IMPROVING OPERATIONAL AWARENESS OF AUTONOMOUS DETECTION AND TRACKING OF HUMANS USING UNMANNED AERIAL VEHICLES

Sambit Bhattacharya, Fayetteville State University, USA
Bogdan Czejdo, Fayetteville State University, USA
Joshua Cutolo, Fayetteville State University, USA
Allen Dudek, Fayetteville State University, USA
Joshua Pinos, Fayetteville State University, USA

ABSTRACT

Currently, there are no commercial autonomous devices that can track humans reliably in real time in all situations. There are two main reasons for that: the inefficiencies of the human tracking algorithms and not sufficient processing performance of a drone. As of now, there are many relatively accurate (~70%) human detection algorithms in existence, but to run these algorithms in real time a supercomputer is required, or at least highest range GPUs. There are also alternative human detection algorithms with guaranteed real time performance but their accuracy (~50%) is left to be desired. The goal of this project was to assess the image quality with respect to object tracking and inform the research community about such quality. Another related goal was to guarantee the high accuracy if the assessment is positive. In practice, this implies our novel approach can provide information about the presence/absence of humans in an area and the specific accuracy of human detection by a drone in a given situation i.e. lighting, shadows etc. The achieved operational awareness of the accuracy of autonomous human detection can be crucial in some applications when the operator needs to rely on the evaluation of the quality of the drone output. This will give the operator several choices such as taking control of a drone and tracking humans by himself/herself or sending another drone with an infrared camera. We have constructed a large dataset of images with and without humans based on both our experiments and images taken from the UCF (University of central Florida) Lockheed Martin UAV dataset. We have developed an assessment method to identify images of high quality with respect to an object detection. We chose two different methods to detect humans in this subset of high quality images: Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). Both methods provide algorithms to create feature vectors that are needed for classification. After the feature vectors had been computed we then used a Support Vector Machine (SVM) to classify the images based on their respective feature vectors. Our experiments showed that we can identify humans in images assessed as high quality with very high accuracy around 98%.

Keywords: *Unmanned Aerial Vehicles, Detection of Humans, Machine Learning, Operational Awareness.*

INTRODUCTION

The commercial drones currently in use are essentially lightweight flying computers, with the low powered CPUs onboard that can process a moderate amount of data in regard to stabilization, transmission of information, and sensory data. These drones are typically not suited for processing the vast amount of data that would be needed for real time human detection. In the future, even smaller drones with embedded GPU processors may possibly have the power to accomplish this task. This could replace the drone's own processor, allowing the drone's fight system to be managed by the CPU, with the image-based computations being done on the GPU itself.

In addition to limited processing power of drone's CPU, we need to deal with the inefficiencies of the human tracking algorithms. As of now, there are many relatively accurate (~70%) (Dalal and Triggs, 2005) human detection algorithms in existence, but to run these algorithms in real time would require a supercomputer, or at least a highest range GPU. There are also alternative human detection algorithms with guaranteed real time performance but their accuracy (~50%) is left to be desired.

Taking into account future progress, more precise determination of image quality with respect to object tracking will still be needed. It will be crucial to have access to information about such assessment, and also guarantee high accuracy if the assessment turns out positive. Our project addresses these needs and the developed system can provide information about the presence/absence of the humans in an area and the specific accuracy of human detection given variable conditions (Burningham at al, 2002) (Wang and Bovik, 2006) (Sheikh and Bovik, 2006) i.e. background type, lighting, shadows etc. The achieved operational awareness of the accuracy of autonomous human detection is important in applications when the operator wants to rely on drone output for quality evaluation. If the assessment indicates low reliability the operator would have several choices such as taking control of a drone and tracking a human by himself/herself or sending another drone with infrared camera, including other actions that can be allowed by increased operational awareness.

