EVALUATING SECTOR RING HISTOGRAM OF ORIENTED GRADIENTS FILTER IN LOCATING HUMANS WITHIN UAV IMAGES

M. Ghasemi[1], M. Varshosaz[2], S. Pirasteh[3,*]

[1] Dept. of Photogrammetry, Faculty of Geomatics Eng., K. N. Toosi University of Technology, Iran, mrzghasemi@email.kntu.ac.ir
[2] Dept. of Photogrammetry, Faculty of Geomatics Eng., K. N. Toosi University of Technology, Iran, varshosazm@kntu.ac.ir
[3] Department of Surveying and Geoinformatics, Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong, China, sapirasteh@swjtu.edu.cn

KEY WORDS: UAV, Human Detection, SRHOG, Evaluation

ABSTRACT:

Developing systems to find injured people quickly after natural disasters is an important topic. In recent years, special attention has been paid to the use of UAV images for this purpose. In this regard, an accurate and strong feature is required. It is shown that the Sector Ring Histogram of Oriented Gradients, is a feature very much independent from rotation and scale. The aim of this paper is to evaluate the performance of a human detection algorithm which is based on this strong feature. Experiments carried out suggest that using SRHOG feature humans can be detected with an accuracy of 73.69%. However, despite giving good accuracy, SRHOG results contain more than 33.33 % false labels.

1. INTROUCTION

The use of unmanned aerial vehicles (UAVs) in various subject areas and applications has increased dramatically in recent years. It is of great interest to users and scientists to develop systems for the rapid identification of injured persons following natural disasters in unmanned aerial vehicles (UAV) images (Jingxuan et al., 2016). Unfortunately, when UAV images are used for human sensing with other challenges, the use of a UAV as an image platform introduces certain other problems (Blondel, 2013). Because pictures are taken from above, individuals can be quite different from those on the ground. Therefore, the injured person may be partially covered with snow, rock and the like (Liu, 2017).

Work on human detection using photos has concentrated mostly on pedestrian detection (Mihçioğlu et al., 2019, ZhangSr et at., 2019, Karg et al.,2020), study of human movement (Bahri et al., 2019, Zhang et al., 2019) and facial recognition (Zhang et al. , 2019, Ding et al., 2019, Prasad et al., 2020). Benenson, et.al in 2014 compared over 40+ methods and concluded that the main challenge ahead seems to develop a deeper understanding of what makes good features good, so as to enable the design of even better ones (Benenson et al, 2014). Therefore, the principal task is to identify a feature that can define the presence of the human body. Displaying various features, the data such as texture (Ojala et al., 2002, Leibe et al., 2005), colour (Ott et al., 2009, Walk et al., 2010) and edge (Nguyen et al., 2009) is often removed. For instance, Leibe, et al. uses the texture information to identify pedestrians in a crowded scene (Leibe et al., 2005). For this, so called Haar Wavelets (Dollár et al., 2008) are used which are gray level patterns computed based on the magnitude of the difference between neighboring pixel intensities. Another example is Color self-similarity (CSS) feature (Walk et al., 2010) which is defined using the histogram of color tones present in different parts of an image. In general, techniques that are based on texture or color features highly depend on the pixel values and, thus, the image background may disturb their outcomes. To this end, they are usually used along with background subtraction or motion analysis techniques (Cutler et al., 1998).

To find the shape of an object, edge features are of great use as objects can be well be represented through their edges (Nguyen et al., 2011). In contrast to color and texture, the edge features describe objects mainly based on their geometry.

Histograms of oriented gradients (HOG) (Dalal et al., 2005) plus support vector machine (SVM) (Cortes et al., 1995) has been paid great attention and applied to human detection extensively since it was proposed in 2005. Similar to Edge Orientation Histograms (EOH) (Gerónimo et al., 2007) and Scale-Invariant Feature Transformation (SIFT) (Lowe et al., 2004), HOG concentrates on the gradient information of image, but it is different that HOG employs the dense grid of uniformly spaced cells and the overlapping local contrast normalization to strengthen the robustness to illumination and shadow.

