Parts Based Detection of AK-47s for Forensic Video Analysis

Captain Justin Jones, Dr. Mathias Kölsch

MOVES Institute, Naval Postgraduate School

Monterey, CA

jajones@nps.edu, mnkolsch@nps.edu

ABSTRACT

Law enforcement officers, military personnel and forensic analysts are increasingly reliant on imaging systems in order to perform in a hostile environment. From surveillance systems to computer forensics, intelligence personnel require a robust method to quickly and efficiently locate objects of interest in images and videos. Most current approaches require a full-time operator to monitor a surveillance video or to sift a hard drive for images that could implicate a suspect.

In this paper, we demonstrate the effectiveness of automated image analysis tools to spot weapons in videos through a standalone application and user-interactive analysis. By training multiple appearance-based classifiers on a large corpus of labeled data, and by combining classifiers through machine learning techniques, we can create an overall classifier to detect individuals holding an AK-47. By automatically locating people carrying weapons in video, operators can be tipped to suspicious events. Current results indicate the value and importance of automated threat detection for forensic analysis in support of the law enforcement and intelligence communities.

ABOUT THE AUTHORS

Justin Jones, Captain, USMC is Signals Intelligence Officer studying in the Computer Science curriculum at the Naval Postgraduate School in Monterey, Ca. Prior to his current assignment, he served as a platoon commander, company executive officer, and battalion collection manager while assigned to 3D Radio Battalion, Kaneohe Bay, Hi. He has operational experience from assignments in Iraq, the Joint Special Operations Task Force-Philippines, Kosovo, and amphibious operations with the 24th Marine Expeditionary Unit. His research focus is also in the application of computer vision techniques in support of modern military operations.

Mathias Kölsch (Ph.D., University of California, Santa Barbara, 2004) is an Assistant Professor of Computer Science at the NPS in Monterey, CA. He is also affiliated with the MOVES Institute, Chair of the MOVES Academic Committee and the Academic Associate for the MOVES curriculum. His research interests include computer vision, hand gesture recognition, augmented and virtual environments, sensor networks and mobile/embedded computing.

Parts Based Detection of AK-47s for Forensic Video Analysis

Captain Justin Jones, Dr. Mathias Kölsch

MOVES Institute, Naval Postgraduate School

Monterey, CA

jajones@nps.edu, kolsch@nps.edu,

INTRODUCTION

Intelligence personnel supporting modern military operations are increasingly reliant on imaging systems to perform in hostile environments. From computer forensics to surveillance, imaging systems impact commander decision making. Correspondingly, increasing amounts of data must be processed to produce an analyzed intelligence product. With the explosion of UAV technology, wide area surveillance assets, and deployment of ground based surveillance platforms, intelligence personnel do not have the manpower to monitor all video feeds for real time decision making. Intelligence personnel also require mechanisms to detect suspicious material during collection operations in order to identify suspicious websites or other targets for monitoring. At the same time, ground forces must increasingly cope with exploiting useful data from electronic devices and media captured during raids on insurgent or terrorist safe houses.

This paper demonstrates computer vision techniques for use in surveillance, computer forensics, and collection operations, by detecting suspicious objects in images. It focuses on the detection of AK-47s, due to their prevalence and potential for a variety of intelligence applications. Using computers to analyze images and video streams decreases the response time to suspicious activity, allowing for real time alerts to be sent to forces and directly leading to lives saved and the disruption of enemy activities.

OPERATIONAL NEED

In recent years, new threats have emerged against the United States of America. This new threat can easily hide among the populace, thwart traditional combat intelligence gathering methods, and exploit seams in the authorities and capabilities of military and government intelligence organizations. With terrorist groups, insurgencies, piracy, drug cartels and other organized crime groups, as well as the emergence of potential peer competitors all posing significant threats to U.S. national security, intelligence professionals require new methods that flexibly support a variety of environments and facilitate rapid decision making.

In order to prosecute threats in these complex environments, forces must be able to precisely locate an individual operating within a large population. Compounding this problem is the amount of information entering an intelligence cell. In 2009 alone, UAV's from the United States produced 24 years worth of video, if watched continuously (Defense Industry Daily, 2010), with new UAV models projected to increase data streams many times over. While UAV technology does provide the warfighter with significant advantages, more data does not necessarily equate to better information. The same applies to the volume of forensic materials collected by ground forces, and intelligence and propaganda intercepted by intelligence collectors. See Figure 1. In order to be relevant to operational forces, the data must be processed, which is a significant weak point with modern intelligence mechanisms.

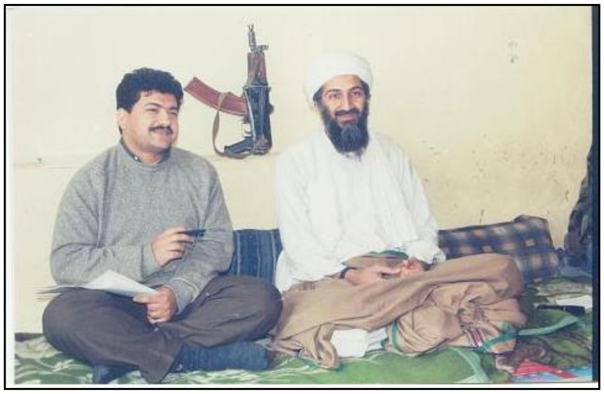

Figure 1. Terrorist material often contains weapons that can be used to focus intelligence efforts.

