Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach

Vandit Gajjar, Ayesha Gurnani and Yash Khandhediya
L. D. College of Engineering

International Conference on Computer Vision 2017

Abstract

- With crimes on the rise all around the world, video surveillance is becoming more important day by day. Due to the lack of human resources to monitor this increasing number of cameras manually, new computer vision algorithms to perform lower and higher level tasks are being developed. We have developed a new method incorporating the most acclaimed Histograms of Oriented Gradients, the theory of Visual Saliency and the saliency prediction model Deep Multi-Level Network to detect human beings in video sequences. Furthermore, we implemented the k – Means algorithm to cluster the HOG feature vectors of the positively detected windows and determined the path followed by a person in the video. We achieved a detection precision of 83.11% and a recall of 41.27% in a time as short as 23.0047 seconds.

Human Detection using HOG Features

- We firstly emulated the results achieved for comparison purpose. We trained the Support Vector Machine (SVM) classifier as a two class classifier on the dataset by giving cropped human images as positive samples and random images of the same size as negative samples. The classification accuracy on the training dataset was found to be very high in the order of 98.32%.
- We tested the SVM classifier on other unseen images of the dataset. To detect humans from the whole image, the basic algorithm assumes a window of a particular size. Windows of this size are cropped out from the image one by one in a convolutional manner and HOG features are computed. As the size of the image increases, total number of windows to be classified increases and due to extraneous data, the computational time to compute HOG features as well as for classification increases as well. The computation time taken was 1768.29 seconds for a single image in the best case. This method is hence not viable for real-time applications.
- Regardless, we evaluated the results of the classification and found that the many false positive for Human Detection in Normal Images using HOG Features existed in the inferences of the classifier. This was due to vaguely similar HOG features of some parts in the background with the HOG features of a general human being. The recall was computed to be 0.93 which is very high. A proper model for human detection in surveillance videos required a perfect blend of precision as well as recall. Therefore, we did a bit of research to find new region proposal algorithms which would be helpful in reducing the recall as well as the computation time.

Saliency Model

- We used the Deep Multi-Layer Network for saliency prediction. It is called the ML – Net. The purpose of using a Visual Saliency model was to propose possible regions where humans might be in the frame. Visual Saliency maps would indicate higher intensity where there are humans in the image because of results supported. The Deep Multi-Layer Network described in outperforms all other models on the SALICON Dataset and also performs better on the MIT Saliency Benchmark. This model is sufficient for our use i.e. human detection for video surveillance.

Fig. 1 Example of Salience – Windowed Image

End Results and Conclusion

- We observed that detection and motion patterns of humans can be found out efficiently by using the proposed method in this paper. Using Visual Saliency as a region proposal algorithm proved to be beneficial for the research work. Human Detection improved by introducing the salience-windowed frames of the video to the HOG + SVM classifier. The performance improved manifolds as compared to classification on normal images. This happened due to unnecessary data which was in a way increasing the recall of the model and also the computation time and thus making the model not fit to be used in real time. Using the Deep Multi-Layer Network for saliency prediction to propose regions of interests makes the model much more feasible in terms of computational requirements as well as precision and recall. We also observed the usage of k – Means algorithm in clustering feature vectors of the same person and how it can be used to compute the path followed by a person in the video sequence.

References

[1] M. Cornia, L. Baraldi, G. Serra and R. Cucchiara. A deep multi-level network for saliency prediction. In 23rd International Conference on Pattern Recognition (ICPR), 2016, pages 3488-3493. IEEE, 2016.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In (CVPR), 2005, pages 886-893.

[3] N. Dalal, B. Triggs and C. Schmid. Human detection using oriented histograms of flow and appearance. In (ECCV), 2006, pages 428-441.

[4] J. W. Davis and V. Sharma. Background-subtraction using contour-based fusion of thermal and visible imagery. In Computer Vision and Image Understanding 106, no. 2 (2007): 162-182.

[5] T. Judd, K. Ehinger, F. Durand and A. Torralba. Learning to predict where humans look. In (ICCV), 2009, pages 2106-2113.

[6] T. Judd, F. Durand, and A. Torralba. MIT saliency benchmark.

[7] K. Mikolajczyk, C. Schmid and A. Zisserman. Human detection based on a probabilistic assembly of robust part detectors. In (ECCV), 2004, pages 69-82.