Our novel approach to using the drones for autonomous detection and tracking of humans is illustrated in Figure 1. The drone sent on a mission to detect humans in a defined area provides background image that is first analyzed with respect to quality (Step 1). There are several measures that were chosen by us to estimate quality conditions (Burningham at al, 2002) (Wang and Bovik, 2006) (Sheikh and Bovik, 2006) such as absolute resolution, relative resolution with respect to human size, contrast, lighting, and shades. If the image quality (reliability estimation) is sufficient, the drone can work in the autonomous mode (Step 2b and skip Step2a) providing all necessary information for an operator (Step 3) (Ahonen and Pietikainen, 2004) (Dalal and Triggs, 2005) (Zivkovic and Van der Heijden, 2006). If this is not the case the drone can autonomously change the location/position to improve the image quality of observed area (Step 2a) and provide information about detected humans. If the reliability of detection is still not sufficient an operator is made aware of that and can decide (Step 3) whether to take over the drone operations or do another appropriate action e.g. send other drone with enhanced capabilities such as an infrared sensor. The process continues within such loop with the operator provided with operational awareness and given some options for a compensating action in each cycle. Let us emphasize the improved situational awareness provided by our drone system. The operator is not only informed about presence/absence of any humans in the defined area but also about the accuracy of this information.

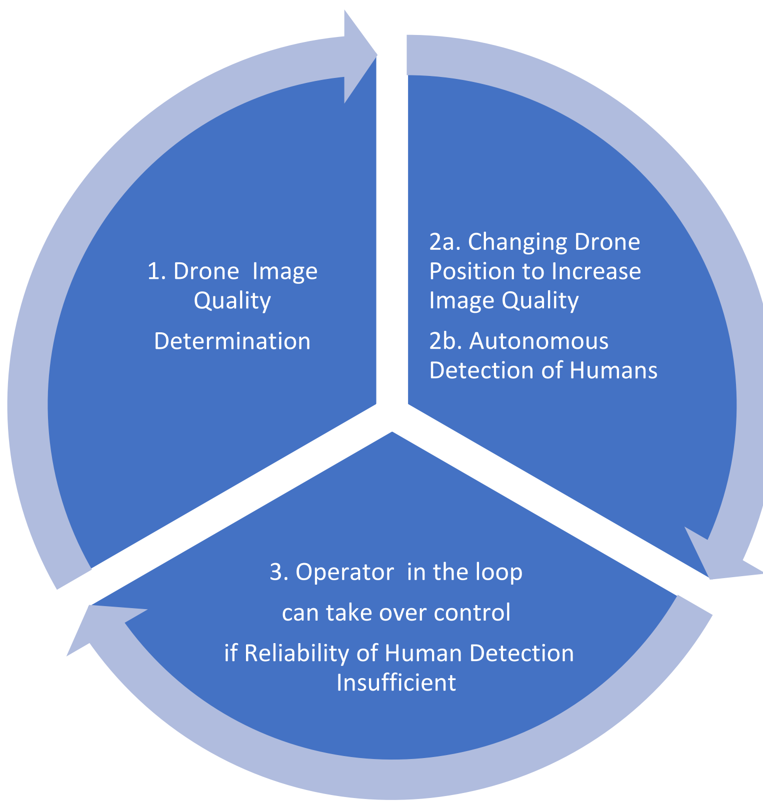


Figure 1. Operational Awareness for Human Detection System

SELECTION OF DATASETS BASED ON IMAGE QUALITY ASSESSMENT

General Image Quality can be defined as "the weighted combination of all of the visually significant attributes of an image" (Burningham at al, 2002). Such a definition emphasizes the general perceptual assessments performed by human viewers. Our interest is rather in Specific Image Quality with respect to an object detection. Such specific image quality requires an extended definition to cover all aspects of an assessment to determine difficulty in detection of a specific object i.e. a human.

In practice, assessment of our specific image quality is a combination of visually significant attributes for general image quality plus visual attributes having impact on the detection step. Since the process of determining the level of general accuracy is called Image Quality Assessment (IQA) we will coin a new term Specific Image Quality Assessment (SIQA). Image quality assessment is part of the quality of experience measures. SIQA like IQA can be assessed using subjective, objective or combined methods but the objective methods are more appropriate. The reason is that computer might identify differences in quality in a set of images where a subjective person (Sheikh and Bovik, 2006) might not, and vice versa.