RELATED WORK

There are a variety of techniques to choose from for identifying weapons in an image for use in a surveillance or forensics application. While color based learning of an object has significant advantages due to its scale and rotation invariance, it is likely not well suited for a military application where object identification must be conducted in low light conditions (including images via night vision devices). Additionally, given the application must locate objects in still frames, techniques that rely on motion, sound, or frame differencing to classify are also not likely to be appropriate. Rapid techniques that are appropriate for real time surveillance monitoring would be just as applicable for scanning images or sub sampling video for suspicious objects in a forensics application. In 2001, Viola and Jones proposed a method for detecting faces based on Haar wavelets, trained with Adaboost (Viola and Jones, 2001), and combined in a sequence they called a “cascade of features.” This seminal work demonstrated a classifier capable of detecting upright faces at 15 frames per second, offering the potential for real time object detection. Lienhart and Maydt expanded on this by developing a richer feature set that included 45 degree rotated features for use in training a strong classifier (Lienhardt and Maydt, 2002). Though weapons are rigid objects, which are well suited for a template based approach; one issue does arise when trying to locate them in images. Weapons typically are recognized by silhouette; therefore, background objects and occlusion can significantly change an objects shape, thereby making a classifier trained on the entire object less likely to respond as a positive detection. This method was chosen in this paper due to its speed of detection and the potential for good precision. Two separately trained whole detectors were compared against parts-based Viola Jones classifiers, with the part combinations modeled by a support vector machine (SVM) and a multi-layer perceptron (MLP).

Other detection methods rely on a variety of features. Histograms of Oriented Gradients (HoG) are a common approach where orientations of gradients are summed in a portion of the image. This method is used in (Dalai, Triggs, Rhone-Alps, and Montbonnot, 2005) (Dalal, Triggs, Schmidt, 2006).

Eigenfaces, or the use of principal components analysis to find the vectors of pixel features with the largest eigenvalues of a face, were studied in detail in (Turk and Pentland, 1991) and (Zhang, Yang, and Lades, 1997).

Edge features are also common to parts based object detection. Edge detection using Gabor filters were used in (Serre, Wolf, and Bileschi, 2007) and (Wu, Zhou, Zhou, and Hu, 2008). Shape recognition is also conducted in (Mikolajczyk, Zisserman, and Schmid, 2003) through the use of a Canny edge detector. Other edge detections based on gradient approaches with the Laplacian operator are used in (Mikolajczyk, Zisserman, and Schmid, 2004).

After choosing a feature set, the best features must be selected out of all the features for learning. Strategies for part learning vary in accordance with the overall goals for object detection. Some strategies, such as found in (Agarwal, Awan, and Roth, 2004), emphasize that parts are clustered into feature sets that are as different as possible, which provides for learning the abstract idea of a part. Another part learning technique using clustering to place parts into a tree for fast object retrieval is used in (Nister and Stewenius, 2006). Other methods, seek to select those features that best classify a validation set. Statistical boosting is common mechanism used to find and train those features that best classify positive and negative examples. Feature selection using boosting is used in (Viola and Jones, 2001), (Mikolajczyk, Zisserman, and Schmid, 2004), and (Serre, Wolf, Bileschi, Riesenhuber, and Poggio, 2007). Rather than choosing a part based on its abstraction, or ability to best classify a validation set, other techniques use part learning for the final classification performance of an object. In (Leibe, Seemann, and Schiele, 2005), a method for detecting pedestrians uses global and local cues via probabilistic top-down segmentation. In (Felzenszwalb, Girshick, McAllester, and Ramanan, 2009), a technique for maximizing over latent part locations is used to determine the presence of a whole object, as opposed to a discriminative process where an object is determined to not be there if enough criteria is not met. Random local feature sampling is used in (Wu, Zhou, Zhou, and Hu, 2008), where random part sampling is matched to randomly trained parts through a response to Gabor filters and a Euclidean distance measurement of the local maxima responses.

Although this work incorporated manual part labeling, automatic and semi-automatic methods exist (Agarwal, Awan, and Roth, 2004), (Kumar and Heber, 2004), and (Leibe, Seemann, and Schiele, 2005).

After selecting the best set of features from a set, detected parts can be combined into a whole object through the use of a trained model. By using the geometry of part detections, the number of false detections and the amount of feature space to search can be reduced. In (Agarwal, Awan, and Roth, 2004), the Sparse Network of Winnows (SNOW) architecture is used to learn associated distance and direction combinations between parts at various scales. Gaussians are used in (Mikolajczyk, Zisserman, and Schmid, 2004) and (Felzenszwalb, Girshick, McAllester, and Ramanan, 2009) to model part combinations.

Markov random fields are also used in (Wu, Zhou, Zhou, and Hu, 2008) for modeling spatial part locations as undirected graphs representing the dependencies between detected parts.

In (Felzenszwalb, Girshick, McAllester, and Ramanan, 2009), parts combinations are learned through the use of star based topology with the total score for detection being a combination of root filter detection location, parts detected, and their associated locations in relation to the trained model.

In (Leibe, Seemann, and Schiele, 2005), the spatial relationship among detected parts was represented by having extracted patch sections compared to a codebook and then having each matched codebook entry ”vote” for the probable location of the object.

