	0.002715	0.014079	75	0.002939	0.025189
16	0.047364	0.003853	36	0.003853	0.003386	56	0.003386	0.029795	76	0.003853	0.026751
17	0.048394	0.003971	37	0.003971	0.003564	57	0.003564	0.011547	77	0.003971	0.028978
18	0.046088	0.004928	38	0.004928	0.003652	58	0.003652	0.013956	78	0.004928	0.029553
19	0.044988	0.004453	39	0.004453	0.00386	59	0.00386	0.016009	79	0.004453	0.030947
20	0.046088	0.002994	40	0.002994	0.001635	60	0.001635	0.018595	80	0.002994	0.014992

Table 3 Result of overall Classification

S/N	SVM RESULT		EXPECTED RESULTS	S/N	SVM RESULT		EXPECTED RESULTS
1	0.00411	non person	Non person	21	0.050304	non person	person
2	0.004709	non person	Non person	22	0.050291	Non person	non person
3	0.006587	non person	Non person	23	0.04274	Non person	non person
4	0.00478	non person	Non person	24	0.046726	Non person	non person
5	0.004757	person	Person	25	0.041921	Non person	Person
6	0.004136	person	Non person	26	0.048396	person	Person
7	0.006597	person	Person	27	0.041953	Non person	Person
8	0.004309	person	Non-person	28	0.029928	person	Person
9	0.005182	person	Non-person	29	0.032168	Non person	Person
10	0.004041	person	Person	30	0.038191	Non person	Person
11	0.004192	non-person	Person	31	0.03238	person	non person
12	0.003748	person	Person	32	0.030202	person	Person
13	0.00464	non-person	Person	33	0.026664	person	non person
14	0.003606	non-person	Person	34	0.026664	Person	non person
15	0.039772	non-person	Person	35	0.025019	Person	non person
16	0.042776	person	Person	36	0.032413	Person	Person
17	0.045918	person	Person	37	0.029795	Person	Person
18	0.044889	non-person	Person	38	0.011547	Person	non person
19	0.04109	non-person	Person	39	0.013956	Person	non person
20	0.050304	non-person	Person	40	0.029928	Person	non person

Table 4. Testing Image SUR-Features

S/N	MEAN OF SUR-FEATURES	S/N	MEAN OF SUR-FEATURES	S/N	MEAN OF SUR-FEATURES
1	0.00411	15	0.039772	29	0.029928
2	0.004709	16	0.042776	30	0.032168
3	0.006587	17	0.045918	31	0.038191
4	0.00478	18	0.044889	32	0.03238
5	0.004757	19	0.04109	33	0.030202
6	0.004136	20	0.050304	34	0.026664
7	0.006597	21	0.050304	35	0.026664
8	0.004309	22	0.050291	36	0.025019
9	0.005182	23	0.04274	37	0.032413
10	0.004041	24	0.046726	38	0.029795
11	0.004192	25	0.041921	39	0.011547
12	0.003748	26	0.048396	40	0.013956
13	0.00464	27	0.041953		

14	0.003606	28	0.041953		

The values in table 1 were used to compute the performance metrics for the evaluation of the system. The values of the sensitivity, specificity, precision and accuracy is given below in equations 9, 10, 11 and 12.

$$\text{Sensitivity}\left(\frac{\text{[illegible]}}{\text{[illegible]}}\right) = 46.2\% \quad (9)$$

$$\text{Specificity}\left(\frac{\text{[illegible]}}{\text{[illegible]}}\right) = 57.1\% \quad (10)$$

$$\text{Precision}\left(\frac{\text{[illegible]}}{\text{[illegible]}}\right) = 66.7\% \quad (11)$$

$$\text{Accuracy}\left(\frac{\text{[illegible]}}{\text{[illegible]}}\right) = 50\% \quad (12)$$

The classifier was randomly tested with forty images, twenty human images and twenty non-human images. Table 1 summarizes the values of the result. The true positive cases imply the number of human images classified as humans, twelve of these images were classified as human images. The true negative implies human images classified as non-humans, a total of fourteen images was truly identified as non-human images. The false positive represents non-human images classified as humans a total of eight images was classified wrongly as humans. The false negative represents non-human images classifies as humans a total of six images fell in this category. From table 2, the mean of the features of human images and non-human images were distinctly different from each other, although, there were some values that were approximately the same, this represents why the level of accuracy and precision.

V. Conclusion

In this research, we proposed human detection from an aerial image using the SURF feature and SVM for human detection. Human detection is a challenging classification problem, unlike object detection, human detection from an aerial image has different posture appearance of humans. More so, at high altitude human shape appear deformed. Different features selection and different algorithm have been proposed. Although effective, but limited due to, characteristic of human posture in an image. The objective of this paper is to develop an algorithm for better human detection form an aerial image. To achieved this, an aerial images were acquired preprocess and segmented using Otsu segmentation. A database comprises of two hundred images was created; 70 percent (140 images) of it was used in training the classifier and 30 percent (60 images) for testing the classifier. Accuracy of 50%, specificity of 57.1%, sensitivity of 46.2% and precision of 66.7% was achieved. These results can be used for a better human detection from an aerial image irrespective of the position or movement. The objective of the research was achieved as desired. For future improvement, the following area approach can be considered:

- Combination of different local features and global feature can be considered.
- Hybridization of detection algorithm can also be considered.