Let us start from discussion of the visually significant attributes covered by General Image Quality assessment using objective methods. It can be performed by different algorithms that analyze the distortions and degradations introduced in an image. The typical methods (Sheikh and Bovik, 2006) are based on experimental measurements of image quality attributes. Sharpness, as one of the most important attributes, determines the amount of detail an image can convey. Drone image sharpness is affected by the drone camera's lens, sensor and focusing ability. Sharpness can also be affected by movement or shaking of the drone. There are algorithms to sharpen the image but there are serious limitations especially when the detection of objects is the task at hand. There are many other image quality attributes like noise, dynamic range (or exposure range) and contrast, which usually play a lesser role but need to be considered when their distortions are excessive. Some of these can result in misclassification of highlights or shadows as humans. Yet another distortion can be caused by software for data compression and transmission loss. This distortion is especially important for drone systems with external image processing that need to use heavy compression to be able to send data in real-time.

In our project we used an AR drone that was developed by the Parrot company. The AR Parrot's drone recorded video via the drone's onboard webcam. We have constructed a large dataset of images with and without humans based on both our experiments and images taken from the UCF (University of central Florida) Lockheed Martin UAV dataset. Most of the images were taken from a height approximately 450 feet and contained several humans entering and leaving at various time during the recording. Using our image assessment method, we classified the images into high quality and low quality. After this process we got 90,000 high quality images.

DETECTION OF HUMANS IN IMAGES

Histogram of Oriented Gradients, or HOG, (Dalal and Triggs, 2005) is one way of creating feature descriptors in image processing and computer vision to detect objects in an image. For the purpose of this project, the objects in question were humans. In theory, HOG splits an image into smaller and smaller connected regions and compiles a histogram of gradient directions for each cell in an image which, in combination with one another, creates the feature descriptors.

Local Binary Patterns, or LBP, is another way of creating feature descriptors. In theory, LBP divides the image into 16x16 cells. Each pixel in the cell is compared to its nearest 8 neighbors. An 8 bit binary number is created from this comparison, with a label of 1 if a neighbor is greater than the center pixel and a 0 if the neighbor is smaller. Next, a histogram is created of the frequency of each 1 or 0 in the binary number. This is done for all pixels in an image and then the histograms are combined creating feature vectors for the image.

After obtaining the feature vectors, a Support Vector Machine (SVM) was used for human detection. SVM is a supervised learning model that recognizes patterns and uses them for classification. In our case, the SVM classified whether an image contains human or not. We randomly split up our data, with 30% of it for testing and 70% of it for training. From there, the program constructed by us created an SVM model based on training data and after testing printed out the confusion matrix and the classification report. To summarize, the information was provided on how well the SVM classified the data.