Besides object detection for identifying suspicious activities, a number of other techniques have been used to identify potentially suspicious events or human surveillance in video applications. Automated video analysis techniques for finding violent events or surveillance of humans in video were explored in detail in (Kuno, Watanabe, Shimosakoda, and Nakagawa, 1996), (Nam, Alghoniemy, and Tewfik, 1998), and (Vasconcelos and Andrew Lippman, 1997). The methods in these seminal papers rely on audio visual cues, such as the sudden flash or sound of an explosion, dynamic changes between frames, or motion accelerations in relation to human silhouettes to determine the presence of a violent event. While suitable for video, these techniques are not likely to be as applicable to forensic applications searching still frames for suspicious objects.

METHODS

The approach described in this paper combines the benefits of Viola-Jones detections with the flexibility of parts-based methods. Two detectors were trained, one for the rear end of the AK-47 and one for the barrel in the front. The relative spatial distribution of part detections was then learned with a support vector machine (SVM) and a multi-layer perceptron (MLP).

Sources of Training Data

All images for training and testing were provided by selecting frames from videos, which were obtained by searching the Internet. The strategy for training a classifier to detect an AK-47 was to provide a number of imagesof AK-47s, all pointing right, with in-plane and out-of plane rotations of no more than approximately 10 degrees. This trained the classifier to recognize the AK-47 in a specific orientation that could then be used to recognize in plane rotated AK-47s in other orientations through rotation of an image. For the negative image set, scenes were used for training, mostly from video sources. This includes crowded city streets, villages, as well as people holding objects, so that a classifier would not inadvertently learn the hands of people holding AK-47s.

Division of Data

Of the 18 total videos, 13 videos with 1146 annotated images were selected to train the classifiers. These images included a variety of backgrounds and configurations of the weapon. Due to the wide variety of stock configurations, as well as occlusion in a number of images, the stock was not used for training. For the negative image set, 5668 images from 23 videos were used. For testing, 5 positive videos with 687 annotated images, and 24 negative videos with 7045 images were used.

Training

OpenCv utilities were used to create robust classifiers trained on the whole AK-47, and classifiers for parts of a weapon. During the annotation process, a rectangular area containing the object was extracted from each positive image. Prior to training, annotated sections were normalized to a specific size and converted to grey scale. For classifiers trained on the whole object, annotated sections were normalized to 20x40 pixels. Parts were created by taking the annotated section and dividing the width in half. These images were then normalized to 20x20 pixels before training. See Figure 2.

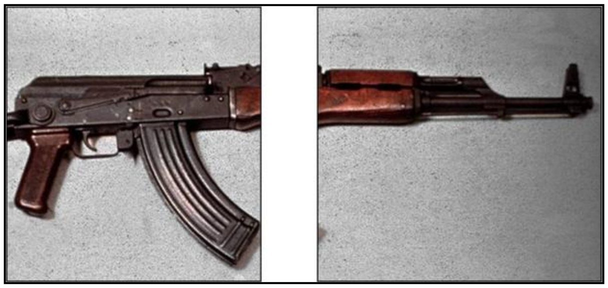

Figure 2. Part training for left and right AK-47 parts.

Number of Images for Training

When training classifiers, more training data is typically better. The first whole classifier, designated Whole_AK, was trained with the 1146 positive and default 2000 negative examples per stage. Another whole classifier, Whole_AK_Negative_Resistant was trained with more negative samples in order to improve the false positive rate. Whole_AK_Negative_Resistant was trained with the 1146 positive samples and 5660 negative samples per stage. The Left_Half_Detector and Right_Half_Detector were trained with the same images as the Whole_AK_Negative_Resistant classifier.

Adaboost with Haar Features

After preparing the images for training, OpenCV's Haar training utility was used to produce the above classifiers. A complete overview of the boosting process is contained in (Viola and Jones, 2001). All classifiers were trained with the extended Haar feature set, with each classifier's cascade containing 20 stages. The specified minimum hit rate for the all classifiers was 0.998 per stage with a maximum false alarm rate for each stage of 0.5.

Training the SVM and MLP

After training part classifiers from the set of annotated training images, another mechanism is needed to determine whether parts detected match a model of orientation and scale consistent with the presence of a known object. In order to train this model, each classifier was passed over the 1146 positive images, and 5568 negative images, with post processing turned on. All combinations of left and right detections in a photo were kept. Detections inside the annotated box of the training set were considered to be true detections, while detections outside of the annotated box or in a negative photo were considered to be false detections. This produced a vector consisting of 3 elements:

- Difference between the left half detection center x value and the right half detection center x value normalized by the mean radius of the 2 detections.
- Difference between the left half detection center y value and the right half detection center y value normalized by the mean radius of the 2 detections.
- Difference between the left and right radii normalized by the mean radius of the two radii.

A graph of the geometry of the detections is provided in Figure 3, showing the cluster of positive detections versus negative detections. Note that the normalized radii difference is not included in the graph. After training over the positive and negative training image set, the results were used to create a hyperplane separating the positive and image sets for use with a support vector machine (SVM). The same data set was also used to train a multi-layer perceptron (MLP) for comparison and in the case the data set was not linearly separable.

