2020 16th IEEE International Colloquium on Signal Processing & its Applications (CSPA 2020), 28-29 Feb. 2020, Langkawi, Malaysia

Detection of Human Bodies in Lying Position based on Aggregate Channel Features

Mohammad Aidil Shah Sajat, Habibah Hashim and Nooritawati Md Tahir
Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM)
40450, Shah Alam, Selangor, Malaysia.
Corresponding authors: m.aidilshah@gmail.com, habib350@uitm.edu.my, nooritawati@ieee.org

***Abstract*— In recent years, detection of human body has drawn a lot of attention from researchers in the field of image recognition, with most work focused on pedestrian detection. The detection of human bodies in lying position also received numerous attention in applications such as elderly fall detection, sleep studies as well as in search and rescue (SAR) operations. Thus, in this paper, feature extraction performed by the Aggregate Channel Features (ACF) algorithm is explored for detection of human bodies in lying positions. ACF makes use of a Boosted Decision Tree (BDT) classifier that has resulted in increase in speed of detection. The classification was carried out using a dataset developed from aerial images of human bodies obtained from the internet. Initial result showed that the accuracy of ACF using the given dataset is 88% and the value of F-measure obtained was 0.9231. This proposed method will be further explored on a more advanced dataset.**

Keywords — Aggregate Channel Features; human body detection; lying position; object detection; search and rescue

I. INTRODUCTION

The detection of human body in lying position has a very great impact in aerial detection of victims in search and rescue (SAR) operations. While significant progress has been made in the development of ground robots for SAR operations, most of these robots still lack the mobility needed to explore disaster sites autonomously. Unmanned aerial vehicles (UAVs) could provide visual information in real-time mode specifically for SAR, for instance data related to aftermath due to earthquake or hurricane. In addition, UAV can be remotely managed via radio however recently new control system has been developed that is able to maintain the stability of UAVs along the proposed route [1]. The catastrophic site can be quickly explored from the air to identify people who need help or to locate the deceased [2]. Having on-board systems on UAVs that are able to detect and specifically locate victims of mass disasters can help recover the injured quickly and in a forensic perspective, help to bring some of the early processes of victim identification by cutting down on the delay of retrieving bodies prevent difficulties in identification due to severe decomposition.

In recent years, a number of solutions for human body detection have been proposed. The histograms of oriented gradient detector (HOG) are one of the more popular and effective models for human detection proposed to date [3]. In this model, image gradient histograms are calculated and standardized in a local and overlapping block scheme and concatenated into a single detection window descriptor that is densely scanned in a test image across all scales and locations. Further, the HOG features are extracted from the whole compact grid parts based on region of interest. Next, the combined features are classified using Linear Support Vector Machine (SVM). It is recognized that HOG provides an accurate description of human body contour, however it requires significantly higher computational time. Patwary et.al. [4], proposed that HOG features can be reduced by considering useful features that contain edge information, and by eliminating blocks which have less significance in extracting the HOG features, they managed to reduce the computational time required by conventional HOG.

Moreover as reported in [5, 6] family of channel features showed promising record performances for pedestrian detection. Additionally, based on the original images, channel features computed the registered maps that include gradients oriented gradients histogram followed by feature extraction at these extended channels.

Hence, in this paper, variant of channel features called ACF [6] that directly extracted pixel values on sub-sampled channels is explored. It is well known that channel extension contributes rich representation capacity whilst simple feature form assures fast computation. With these two superiorities, the aggregate channel features is investigated and validated as potential for human body detection.

II. RELATED WORK

Since the early days of computer vision, human body detection has a lot of researcher's attention. Most of the work focused on the pedestrian detection [3], [7]–[16]. There are generally two categories of current methods that have been used for pedestrian detection. First category is known as model-based on hand-crafted features [3], [6], [17]–[19] and deep models [8], [20]–[23].