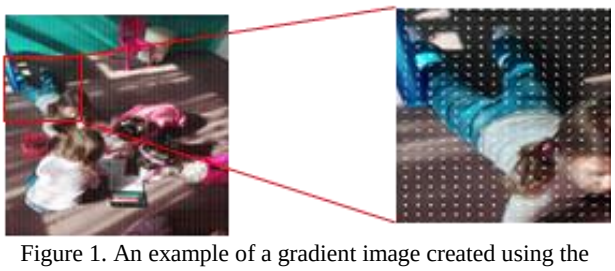

Figure 1. An example of a gradient image created using the HOG feature

UAVs move in a 3D world. A drone's camera undergoes rolling, pitching, heading or a combination of all and this makes the detection more complex. So the feature should be rotation invariant. SRHOG (Liu et al., 2017), Inspired by HOG (Dalal et al., 2005), which utilizes a dynamically defined polar coordinate system (Figure 2) to calculate the gradients via

* Corresponding author

Approximate Radial Gradient Transformation (ARGT) (Takacs et al., 2013), is a rotation invariant feature which can solve the rotation in a plan.

According to the Radial Gradient Transform (RGT) (Takacs et al., 2013) coordination system, which varies with the pixel position, instead of the fixed global (X, Y) system to describe the pixel's gradient, makes the feature vector resistant to the image rotations. Orthogonal bases of local frame are the radial and tangential unit vectors at the pixel p relative to the detection window center C. The components of the gradient (Gr, Gt) are described in directions r and t. Therefore, when this local coordinate system is rotated the resulting vector does not change and makes the feature vector resistant to the image rotations.

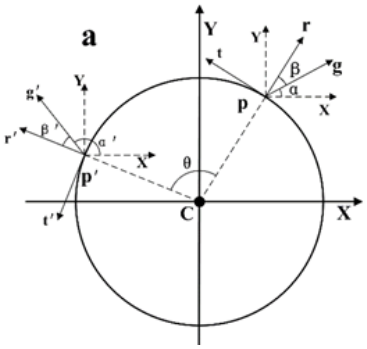

Figure 2. Definition of local coordinate system for each search window in the SRHOG method

As shown in Fig.2, the gradient orientation is redefined as the angle β between the gradient and local basic vector. After the image rotation, the new gradient orientation β' still is equal to the β, which guarantees the rotation-invariance of statistic at each pixel.

The aim of this paper is to assess SRHOG for human detection in UAV images. This paper includes three parts: Part one discusses how to implement the SRHOG function. Experiments performed to test it are recorded in the following. Ultimately conclusions are drawn and recommendations are proposed for future works.

In the remainder of this article, the second part describes how to carry out evaluations.

2. METHODOLOGY OF COMPUTING SRHOG FEATURE AND ITS IMPLEMENTATION FOR HUMAN DETECTION

In this section the methodology is explained briefly. At first, the image is scanned by a 128*128 search windows. Then the radial and tangential gradients (Figure 2) of each pixel are calculated via Approximate Radial Gradient Transformation method. The magnitude and direction of gradient vector is achieved by:

$$\text{Gradient magnitude} = \sqrt{G_r^2 + G_t^2} \quad (1)$$

$$\text{Gradient direction} = \arctan\left(\frac{Gt}{Gr}\right) \quad (2)$$

Where Gr and Gt are the radial and tangential gradients respectively. Once the gradients are calculated, using 15 co-centered circles and 16 angular sectors, the search window is split into several sector rings. It is worth noting that these blocks have some overlaps (Figure 3) to make the search window feature more robust against changes in illumination (Liu et al., 2017). The next step is measurement of each block's gradient histogram. Once the blocks are formed, the gradient histogram of gradients shall be determined. The horizontal axis of this histogram corresponds to gradient directions ranging from 0^o to 160^o (i.e. 9 bins per 20^o). Pixels in the block whose gradient directions are within the bin range, are assigned to that bin. The height of each bar is equal to the weighted sum of pixel gradient magnitudes. At the end, the feature vectors of block histograms are put together to form the final feature vector of the search window.