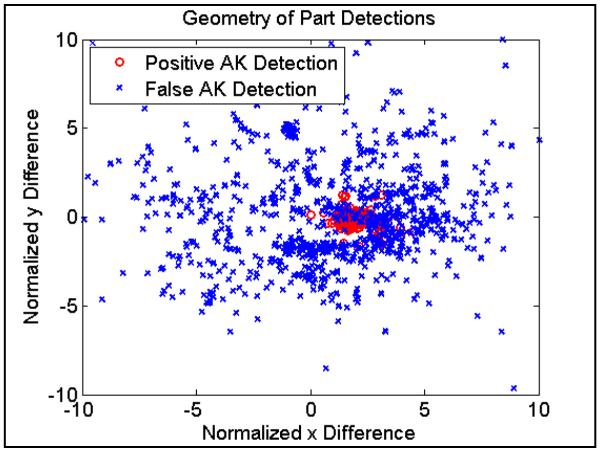

Figure 3. Plot of part detections for SVM/MLP.

Performance Measures

The following equation was used to determine the percentage of weapons detected in the image set:

$$Recall = \frac{NumberOfWeaponsDetectedInSet}{TotalWeaponsInSet} \qquad (1)$$

The following equation was used to determine the percentage of falsely identified weapons in the image set:

$$FalsePositiveRate = \frac{NumberOfFalsePositivesInImageSet}{TotalAreasChecked} \quad (2)$$

While the test set in this paper uses a number of image sizes, the performance metric is compared against a standard video with a frame size of 640 x 480 pixels, running 15 frames per second for one minute. This is a total of 900 frames per minute with several hundred thousand sub-windows checked per frame. The total number of predicted detections is then the false positive rate times the amount of subwindows checked. This is the expected number of false returns that could be expected in a minute of video. For a 640x480 image, a 20x40 whole trained detector is scanned over 314,319 areas. For the smaller 20x20 part detections, 352,718 areas are checked in each 640x480 image. For the entire test image set, 877,639,110 negative areas were checked for each whole detector. For the part detectors, 1,033,439,228 negative areas were checked over the image set. To get the number of areas checked in a minute of video, the number of areas checked in an image for a detector was multiplied by the amount of frames checked in a minute. The total number of areas checked in a minute is then multiplied by the false positive rate in order to receive a number of predicted false positives per minute of video. The following equation gives the number of predicted false positives per minute of video:

$$FalsePositivesPerMinute = FPR * AreasPerFrame * 15 * 60 \quad (3)$$

The predicted false positives per minute of 640x480 video provides a more intuitive way to compare the detectors than the actual false positive rate.

Improving Recall by Reducing the Number of Stages

In order to improve the recall of a trained classifier, stages can be removed from the Viola-Jones cascade. While this makes the overall whole or part classifier more likely to indicate that an object is present, there is also an increase in the number of false positives that will be detected by the classifier. To create the ROC curves in the following section, classifiers were removed from the Viola-Jones cascades until a recall above 85 % was obtained.

Detecting an AK-47 with Haar Classifiers and Machine Learning Techniques

After training Haar classifiers for the Left and Right Halves and a machine learning technique to combine part detections, all classifiers were evaluated against a test set. In order to detect an AK-47, the Left and Right Half detectors are first scanned over an image. A vector containing the normalized x center difference, normalized y center difference, and normalized radii difference is created for all combinations of left and right detections in an image. In general, the Left Half Detector is much more discriminative than the Right Half. All combinations of the left and right halves are evaluated with either the support vector machine or a multi-layer perceptron, producing a final classification for the detection.

EXPERIMENTS

The test image set consisted of a set of 687 positive images containing AK-47s and 7045 negative images, all from internet sources. All positive images were annotated by a human. All test AK images were consistent with the training set: barrel pointing right, with in-plane and out-of plane rotations of no more than approximately 10 degrees. The test set was not rotated to achieve these results.

The following experiments were conducted using the above approaches. Two Viola-Jones classifiers for the whole object were created to verify that a Viola-Jones classifier could be trained to find AK-47s in images, and as baseline for comparison to the parts-based detection. The second whole trained classifier was trained with more negative images per stage in an attempt to achieve a lower false positive rate. Parts trained for the AK-47 were expected to have a higher recall, but poor false positive rates in comparison to classifiers trained on the whole object. Part combination techniques with a support vector machine and a multi-layer perceptron were expected to leverage the higher recall capability of part detections, but keep the false positive rate lower in comparison to the part detections.

RESULTS

Whole Detections

Results for the image set with the Whole_AK and Whole_AK_Resistant detectors are shown in Figure 4. Note that Whole_AK has a higher recall but a higher false positive rate, due to being trained with less negative images.

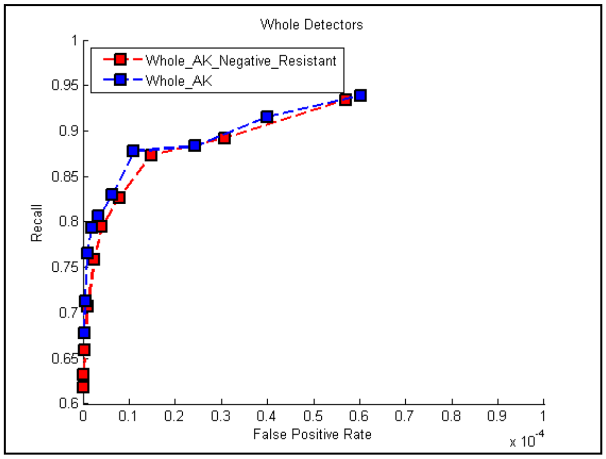

Figure 4. Whole image detections.