As reported in [3], conventional methods extraction includes Haar, HOG [3], or HOG-LBP [17] from images to train the SVM [3] or boosting classifiers [8]. The hand-crafted features achieved good accuracy for pedestrian detection. For instance, for partial occlusion of pedestrian detection, LBP and HOG has been utilized as discussed in [17]. Also, algorithms developed based on Integral Channel Features (ICF) [5] and ACF [6] has been proposed that comprised of gradient histogram, gradients and LUV that can be extracted efficiently.

On the other hand, as reported in [22], the deep learning approaches are based on learning raw pixel features to boost pedestrian detection efficiency. For example, The Joint Deep model has developed a hidden layer of deformation to provide details for Convolutional Neural Network (CNN) to model mixture. In contrast to previous deep models that formulated pedestrian detection as a single binary classification task, TA-CNN is jointly optimizing pedestrian detection with related semantic tasks. ConvNet [8] used convolutionary sparse coding for pedestrian detection to unsupervised pretrain CNN.

III. Proposed Object Detection

The ACF has been chosen as the algorithm to detect the human body in lying position based on the investigating the aggregate channel features related to human body detection.

A. Feature description

Channel: Firstly, it is good to understand the ACF feratures. The channel is the basic structure of ACF. Secondly similar channels of ACF is utilized as discussed in [10] namely normalized gradient magnitude, histogram of oriented gradients (six channels), and LUV color channels. Color channels of images are the most common type category with Gray-scale and RGB being typical ones. Additionally, for more challenging situation, several different types of channels were developed for transmitting different types of information. Generally, channels can be defined as a registered map of the original image with the pixels determined from the corresponding original pixel patches. [5]. With linear or non-linear transformation of the original image, different channels can be determined.

Computation: Computation of the aggregate channel features is quite simple. For instance, for one color image all the channels specified are computed with the preset factor sub-sampled. Further, the aggregate pixels in all these sub-sampled channels are vectorized into one pixel look-up table.

B. Feature Design

In general, the feature design is done mainly based channel types, window size and subsampling method.

Channel: Normally, there are three types of channels used namely gradient magnitude, color channel and gradient histograms. The computation of gradient magnitude as well as gradient histogram channels is actually the generalized version of the HoG features. Further, gradient magnitude is the largest response on all three-color channels followed by gradient-oriented histograms and finally the HoG.

Detection window size: The scale to resize both lying and non-lying position body samples is done via detection window size followed by training the detector. Larger window size includes more pixels in feature pool and thus may improve the human body in lying position detection performance. However if the window is too large, some small lying body will be missed detected that could affect the efficiency of detection.

Subsampling: The perceptive scale to monitor the size of the aggregation is known as the subsampling factor. Note that varying the factor from large to small will cause the representation of the features to shift from coarse to fine, thus increase the size of the feature pool. Average pooling is one of the way to perform sub-sampling for original aggregate channel features.

C. Training Design

The ACF object detector training will take 4 stages. At each stage of the training, the number of positive samples remain the same but the number of negative sample and the weak learners keep increasing. The model size used for training is 126x63.

Training data: The dataset used consists of the aerial view image of human body in lying position. The used dataset only contains about 300 pictures. Each image must contain only one lying body. The more image, the better result for the detection. The minimum size for training is 126x63 pixel. The lying body in each image were label or stated using boundary box as the positive sample. The negative sample will be automatically created referring to the given positive sample. All the images are labelled using image labeler application in MATLAB. In ensuring performance of the classifier, suitable data training is required.

Weak learners: The number of weak classifiers contained in the soft-cascade can be varied based on the pool size of the features. If the weak learners increased, this resembled that more classifiers could generate better performance, however if the number increased, the performance begins to saturate. There is a trade-off between accuracy and speed as more classifiers slow down the detection rate. The search for the optimum saturate point is important during training in this context [24].

IV. Methodology

In this section, the training details used for the human body in lying position detector will be described, also explained are the overall framework for the navigation setup.

A. Dataset for the training

The images for the dataset was collected from the internet and was put together as part of this work. It consists of the aerial view images of human body in lying position. The used dataset only contains about 300 images. All the images obtained from the internet. Each image contains only one lying human body. The more images that was included in the dataset, the better result for the human body in lying position detection. All the obtained images from the internet must be resized before it can used as part of the dataset. The minimum size for training image is 126x63 pixels. The size of the image cannot be below 126x63 because the size of the training model that we use is 126x63 pixels. The size also cannot be too large. We use Adobe Photoshop CS6 to resize all the images.