In this paper the supervised classification method is used to identify humans, the Support Vector Machine (SVM). SVM is conducted in two main phases: training and testing. During the training phase, a classification model is used later to assign each test picture to the human or non-human tag. For this, features of many positive (completely or partially containing humans) and negative (containing no humans) are extracted using SRHOG. These features set out the training data that SVM needs to build up its test model. To find the label of any test image its features are extracted and passed to the SVM classifier. In the end, all test images that contain a one or more humans are assigned to the human category whereas the test images that contain no humans are assigned to the non-human category. To find a human in scene a sliding window approach have been used which label each window by means of the mentioned method. There exists large gap between the data input speed and processing speed in large-size images. To shorten this gap, a parallel processing scheme is used. In this method several process (the number of process depends on the CPU) can be perform at the same.

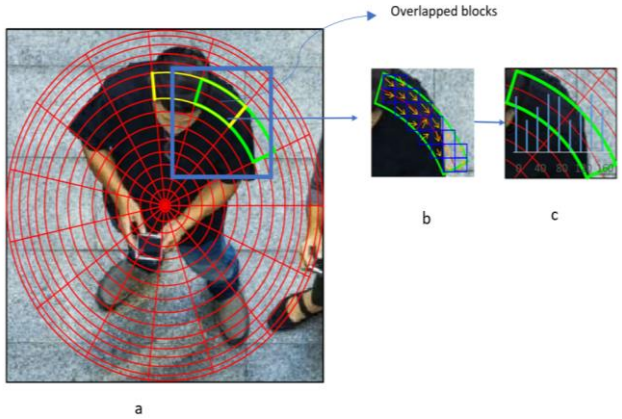

Figure 3. a) spatial configurations in SRHOG, b) gradient of pixels in one of the blocks in the search window, c) histogram of gradients of the block

3. EVALUATIONS

Three datasets were used for the evaluations. The first is the proposed INRIA dataset along with HOG feature (Dalal et al., 2005), which was frequently used in numerous studies as a systematic benchmark for testing human and pedestrian detection algorithms (Benenson et al., 2014). The collection

contains only standing, walking and upright views of people. However, as already described, an individual is mostly viewed from a top-down angle on a UAV image. Therefore, humans are deformable objects, and thus have variations in the class. The training dataset should therefore be detailed to allow for accurate classification. Thus, we acquired and used many additional images taken by AR Drone 2.0, DJI Tello, DJI Inspire 2.0, and DJI Phantom 4 Pro drones.
Figure 4 shows some example image from INRIA data set, from which we used 2164 positive and 432 negative samples for training and 1126 positive and 453 negative images for the test.

Figure 4. Some examples of INRIA dataset

pixels. However, as mentioned before, the samples need to be 128x128 pixels. Therefore, to increase the dimensions, the border parts of samples were duplicated and merge to them (Figure 5a) to make them 128x128 pixels image.

Figure 5. Some examples of Drone images

The additional drone images acquired by the authors were 592 and 200 positive and negative samples respectively. The positive images included humans either completely or partially. Figure 5 shows some examples of both positive and negative images.

Figure 6. Some examples of Tello images.

The Tello images were taken at lower attitudes were to make the training dataset even more comprehensive. In the training phase, 162 positive and 92 negative images taken by Tello were added to other two data sets. Some examples are shown in Figure 6.

