# An analysis of Feature extraction and Classification Algorithms for Dangerous Object Detection

Sakib B. Kibria
Department of Electrical and Computer Engineering
North South University, Bangladesh

Mohammad S. Hasan
The School of Computing and Digital Technology
Staffordshire University, UK

***Abstract*— One of the important practical applications of object detection and image classification can be for security enhancement. If dangerous objects e.g. knives can be identified automatically, then a lot of violence can be prevented. For this purpose, various different algorithms and methods are out there that can be used. In this paper, four of them have been investigated to find out which can identify knives from a dataset of images more accurately. Among Bag of Words, HOG-SVM, CNN and pre-trained Alexnet CNN, the deep learning CNN methods are found to give best results, though they consume significantly more resources.**

***Keywords—Image Classification, Object Detection, Knife Detection, Bag of Words, SURF, HOG, SVM, CNN, Alexnet.***

## I. Introduction

Object detection and recognition from images or videos is one of the most popular and worked on branches of computer vision. It can be applied in various cases e.g. automatic traffic control, face detection and recognition, vehicle number plate identification etc. One of the sectors where this can leave a significant impact is security. At this day and age, CCTVs are very common in most public places as well as private areas e.g. homes. The footages from these cameras play a big role in solving a lot of crimes and bringing the guilty to justice. However, these are used for reactive measures i.e. to find about an incident after it has already occurred. Methods of object recognition and image classification can be used in this case to build systems that would alert the authorities when any dangerous object e.g. a knife, a gun etc. is detected in the captured image. Keeping this goal in mind, this paper investigates some of the best methods of image classification to find out which process performs best for this purpose. Among all the object detection and classification methods, Bag of Words (BoW) i.e. Speeded Up Robust Features (SURF) extractor and Support Vector Machine (SVM) classifier, Histogram of Oriented Gradients (HOG) extractor with SVM classifier, Convolution Neural Network (CNN), and pre-trained CNN with SVM have been considered and the performances are analysed.

This paper is organized as follows. Previous works in the similar field are discussed in section 0. Section II describes the background and relevant algorithms. Section III explains the experiment setup, datasets, software and algorithms used. The results are shown and analyzed in section IV, and finally, some conclusions are drawn in section V.

## Existing Works

Quite a lot of work has been done on the subject of object detection and image classification. The article [1] describes a system for detecting knives and handguns from CCTV image. Here feature extraction has been done from MPEG 7 video and SVM has been used for classification. In the article [2], the authors have compared Scale-Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Convolutional Neural Network (CNN) algorithms for image recognition, among them CNN performed the best. Article [3] compares SIFT, KAZE, Accelerated KAZE (AKAZE) and Oriented Fast and Rotated Brief (ORB). In general, SIFT performs best according to their tests, but ORB is the fastest. Article [4] conducts a comparison of SIFT and SURF algorithms based on face detection. The results describe that even though SIFT is more robust, SURF is much faster and has good accuracy rate. Object detection and recognition methods using SURF and BoW are described in the article [5]. According to its results, while using BoW method, SURF feature extractor and SVM classifier give best results for recognition. Hence BoW method is chosen as one of the aspects of comparison in this paper. Another popular feature extractor is HOG which can be used with SVM for object detection. In the article [6], the authors have used this technique for detection of vehicle logo. An alternate way of object detection is deep neural networks, as [7] uses CNN and SVM for face detection. The authors have shown that CNN and SVM give better results than other popular methods for face recognition. Article [8] also uses a hybrid CNN-SVM approach recognition of hand-written digits. Another possible approach for image classification is using a pre-trained CNN. Article [9] shows the high accuracy of the Alex-Net neural network. This paper investigates the performances of BoW, HOG, SVM, CNN etc. to detect knives in images.

## II. Preliminaries

Widely used classifiers with the high accuracy are chosen for this research and are explained in the following sub-sections.

### *A. Bag of Words (BoW)*

BoW is a method in computer vision which categorizes image features as ‘words’ and is applied for image classification. It generates a histogram of visual word occurrences that represent image [10]. BoW follows three steps: feature detection, feature description, and codebook generation [11]. For feature extraction, by default, bagOfFeatures class in Matlab extracts upright SURF features [12].