Acknowledgment

We would like to thank Computer Engineering Department, Federal University of Technology of Minna for providing us with the MATLAB software and space in the laboratory to actualize the research objective.

References

[1] F. Gökçe, G. Üçoluk, E. Şahin, and S. Kalkan, "Vision-based detection and distance estimation of micro unmanned aerial vehicles," *Sensors,* vol. 15, pp. 23805-23846, 2015.

[2] H. Beiping and Z. Wen, "Fast Human Detection Using Motion Detection and Histogram of Oriented Gradients," *JCP,* vol. 6, pp. 1597-1604, 2011.

[3] X. Chao-jian and G. San-xue, "Image target identification of UAV based on SIFT," *Procedia Engineering,* vol. 15, pp. 3205-3209, 2011.

[4] E. T. Gilmore, P. D. Frazier, and M. Chouikha, "Improved human detection using image fusion," in *Proceedings of the IEEE ICRA 2009 Workshop on People Detection and Tracking, Kobe, Japan*, 2009.

[5] A. Gaszczak, T. P. Breckon, and J. Han, "Real-time people and vehicle detection from UAV imagery," 2011.

[6] R. Montanari, D. C. Tozadore, E. S. Fraccaroli, and R. A. Romero, "Ground vehicle detection and classification by an unmanned aerial vehicle," in *Robotics Symposium (LARS) and 2015 3rd Brazilian Symposium on Robotics (LARS-SBR), 2015 12th Latin American*, 2015, pp. 253-258.

[7] M. Laroze, L. Courtrai, and S. Lefèvre, "Human Detection from Aerial Imagery for Automatic Counting of Shellfish Gatherers," in *VISIGRAPP (4: VISAPP)*, 2016, pp. 664-671.

[8] N. Ammour, H. Alhichri, Y. Bazi, B. Benjdira, N. Alajlan, and M. Zuair, "Deep Learning Approach for Car Detection in UAV Imagery," *Remote Sensing,* vol. 9, p. 312, 2017.

[9] O. Meynberg, S. Cui, and P. Reinartz, "Detection of high-density crowds in aerial images using texture classification," *Remote Sensing,* vol. 8, p. 470, 2016.

[10] M. Barekatain, M. Martí, H.-F. Shih, S. Murray, K. Nakayama, Y. Matsuo*, et al.*, "Okutama-Action: An Aerial View Video Dataset for Concurrent Human

Action Detection," *arXiv preprint arXiv:1706.03038,* 2017.

[11] S. A. Medjahed, "A Comparative Study of Feature Extraction Methods in Images Classification," *International Journal of Image, Graphics and Signal Processing,* vol. 7, p. 16, 2015.

[12] V. S. E.Komagal , K.Anand and C.P.Anand raj, "Human Detection in Hours of darkness Using Gaussian Mixture model Algorithm" *International Journal of Information Sciences and Techniques (IJIST),* vol. 4, pp. 83 - 89, 2014.

[13] N. K. Verma, A. Goyal, A. H. Vardhan, R. K. Sevakula, and A. Salour, "Object Matching Using Speeded Up Robust Features," in *Intelligent and Evolutionary Systems*, ed: Springer, 2016, pp. 415-427.

[14] M. Radovic, O. Adarkwa, and Q. Wang, "Object Recognition in Aerial Images Using Convolutional Neural Networks," *Journal of Imaging,* vol. 3, p. 21, 2017.

[15] B. Anand and M. P. K. Shah, "Face Recognition using SURF Features and SVM Classifier," *International Journal of Electronics Engineering Research,* vol. 8, pp. 1-8, 2016.

[16] A. A. Fathima, R. Karthik, and V. Vaidehi, "Image stitching with combined moment invariants and SIFT features," *Procedia Computer Science,* vol. 19, pp. 420-427, 2013.

[17] H. Joshi and M. K. Sinha, "A survey on image mosaicing techniques," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),* vol. 2, pp. pp: 365-369, 2013.

[18] B. Scholkopf, K.-K. Sung, C. J. Burges, F. Girosi, P. Niyogi, T. Poggio*, et al.*, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," *IEEE transactions on Signal Processing,* vol. 45, pp. 2758-2765, 1997.

[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 886-893.

[20] A. M. Garcia, M. A. Rufino, L. C. Sangalang, J. A. Teodoro, and J. Ilao, "Application of Histogram of Oriented Gradient in Person Detection from Aerial Images," 2014.

[21] I. Riaz, J. Piao, and H. Shin, "Human detection by using centrist features for thermal images," in *International Conference Computer Graphics, Visualization, Computer Vision and Image Processing*, 2013.

[22] V. B. Subburaman, A. Descamps, and C. Carincotte, "Counting people in the crowd using a generic head detector," in *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, 2012, pp. 470-475.

[23] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE transactions on pattern analysis and machine intelligence,* vol. 32, pp. 1627-1645, 2010.

[24] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. I-I.