Part Detections

Results for part detectors are shown in Figure 5 comparing the Left_Half classifier against the Right Half classifier. This is a comparison of detecting AK-47s through just detection of one of the parts. While parts have higher recall than whole detectors, part detectors alone generate far more false positives than whole trained and the part-based SVM/MLP techniques. The Right_Half detector has the highest recall of any detector, but a very poor false positive rate.

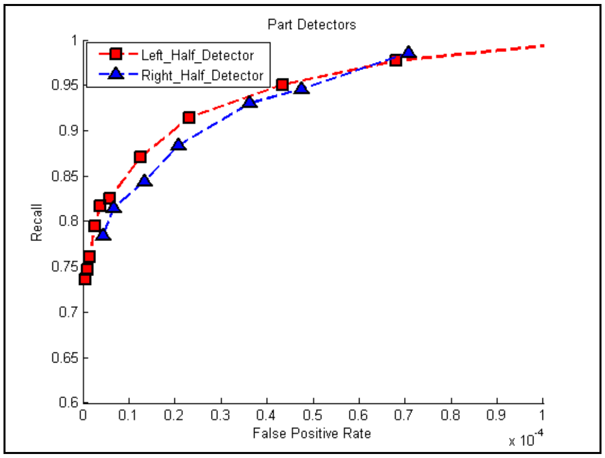

Figure 5. Part image detections.

Part Detections with an SVM

Results for part detectors and an SVM are contained in Figure 6. The Left_Half and Right Half classifiers were applied to the image, with a Support Vector Machine used to evaluate all combinations of detections. In order to achieve a recall rate approaching 85%, stages must be removed from both classifiers. At the upper end of the recall scale, as both stages are progressively removed from both classifiers, the false positive rate greatly increases. This is due to the combination of all left and right detection combinations being applied and checked against the SVM model.

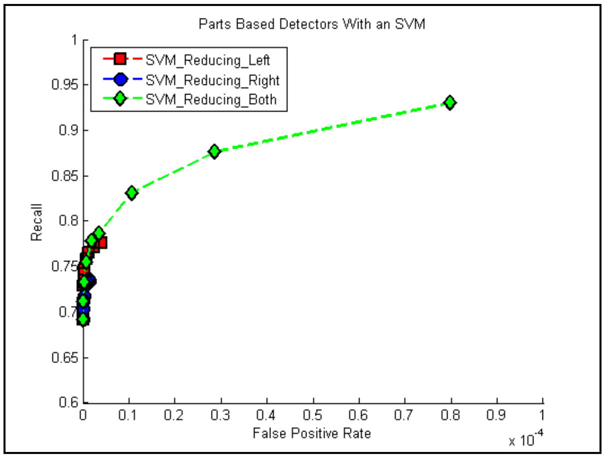

Figure 6. Parts-based SVM image detections.

Part Detections with an MLP

Results for the image set with part detectors and an MLP is contained in Figure 7. The Left_Half and Right Half classifiers were applied to the image, with a multi-layer perceptron used to evaluate all combinations of detections. At the upper end of the recall scale, as both stages are progressively removed from both classifiers, the false positive rate greatly increases. This is due to the combination of all left and right detected combinations being applied and checked against the MLP model.

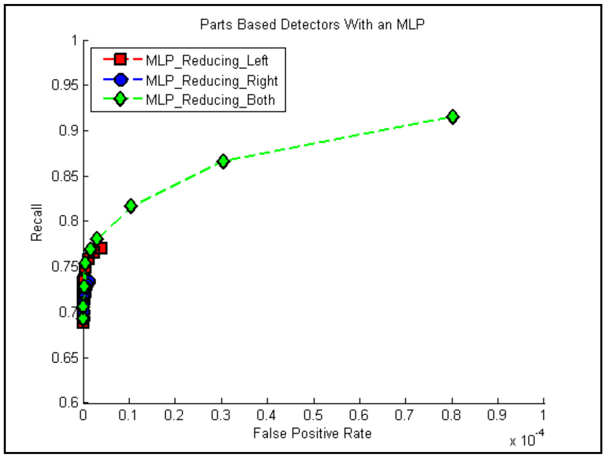

Figure 7. Parts-based MLP image detections.

All Detector Performance - Recall vs. FPR

In Figure 8, all detectors all compared. Due to the wide variation in false positive rates generated, detector performance is best evaluated where a chosen operational system would likely operate. In Figure 9, the graph is capped at an FPR of $2{*}10^{-7}$ or about one false positive per 5 million areas checked.

All Detector Performance: Recall vs. FPs per Minute of Video

In order to better represent the false positive rate of an operational system, the FPR is also shown in terms of the number of false positives per minute of video at a 15 frames a second at 640 x 480 resolution. See Figure 10. In Figure 11, the graph is capped at 200 false positives per minute of video to show classifier performance at the lowest false positive rates.

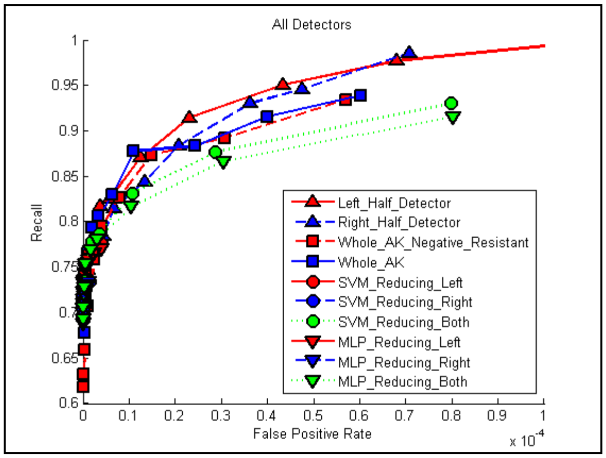

Figure 8. All detector performance total ROC.