SURF follows three main steps, first 'interest points' need to be identified from different distinctive locations of the image [13]. This detection is based on Hessian matrix and blob-like structures are identified from different parts of the image. The algorithm puts emphasis on the scalability of the image and the features. Next, the neighbourhood of every interest point is represented by a feature vector. Finally, the descriptor vectors are matched between different images.

The BoW method then creates codebook from these feature vectors obtained from SURF descriptor through k-means clustering. Finally, an SVM is trained with these features for classification.

### B. *Support Vector Machine (SVM)*

SVM is a supervised machine learning algorithm. Given labeled training data, the algorithm outputs an optimal hyperplane which categorizes new examples [14]. There are quite a few kernel versions of SVM, in this paper 'linear' SVM is used in every case. Linear SVM maximizes the margins from nearest training point [15].

### C. *Histogram of Oriented Gradients (HOG)*

HOG is another feature descriptor like SURF and is used in many kinds of image recognition and detection purposes. This algorithm counts occurrences of edge orientations in a local neighborhood of an image [16]. The HOG descriptor represents a local statistic of the orientations for the image gradients for a keypoint. In other words, each descriptor is a collection of histograms composed of pixel orientations given by their gradients [17].

### D. *Convolutional Neural Network (CNN)*

CNN is a type of deep learning algorithm that is efficiently and widely used in computer vision. CNN usually has a combination of convolutional, pooling or fully connected layers between an input and an output layer [18].

### E. *Alexnet*

For image classification and recognition, using a pre-trained neural network can give much better results in case of deep learning [19]. One of the most suitable pre-trained networks for image classification purposes is Alexnet. Alexnet is a deep CNN created by Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton which is trained with over a million images from 1000 different classes [20]. Alexnet consists of five convolutional layers and three fully connected layers [21].

## III. Experimental Setup

Figure 1 shows all the steps of the experiment. The first step is to extract features from these images. Next, a classifier is trained with these features to recognize and classify them correctly. Finally, the classifier is able to predict the category of an input image. The classifier tested with a test set to view its accuracy.

### A. *Environment*

For this experiment, all the algorithms are implemented in Matlab [22]. The version used is R2017a, with Computer Vision Toolbox, Machine Learning Toolbox, Neural Network Toolbox and Parallel Processing Toolbox installed.

### B. *Data Set*

A large number of images are used as data set for this experiment. As an example of a dangerous object, knife images have been chosen. There are 1000 images of knives and 1000 background images without knives in total. The images are obtained from the database provided in [1]. Figure 2 shows examples of knife images and background images used in this paper for the experiments. The knife images are cropped from CCTV footage frames. The knives that are considered in this paper are the daily-use ones and are easily accessible from the supermarkets.

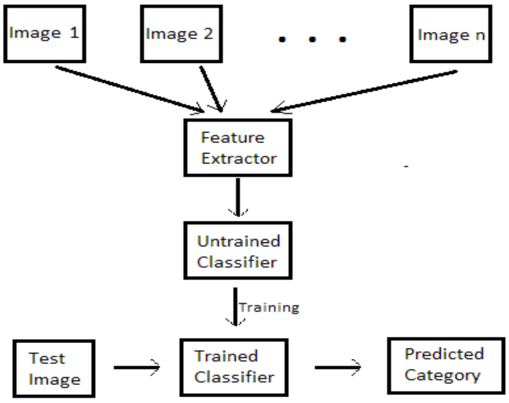

Figure 1: Experiment workflow

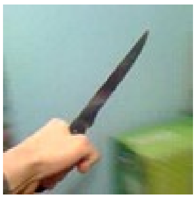
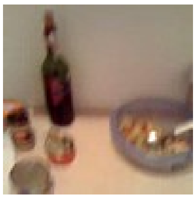

Figure 2: Sample image with knife and background image without knife [23].

### C. *Algorithms*

Four algorithms have been investigated in this paper.

1. BoW: SURF feature descriptor is used with linear SVM in the BoW method.

2. HOG+SVM: Features are extracted with HOG features with cell size 4x4. Linear SVM is used as the classifier.

3. CNN: Deep convolutional network used for classification with MaxEpochs set to 40. Parallel processing used to run the algorithm simultaneously over all the available CPUs.