To evaluate the performance of the algorithm, three indices Recall, Recall_neg, and Precision (Wójcikowski, 2016), were used which are computed using TP, TN, FP, and FN figures (Table 1). The Recall index shows the ratio of the correctly identified positive windows over the total number of positive windows and is computed by:

$$\text{Recall}=\frac{TP}{TP+FN} \qquad (3)$$

In effect, Recall shows how strong the proposed feature is in identifying the positive samples.

The second index, Recall_neg, refers to the ratio of correctly identified negative samples over the total number of all negative samples. and is calculated by:

$$\text{Recall\_neg}=\frac{TN}{TN+FP} \qquad (4)$$

A bigger Recall_neg suggests that the procedure is stronger and makes less mistakes. The last, Precision, refers to the ratio of correctly identified positive windows over all of the positive windows and is computed by:

$$\text{Precision}=\frac{TP}{TP+FP} \qquad (5)$$

Precision shows the overall accuracy of the method.

Index	Meaning
TP	Positive windows correctly identified
FN	Negative windows identified as negative
TN	Negative windows correctly identified
FP	Negative windows identified as positive

Table 1. Meaning of indexes

For the first two tests, ROC curve and the area below the curve was also calculated which is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. And the AUC (area under curve) tells how much model is capable of distinguishing between classes.

4. RESULTS AND DISCUSSIONS

At first, the overall efficiency of the SRHOG feature was evaluated using the INRIA data set. Then this test was carried out once again but this time with Drone images to test the ability of it in detecting humans in nadir images. Second, the ability of SRHOG in detecting humans appearing in different situations like standing and sitting was evaluated.

The results of the first experiment are presented in Table 2.

Method	TP	FN	TN	FP	Recall	Recall_neg	precision
SRHOG	846	280	151	302	0.7513	0.3333	0.7369

Table 2. SRHOG Performance on INRIA dataset

As shown in Table 2, the precision of SRHOG is 73.69% which is not too bad. However, the Recal_neg has a very low value.

Indeed, here out of 453 negative samples, SRHOG classified only 151 cases correctly. This means despite its relatively good accuracy in detecting correct positive labels, SRHOG leads to too many false negative labels too.

Figure 7 shows ROC curve of applying the SRHOG +SVM on INRIA data set. The area under the curve is 0.2040 which shows the somehow low capability of distinguishing between classes.

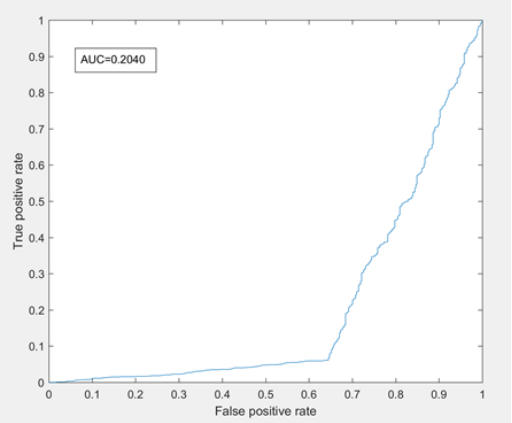

Figure 7. ROC curve for INRIA

Table 3, shows the next experiment were the performance of SRHOG for detecting humans in UAV images is studied. In this experiment, 803 positive and 150 negative samples taken by a Parrot Ar. Drone camera (Fig. 6) are used.

Method	TP	FN	TN	FP	Recall	Recall_neg	precision
SRHOG	597	206	44	106	0.7435	0.2933	0.8492

Table 3. Performance SRHOG on Drone images

Among the 803 positive samples, the SRHOG classified 597 cases correctly. However, among 150 negative samples, once again, SRHOG classified only 44 cases correctly. This suggests the same conclusion as those of the previous experiment.