The negative samples that were put aside was randomly extracted from the given dataset. Before the negative sample can be extracted, the positive sample must be labeled using the boundary box method in each of the dataset images. The boundary box will be labelled as *lying*. The reading of each edge of the boundary box will be recorded in a table along with the image path. All the labelling process are done using image labeler application that can be used in the MATLAB. The negative sample must be bigger than the positive sample in order for the program to easily differentiate between the lying bodies and the background image.

B. Model for the training

The model size used for the training is 126x63. The training will take 4 stages. The first stage of the training will start after all the sample of positive examples completely loaded, the computation of the approximation coefficients complete and the computation of the ACF complete.

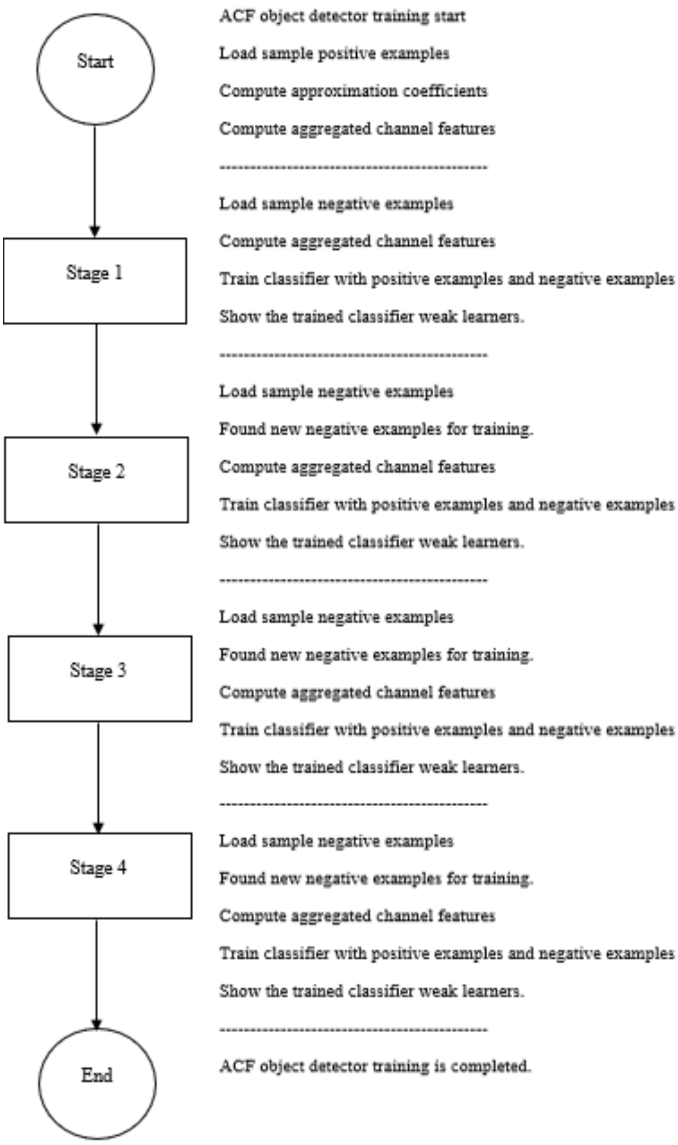

Fig. 1. Flowchart of the ACF object detector training

Figure 1 shows that the flowchart of the ACF object detector training. At each stage of the training, the extraction sample of the negative example will be done. This will enable the program to extract new negative samples for the training. After the negative sampling has been updated, the ACF will recompute. Then, the classifier will be train with the given positive sample and the new negative sample. The number of weak learners will be determined after the classifier train.

C. Implementation

Algorithms for detection of human body during lying position was developed using MATLAB Toolbox (2014, version 3.40) with CPU mode on 2.50 GHz Intel Core i7-4710 HQ, 12 GB of RAM and 64-bitaarchitecture.