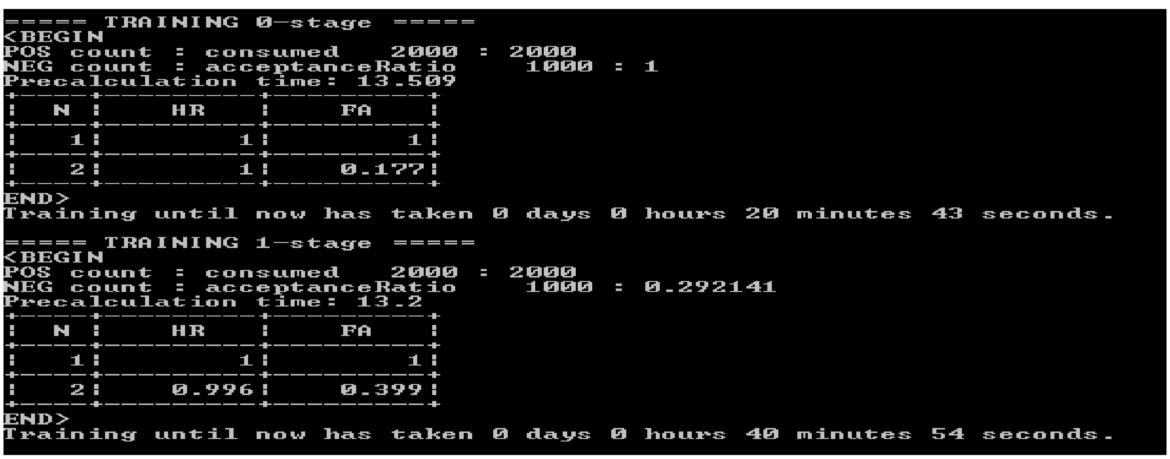
RESULTS

The results from our HOG+SVM training were very good as detailed in the confusion matrix and classification report below. Also, our LBP training had comparable accuracy as shown in Figure 2. With our dataset of 30% of ~90,000 images we obtained a 99% precision score of detecting images without humans and a 96% precision score of detecting humans. We thus demonstrated that with proper image quality assessment we can guarantee very high reliability of human detection.

Confusion Matrix:
[[17751   369]
 [219 8824]]

Classification Report:
SVC(C=0.01, cache_size=200, class_weight='auto', coef0=0.0, degree=3,
  gamma=0.0, kernel='linear', max_iter=-1, probability=False,
  random_state=None, shrinking=True, tol=0.001, verbose=False)

	precision	recall	f1-score	support
No Human	0.99	0.98	0.98	18120
With Human	0.96	0.98	0.97	9043
Avg / total	0.98	0.98	0.98	27163



.

Figure 2. Report on LBP Classification

CONCLUSIONS

Real time tracking of objects using drones is a growing field with many applications. In this project we explored methods of accomplishing this task with high reliability. We used image assessment techniques that allowed us to classify images into two classes: good and bad quality subsets. For the good quality subset, we used two methods to create feature vectors for SVM: HOG (Histogram of Oriented Gradients) and LBP (Local Binary Patterns). Using these methods, we were able to successfully train/test the SVM with very accurate results.

A solution to improve HOG detection in the future for bad quality images might be to use the full power of a GPU in a drone's computer. This would allow much faster video processing and better HOG and SVM based detection. At the moment, there are no Python implementations of HOG detection using the GPU. We believe GPUs would also help to speed up the training phase. Another human detection method that is becoming more and more popular is the use of convolutional neural networks (CNNs). We plan to investigate CNNs for human detection with GPU computing equipped drones.

ACKNOWLEDGEMENTS

We would like to acknowledge partial support to this project by the Belk Foundation, and the National Science Foundation (Award Id: 0959958).

REFERENCES

Burningham, Norman; Pizlo, Zygmunt; Allebach, Jan P. (2002). "Image Quality Metrics". In Hornak, Joseph P. Encyclopedia of imaging science and technology. New York: Wiley. doi:10.1002/0471443395.img038. ISBN 978-0-471-33276-3.

Wang, Zhou; Bovik, Alan C. (2006). "Preface". Modern image quality assessment. San Rafael: Morgan & Claypool Publishers. ISBN 1598290223.

Sheikh, Hamid Rahim; Bovik, Alan C. (2006). "Image Information and Visual Quality". IEEE TRANSACTIONS ON IMAGE PROCESSING. 15 (2): 430–444. doi:10.1109/TIP.2005.859378. Retrieved 26 July 2018.

Ahonen, T., Hadid, A., and Pietikainen, M. (2004). “Face Recognition with Local Binary Patterns”. Computer Vision - ECCV 2004, 469–481.

Dalal, N.; Triggs, B. (2005). "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference, vol.1, no., pp.886-893 vol. 1, 25-25 June 2005.

Z.Zivkovic, F. Van der Heijden, (2006). “Efficient Adaptive Density Estimation per Image Pixel for the Task of Background Subtraction”, Pattern Recognition Letters, vol. 27, no. 7, pages 773-780, 2006.

Z.Zivkovic, F.van der Heijden, (2005). “Recursive unsupervised learning of finite mixture models”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.26, no.5, pages 651-656, 2004.