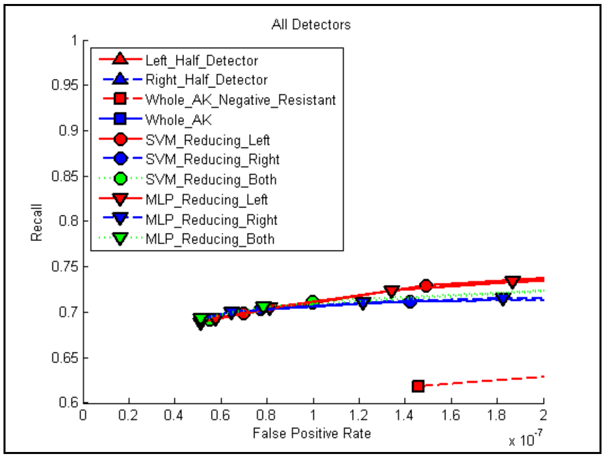

Figure 9. All detector performance capped at FPR of $2.0*10^{-7}$.

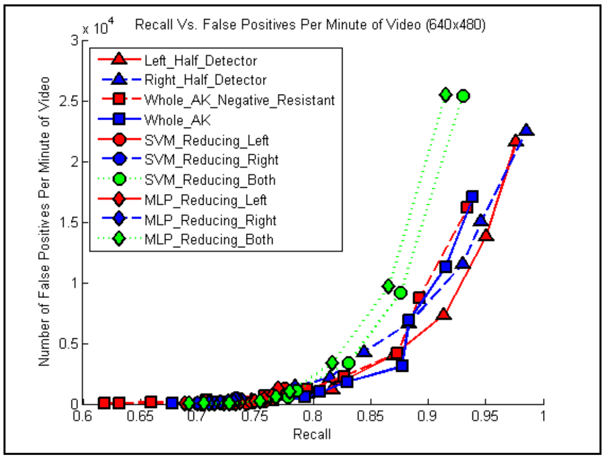

Figure 10. All detectors. Recall vs. FPM

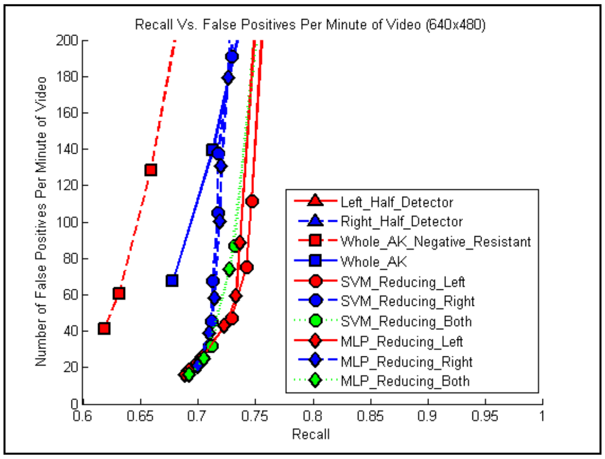

Figure 11. All detector performance. Recall vs. FPM capped at 200 false positives per minute of video.

Overall Evaluation of Methods and Classifiers

For the Whole trained detectors, Whole_AK (all stages) had a recall over the image set of 67.8%, but 67.6 false positives per minute of video (FPM). Whole_AK_Negative_Resistant (all stages) had a starting recall of 61.8%, but generated 41.2 FPM.

For the Part trained detectors, the Left_Half detector (all stages) has a recall of 73.6%, but generates 208.2 FPM. The Right_Half detector (all stages), with the overall highest starting recall, has a starting recall of 78.4%, but generates 1414.8 FPM.

The Left_Half and Right_Half detectors (all stages) using a Support Vector Machine increase the recall over the Whole_AK_Negative_Resistant detector, while significantly decreasing the FPM over whole and part based detectors. The Left and Right Half detectors with SVM had a starting recall of 69.1%, while producing only 17.5 FPM. In comparison against Whole_AK_Negative_Resistant (which was trained with the exact same image set), the parts based SVM classifier increased recall by 7.3%, with a 57.5% reduction in the amount of false positives per minute.

The Left_Half and Right_Half detectors (all stages) using a multi-layer perceptron also increased the recall over the Whole_AK_Negative_Resistant, while significantly decreasing the FPM over whole and part based detectors. The Left and Right Half detectors with MLP had a starting recall of 68.8%, while producing only 16.3 FPM. In comparison against Whole_AK_Negative_Resistant, the parts based MLP classifier increased recall by 7.0%, with a 60.4% reduction in the amount of false positives per minute. While the part detectors coupled with machine learning to combine part combinations have better recall and lower false positive rates than whole detectors, as stages are removed from the part classifiers and become less discriminative, the combination of left and right detections begin to produce a higher rate of false positives than all other detectors. An operational system would likely not be focused on rates at this end due to the extremely high false positive rates making the system completely unusable.