V. Result and Discussion

ACF will automatically detect the lying body in the input image. The desired image contains the human lying body can be used to test the system as an input. The system will produce the output which is the label of the human body in lying position in a boundary box. The score of the confidence also will be shown on each of the boundary box.

Fig. 2. Some of the output result of the ACF to detect human bodies in lying position in aerial view

Figure 2 shows that all the lying bodies has been label. The score of the confidence can be seen at each of the boundary box. This shows that the system was able to detect human lying bodies using ACF based on the trained dataset. There are several errors where the system label laying on the same body twice. This will affect the accuracy of the result.

Further performance measures of the classification results for detection of human body during lying and non lying position will be tabulated using confusion matrix that includes numerical analysis of accuracy, precision, recall as well as F-measure. In addition, results of True Positive (TP), True Negative (TP), False Positive (FP) and False Negative (FN) will be detailed in this confusion matrix.

Table 1: Confusion matrix of 50 test images

n = 50	Predicted: No	Predicted: Yes	**Total**
Actual: No	TN = 8	FP = 2	**10**
Actual: Yes	FN = 4	TP = 36	**40**
Total	**12**	**38**	**50**

Table 1 tabulated that the total number of images used as inputs are 50. The result is that the true negative (TN) is 8, false positive (FP) is 2, false negative (FN) is 4 and true positive (TP) is 36. Using the obtained data, the accuracy, recall, precision and f-measure is calculated. The first measure presented is accuracy which is the ratio of the number of correctly predicted observation to the total number of observations i.e. accuracy = (TP + TN)/ (TP + TN + FP + FN). The value of accuracy which refers to our obtained data is 0.88.

Recall on the other hand can be defined as the ratio of the total number of correctly classified positive observations (or TP), divided by the total number of positive examples, where, recall=TP / (TP + FN). In this case, the value of recall calculated using the obtained data is 0.9.

For precision, we divide the total number of correctly classified positive examples by the total number of predicted positive examples, where, precision = TP / (TP + FP). Here, the value of precision calculated using the obtained data is 0.9474.

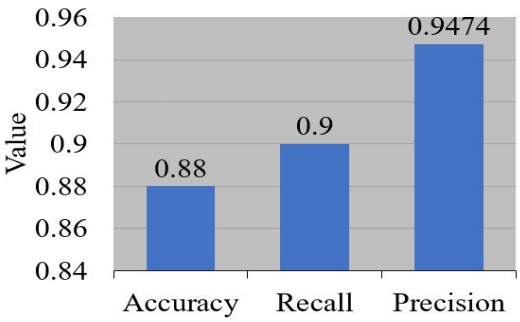

Fig. 3. Graph reading of the accuracy, recall and precision

Figure 3 shows the plotted value of accuracy, recall and precision. Further the value of F-measure that based on Harmonic Mean is computed using equation F-measure = (2*Recall*Precision) / (Recall + Precision). The value of f-measure calculated using the obtained data is 0.9231. The calculated result shows that the accuracy of the ACF using the given dataset is quite high which is 88%. Meanwhile, the value of the f-measure is 0.9231. F-measure uses Harmonic Mean in place of Arithmetic Mean as it punishes the extreme values more.

VI. Conclusion

In conclusion, the effectiveness and suitability of aggregate channel features (ACF) as feature representation for detection of human body during lying and non-lying from aerial-view is implemented. It was found that ACF is indeed apt as feature extraction based on performance measure achieved namely accuracy, recall and precision. Initial results showed that the proposed solution is suitable to implement it to UAV for SAR operations. Note that the dataset used in this study has a total of 300 still images. Next stage of work will proceed with video sequences real time imagery as well as a larger number of image data.

Acknowledgement

The authors would like to thank the Ministry of Education (MoE), Malaysia for funding the research through Grant No: 600-IRMI/TRGS 5/3 (001/2019)-1 and the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia for their support in this research.

References

[1] M. A. Dinesh, S. K. S, S. J. Shetty, K. N. Akarsh, M. G. K. M, and M. Mysore, “Development of an Autonomous Drone for Surveillance Application,” no. June 2018, pp. 331–333, 2018.