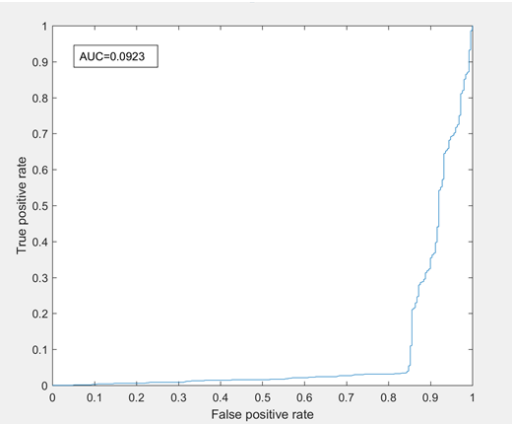

Figure 8. ROC curve for UAV images

Figure 8 shows ROC curve of applying the SRHOG +SVM on images taken by AR. Dron. The area under the curve is 0.0923 which is much less than previous test. Therefor the capability to distinguish between classes dropped.

As mentioned, the third experiment concerned checking SRHOG in images that include a person in various standing or seating positions and lighting conditions. In other words, injured people can appear in a variety of positions, such as standing, sitting, lying down, and a part of the body covered or is in shadow. There is also no guarantee that a human is imaged under proper lighting conditions.

In this experiment, for each situation,500 positive samples taken by Tello drone were used. This is because only in this data set, we were able to capture various standing positions. Examples of these images are presented in Figure 9.

Figure 9. Human in different situations

Table 4 shows the results. In this Table, Occluded body refers to the cases which the upper or lower half of the body appear in the image. Inappropriate lighting conditions refer to the situation which whole body or part of body is in the shadow.

As can be seen in the table, when in a standing and lying position, a human has a better chance of being detected. The Recall values in standing and sitting, positions are 73.42% and 51.63% respectively. sitting position suggest some 21.79% decrease in comparison with standing position. In weak lighting conditions, the Recall value is 55.48%, which reduced by 19.65% compared to standing position.

Another point to note, is the low percentage of occluded human detection. Using the SRHOG algorithm only 26.2% of images contain humans were detected. Obviously, this is not acceptable for an image classification technique. Perhaps, the low accuracy of the results is due to the lack of appropriate training dataset. In our experiments, the algorithms were trained mainly using images containing entire human body. As a result, the classification gave poor results when trying with test data that only partially contained humans.

position	TP	FN	Recall
standing	367	133	0.7342
siting	258	242	0.5163
lying	425	75	0.8510
Weak lighting condition	377	183	0.5548
Occluded body	131	369	0.2620

Table 4. Performance of SRHOG in different situations

5. CONCLUSION

In this paper, several experiment were carried out to examine the SRHOG performance in detecting humans in UAV images. The tests were carried out in all sitting and lying positions, occluded body and inappropriate light conditions were

separately studied. The SRHOG Recall in lying and standing positions were respectively 85.10% and 73.42%.

The biggest weakness of SRHOG was in giving many false labels where the image does not contain any humans. Development of a feature to overcome this issue is desired in the future studies. It was also observed that the feature has a smaller success rate in detecting a sitting position, which shows a different appearance from a human being in the image. Last, perhaps the most significant problem was occlusion, where applying part-based techniques may lead to better results.

.

REFERENCES

Bahri, H., Chouchene, M., Sayadi, F.E. and Atri, M., 2019. Real-time moving human detection using HOG and Fourier descriptor based on CUDA implementation. Journal of Real-Time Image Processing, pp.1-16.

Benenson, R., Omran, M., Hosang, J. and Schiele, B., 2014, September. Ten years of pedestrian detection, what have we learned?. In European Conference on Computer Vision (pp. 613-627). Springer, Cham.

Blondel, P., Potelle, A., Pegard, C. and Lozano, R., 2013, November. How to improve the HOG detector in the UAV context.

Cortes, C. and Vapnik, V., 1995. Support-vector networks. Machine learning, 20(3), pp.273-297.

Cutler, R. and Davis, L., 1998, August. View-based detection and analysis of periodic motion. In Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170) (Vol. 1, pp. 495-500). IEEE.