Rotation

The above tests were run over an image set with right facing AK-47s, with angles similar to those used in training. AK-47s at other angles can be found by rotating the images through fixed angles and scanning the classifier over the rotated images. With an increase in the amount of runs for the detector, the false positive rate was expected to increase. In order to test this hypothesis, a prototype was tested over a separate image set of 3727 images from forensic hard drives using the Left and Right Half detectors (both unchopped) and an SVM, with a rotation angle of 10 degrees. None of the images contained AK-47s. Out of 1,517,722,847,280 areas checked, 571 false positives were detected, with a FPR of 3.76222E-10 or approximately one false positive per 2,658,005,114 areas checked. While the detector has not been tested against a large positive image set at this point (due to a lack of a large image set with in-plane rotated AK-47s), Figure 12 and Figure 13, confirm the capability to find rotated AK-47s with these methods of training. Methods and classifiers contained in this paper are capable of finding weapons similar to an AK-47, including RPKs and AK-74Us. See Figure 14 and Figure 15.

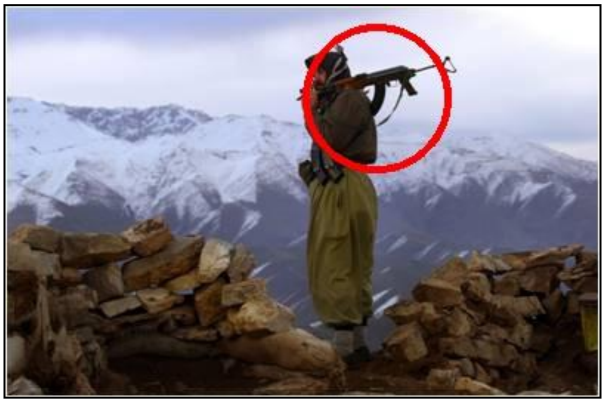

Figure 12. Parts Based Viola-Jones classifiers with SVM true positive.

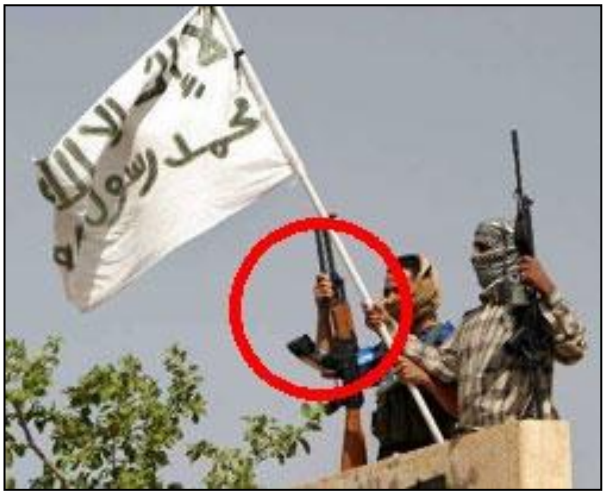

Figure 13. Parts Based Viola-Jones classifiers with SVM true positive.

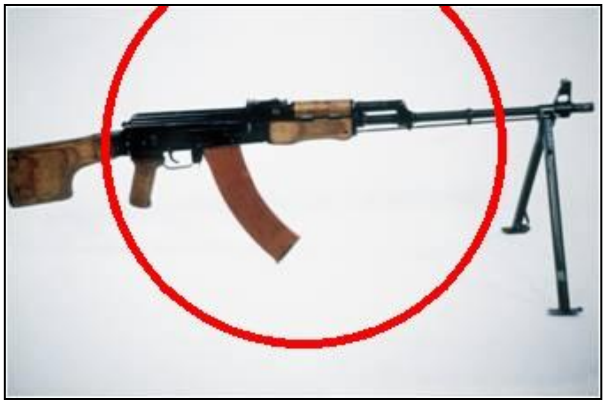

Figure 14. Methods and classifiers in this paper can also be used to find RPKS.

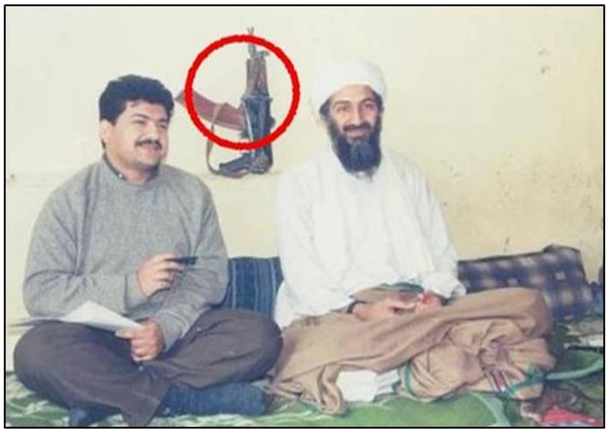

Figure 15. Methods and classifiers in this paper can also be used to find AK-74Us.

CONCLUSION

Our experiments show that parts-based Viola-Jones classifiers combined with either a support vector machine or multi-layer perceptron leverage the high recall capability of part detectors and significantly reduce false positives in comparison to both the part and whole object detectors. Classifiers trained to detect parts of an AK-47 exhibit a high recall, but a poor false positive rate when compared against classifiers trained on the whole object.

This research directly benefits modern operational forces. Intelligence analysts are increasingly reliant on imaging systems, and require capabilities to deal with the growing amounts of data produced by surveillance, collection, and forensic systems. By rapidly locating an AK-47 in a video or image, analysts can focus on exploiting suspicious media and provide timely, relevant intelligence to forces in theatre. Weapon detection in video also supports data fusion efforts and collection management functions to better automate future persistent Intelligence, Surveillance, and Reconnaissance (ISR) systems.