4. Alexnet+SVM: Alexnet has been used as a feature extractor with SVM as a classifier. Alexnet has default image size set to 227 by 227. Hence the images are resized to match.

## IV. Result and Analysis

All the algorithms have been tested in two ways. A dataset of 1000 images with knives and 1000 images without knives has been considered. The performances of the algorithms are investigated in two ways – random data set and two-times-two-folds validation. In random data set, each algorithm takes 500 images randomly from each category to train and 500 from each to test. Then, two-times-two-folds validation has been done for further analysis.

### *A. Random data set*

In this case, all the tests have been done twice and the values have been averaged. From Figure 3 it can be seen that apart from HOG+SVM, all three algorithms have given quite a high accuracy in correctly identifying images with knives. However, highest accuracy is achieved by pre-trained Alexnet. As mentioned section II.E, Alexnet is pre-trained with over one million images for the specific purpose of image classification. This is probably the reason behind the high accuracy it gives. HOG+SVM have given the lowest accuracy. The reason for this could be that HOG cannot extract enough number of features from these kinds of images, so the detection accuracy is low.

To further analyze the results the number of total wrong predictions i.e. False Positives (FP) and False Negatives (FN) are looked at. If the classifier predicts the presence of a target in the image while actually it is absent, it is called FP. Similarly, if the prediction is the absence of a target while it is actually present, the result is FN. Figure 4 shows that Alexnet+SVM has the lowest FP and FN rates. Low FN rate is very important in this case as it means images with knives have not been identified which can be dangerous in real life situations.

Though from numerical results it is clear that pre-trained Alexnet+SVM performs the best, another factor that needs to be considered is the time required to run the algorithms. Figure 5 shows the time taken by each algorithm for training and testing. HOG+SVM takes very little time though the accuracy is not very good as shown Figure 3. BoW i.e. SURF and SVM technique takes more time than HOG, but it is nothing close to the neural networks. Untrained CNN takes very long time to train while extracting features with pre-trained Alexnet and classifying with SVM takes much less.

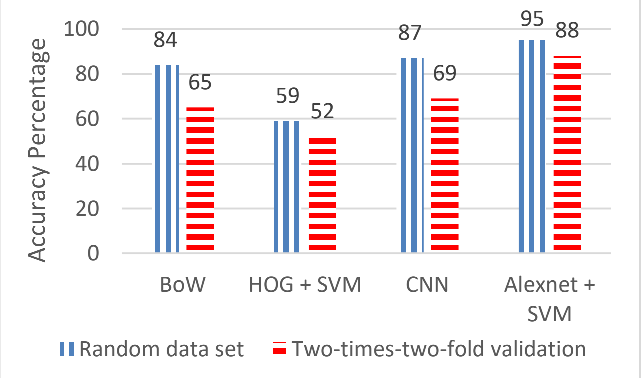

Figure 3: Accuracy levels for a random data set (average of 2 instances) and two-times-two-folds validation.

Figure 5 also shows the testing time for each algorithm. This is also important because after the training, if these algorithms are put to use for building a real-time system, the detection will need to be very quick. Here it shows HOG+SVM takes least time for testing and very similar to training. The interesting result is that even though CNN took a very long time to train compared to other algorithms, its testing time is significantly low. Alexnet+SVM has testing time almost same as training time. This might be a problem in case of real-time systems.

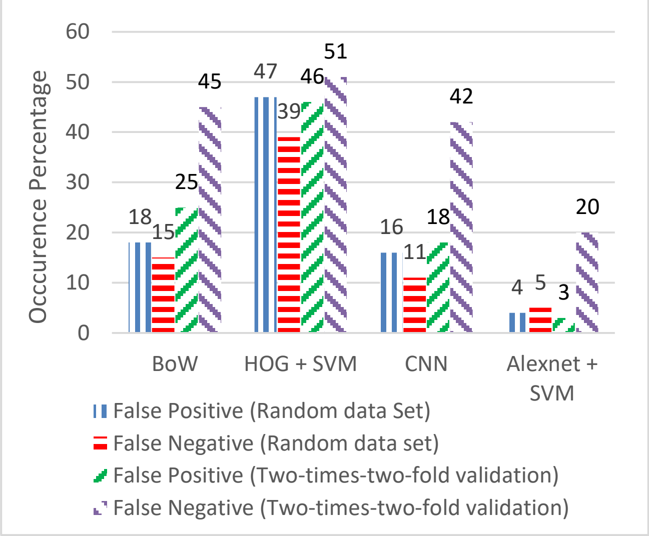

Figure 4: FP and FN rates for a random data set (average of 2 instances) and two-times-two-folds validation.