Dollár, P., Babenko, B., Belongie, S., Perona, P. and Tu, Z., 2008, October. Multiple component learning for object detection. In European conference on computer vision (pp. 211-224). Springer, Berlin, Heidelberg.

Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE.

Ding, J. and Zhao, G., 2019. An improved 3D intelligent dynamic face recognition algorithm based on computer vision.

Gerónimo, D., López, A., Ponsa, D., & Sappa, A. D. (2007, June). Haar wavelets and edge orientation histograms for on–board pedestrian detection. In Iberian Conference on Pattern Recognition and Image Analysis (pp. 418-425). Springer, Berlin, Heidelberg.

Karg, M. and Scharfenberger, C., 2020. Deep Learning-Based Pedestrian Detection for Automated Driving: Achievements and Future Challenges. In Development and Analysis of Deep Learning Architectures (pp. 117-143). Springer, Cham.

Leibe, B., Seemann, E. and Schiele, B., 2005, June. Pedestrian detection in crowded scenes. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (Vol. 1, pp. 878-885). IEEE.

Liu, B., Wu, H., Su, W. and Sun, J., 2017. Sector-ring HOG for rotation-invariant human detection. Signal Processing: Image Communication, 54, pp.1-10.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), pp.91-110.

Mihçioğlu, M.E. and Alkar, A.Z., 2019. Improving pedestrian safety using combined HOG and Haar partial detection in mobile systems. Traffic injury prevention, 20(6), pp.619-623.

Nguyen, D.T., Li, W. and Ogunbona, P., 2009, November. A part-based template matching method for multi-view human detection. In 2009 24th International Conference Image and Vision Computing New Zealand (pp. 357-362). IEEE.

Ojala, T., Pietikainen, M. and Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence, 24(7), pp.971-987.

Ott, P. and Everingham, M., 2009, September. Implicit color segmentation features for pedestrian and object detection. In 2009 IEEE 12th International Conference on Computer Vision (pp. 723-730). IEEE.

Prasad, P.S., Pathak, R., Gunjan, V.K. and Rao, H.R., 2020. Deep learning based representation for face recognition. In ICCCE 2019 (pp. 419-424). Springer, Singapore.

Sun, J., Li, B., Jiang, Y. and Wen, C.Y., 2016. A camera-based target detection and positioning UAV system for search and rescue (SAR) purposes. Sensors, 16(11), p.1778.

Takacs, G., Chandrasekhar, V., Tsai, S., Chen, D., Grzeszczuk, R. and Girod, B., 2013. Rotation-invariant fast features for large-scale recognition and real-time tracking. Signal Processing: Image Communication, 28(4), pp.334-344.

Walk, S., Majer, N., Schindler, K. and Schiele, B., 2010, June. New features and insights for pedestrian detection. In 2010 IEEE Computer society conference on computer vision and pattern recognition (pp. 1030-1037). IEEE.

Wójcikowski, Marek. "Histogram of oriented gradients with cell average brightness for human detection." Metrology and Measurement Systems 23, no. 1 (2016): 27-36.

Zhang, J., Wu, X., Hoi, S.C. and Zhu, J., 2019. Feature agglomeration networks for single stage face detection. Neurocomputing.

Zhang, H.B., Zhang, Y.X., Zhong, B., Lei, Q., Yang, L., Du, J.X. and Chen, D.S., 2019. A comprehensive survey of vision-based human action recognition methods. Sensors, 19(5), p.1005.

Zhang Sr, D., Shao, Y., Mei, Y., Chu, H., Zhang, X., Zhan, H. and Rao, Y., 2019, May. Using YOLO-based pedestrian detection for monitoring UAV. In Tenth International Conference on Graphics and Image Processing (ICGIP 2018) (Vol. 11069, p. 110693Y). International Society for Optics and Photonics.