While initial results from these experiments demonstrate a capability, more work is needed to further increase recall and lower the amount of false positives detected. Our future work will focus on other part combination methods, training on more representative image sets, and conversion of Viola-Jones binary cascades to probabilities to produce an operational system that is needed on the front lines today.

REFERENCES

Ankur Datta, Mubarak Shah, Niels Da, and Niels Da Vitoria Lobo. Person-on-person violence detection in video data. In *In Proc. Int'l Conference on Pattern Recognition*, pages 433-438, 2002.

Antonio Torralba Kevin, Kevin P. Murphy, and William T. Freeman. Sharing features: Efficient boosting procedures for multiclass object detection. In *In CVPR*, pages 762-769, 2004.

B Leibe, E Seemann, and B Schiele. Pedestrian detection in crowded scenes. In *in Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pages 878-885. IEEE Computer Society, 2005.

Bastian Leibe, Aleš Leonardis, and Bernt Schiele. Robust object detection with interleaved categorization and segmentation. *Int. J. Comput. Vision*, 77(1-3):259-289, 2008.

D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, 2006.

Defense Industry Daily. *Too Much Information: Taming the UAV Data Explosion.* May 16, 2010. http://www.defenseindustrydaily.com/uav-data-volume-solutions-06348/ (accessed Jun 9, 2010).

E. Alpaydin. *Introduction to machine learning*. The MIT Press, 2004.

H Rowley, S Baluja, and T Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (20), 1998.

H. Takatsuka, M. Tanaka, and M. Okutomi. Distribution-Based Face Detection using Calibrated Boosted Cascade Classifier. In *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pages 351-356, 2007.

J Nam, M Alghoniemy, and A H Tewfik. Audio-visual content-based violent scene characterization. In *in International Conference on Image Processing (ICIP '98*, pages 353-357, 1998.

J. Wu, Z. Zhou, L. Zhou, and D. Hu. Learning spatial prior with automatically labeled landmarks. In *Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on*, volume 1, 2008.

J. Zhang, Y. Yan, and M. Lades. Face recognition: eigenface, elastic matching, and neural nets. *Proceedings of the IEEE*, 85(9):1423, 1997.

K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65:2005, 2005.

Krystian Mikolajczyk and Cordelia Schmid. An affine invariant interest point detector. In *In Proceedings of the 7th European Conference on Computer Vision*, pages 0-7, 2002.

K Mikolajczyk and C Schmid. Indexing based on scale invariant interestpoints. In *In Proceedings of the International Conference on Computer Vision*, pages 525-531, 2001.

K Mikolajczyk and C Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision(IJCV*, (60):63-86, 2004.

K. Mikolajczyk, C. Schmid, and A. Zisserman. Human detection based on a probabilistic assembly of robust part detectors. *Computer Vision-ECCV 2004*, pages 69-82, 2004.

K. Mikolajczyk, A. Zisserman, and C. Schmid. Shape recognition with edge-based features. In *Proceedings of the British Machine Vision Conference, Norwich, UK*. Citeseer, 2003.

M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71-86, 1991.

M. Weber, M. Welling, and P. Perona. Towards automatic discovery of object categories. In *IEEE Conference on Computer Vision and Pattern Recognition, 2000. Proceedings*, Volume 2, 2000.

N. Dalai, B. Triggs, I. Rhone-Alps, and F. Montbonnot. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, volume 1, 2005.

N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. *Computer Vision-ECCV 2006*, pages 428-441, 2006.

Nuno Vasconcelos and Andrew Lippman. Towards semantically meaningful feature spaces for the characterization of video content. In *in Proceedings of IEEE Int'l Conference on Image Processing*, pages 542-545, 1997.

P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Computer Vision—ECCV'96*, pages 43-58, 1996.

PF Felzenszwalb, RB Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. PAMI, 2009.

P Viola and M Jones. Robust real-time object detection. In *In Second international workshop on statistical and computational theories of vision - modeling, learning, computing, and sampling*, 2001.

P Viola and M J Jones. Rapid object detection using a boosted cascade of simple features. In *In CVPR01*, pages 511-518, 2001.

Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *IEEE ICIP 2002*, pages 900-903, 2002.

R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*, volume 2, 2003.

S. Agarwal, A. Awan, and D. Roth. Learning to detect objects in images via a sparse, part-based representation. 26(11):1475-1490, November 2004.

S. Du, N. Zheng, Q. You, Y. Wu, M. Yuan, and J. Wu. Rotated haar-like features for face detection with in-plane rotation. *Interactive Technologies and Sociotechnical Systems*, pages 128-137, 2006.

S Kumar and M Hebert. Multiclass discriminative fields for parts-based object detection. In *In Snowbird Learning Workshop*, 2004.

T Serre, L Wolf, S Bileschi, M Riesenhuber, and T Poggio. Object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (29):2007.

Y. Deng and G. Su. Face Detection Based on Fuzzy Cascade Classifier with Scale-invariant Features. *International Journal of Information Technology*, 12(5), 2006.

Y Kuno, T Watanabe, Y Shimosakoda, and S Nakagawa. Automated detection of human for visual surveillance system. In *Proc. of Intl. Conf. on Pattern Recognition*, pages 865-869, 1996.