[2] S. Bahadori, L. Iocchi, R. La, and V. Salaria, “Human Body Detection in the RoboCup Rescue Scenario,” Informatica, pp. 1–6, 2002.

[3] N. Dalal, B. Triggs, and Cordelia, “To cite this version : Histograms of Oriented Gradients for Human Detection,” pp. 886–893, 2010.

[4] M. J. A. Patwary, S. Parvin, S. Akter , “Significant HOG-Histogram of Oriented Gradient Feature Selection for Human Detection,” International Journal of Computer Applications 132(17):20-24, 2015

[5] P. Dollár, “Integral Channel Features,” pp. 1–11, 2009.

[6] R. Appel, S. Belongie, P. Perona, and P. Doll, “Fast Feature Pyramids for Object Detection,” vol. 36, no. 8, pp. 1532–1545, 2014.

[7] C. Wojek, B. Schiele, and P. Perona, “Pedestrian Detection : An Evaluation of the State of the Art,” vol. 34, no. 4, pp. 743–761, 2012.

[8] P. Sermanet, “Pedestrian Detection with Unsupervised Multi-Stage Feature Learning,” 2013.

[9] X. Chen, P. Wei, W. Ke, Q. Ye, and J. J. B, “Pedestrian Detection with Deep Convolutional Neural Network Pedestrian Detection with Deep Convolutional,” no. November, 2014.

[10] P. Dollár and P. Perona, “The Fastest Pedestrian Detector in the West,” 2010.

[11] D. Ribeiro, P. Miraldo, J. C. Nascimento, and I. De Sistemas, “A Real-Time Deep Learning Pedestrian Detector for Robot Navigation,” pp. 165–171, 2017.

[12] Tian, Yonglong, Ping Luo, Xiaogang Wang, and Xiaoou Tang. "Pedestrian detection aided by deep learning semantic tasks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5079-5087. 2015.

[13] Hosang, Jan, Mohamed Omran, Rodrigo Benenson, and Bernt Schiele. "Taking a deeper look at pedestrians." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4073-4082. 2015.

[14] Zhang, Shanshan, Rodrigo Benenson, and Bernt Schiele. "Filtered channel features for pedestrian detection." In CVPR, vol. 1, no. 2, p. 4. 2015.

[15] Y. Deng, P. Luo, C. C. Loy, and X. Tang, “Pedestrian Attribute Recognition At Far Distance,” pp. 1–4, 2014.

[16] P. Luo and X. Wang, “Pedestrian Parsing via Deep Decompositional Network Shenzhen Institutes of Advanced Technology , Chinese Academy of Sciences,” pp. 2648–2655, 2013.

[17] X. Wang and T. X. Han, “An HOG-LBP Human Detector with Partial Occlusion Handling,” 2009 IEEE 12th Int. Conf. Comput. Vis., no. Iccv, pp. 32–39, 2009.

[18] P. F. Felzenszwalb, I. C. Society, R. B. Girshick, S. Member, D. Mcallester, and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1627–1645, 2010.

[19] Zhang, Shanshan, Christian Bauckhage, and Armin B. Cremers. "Informed haar-like features improve pedestrian detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 947-954. 2014.

[20] Luo, Ping, Yonglong Tian, Xiaogang Wang, and Xiaoou Tang. "Switchable deep network for pedestrian detection." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 899-906. 2014.

[21] W. Ouyang, X. Zeng, and X. Wang, "Modeling Mutual Visibility Relationship in Pedestrian Detection," vol. 1, 2013.

[22] W. Ouyang and X. Wang, "Joint Deep Learning for Pedestrian Detection," 2013.

[23] W. Ouyang and X. Wang, "A Discriminative Deep Model for Pedestrian Detection with Occlusion Handling," 2012 IEEE Conf. Comput. Vis. Pattern Recognit., pp. 3258–3265, 2012. B. Yang and S. Z. Li, "Aggregate Channel Features for Multi-view Face Detection," IEEE Int. Jt. Conf. Biometrics, pp. 1–8, 2014.