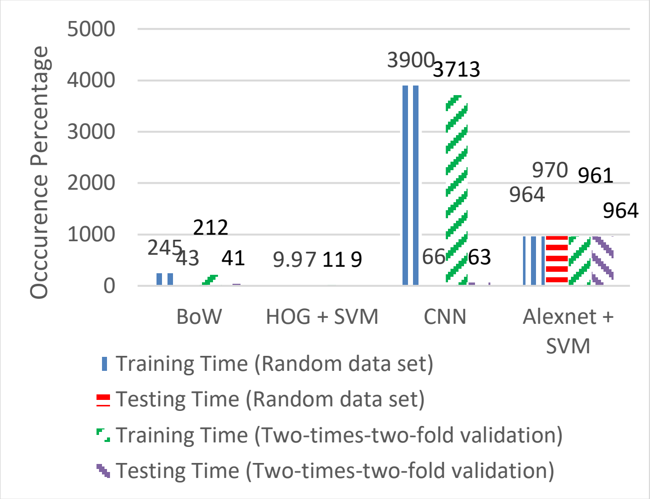

Figure 5: Time Taken for training and testing for a random data set and two-times-two-folds validation.

### *B. Two-times-two-folds validation*

In this case, the dataset is divided into two partitions. The first half is used as the training set and the second half as testing set during the first instance. The second time, the sets are reversed and the first is used for testing while the second for training.

Figure 3 shows the accuracy levels of the four algorithms.

All the accuracies are lower than the random data set results. However, the order of higher accuracy is still maintained in the same exact way i.e. Alexnet+SVM still gives the highest accuracy.

Figure 4 shows that FN rates have increased in each case. However, Alexnet+SVM have the lowest FN rate which means it has missed the least number of pictures with knives. Figure 5 shows the time consumption of the algorithms for the two-times-two-folds validation.

These results do not differ much from the randomly selected data set which is expected as though they run on differently organized data sets, an equal number of training and testing images are used in each case.

The decrease in overall accuracy in case of two-times-two-folds validation might indicate that when it comes to a large variety of images, the feature extraction is not as accurate for some of them. Alexnet can extract correct features from even poor-quality images better than other algorithms.

## V. Conclusion

Although there are many more methods of image recognition and classification, the ones mentioned here are quite well known and successfully applied. Here they have been compared to do a specific task, identify knives in images. For this purpose, deep learning CNN based methods have shown best performance in terms of accuracy, which includes both pre-trained and untrained CNN. The use of pre-trained Alexnet along with SVM gives the best performance among all, but BoW method also gives quite a high accuracy. However, when it comes to time consumed, BoW method is significantly ahead of the neural networks. Using only CNN for classification can take a lot of time to train but very low time for identification, while the testing time for Alexnet based method is quite high. The neural networks are also run across multiple CPUs, so they consume a lot more resources. Overall using pre-trained Alexnet with SVM gives the best performance.

## Acknowledgment

The authors would like to acknowledge an Erasmus+ International Credit Mobility (ICM) fund for Bangladesh awarded to Dr Mohammad Hasan at Staffordshire University, UK in 2016.

## Reference

[1] M. Grega, A. Matiolański, P. Guzik, and M. Leszczuk, "Automated Detection of Firearms and Knives in a CCTV Image," *Sensors*, vol. 16, no. 1, p. 47, Jan. 2016.

[2] S. Achatz, "State of the Art of Object Recognition Techniques," *Sci. Semin. Neurocientific Syst. Theory*, 2016.

[3] O. Andersson and S. R. Marquez, "A comparison of object detection algorithms using unmanipulated testing images Comparing SIFT , KAZE , AKAZE and ORB En j ¨ amf ¨ or andandet av icke-datormanipulerade bilder," 2016.

[4] S. Jain, B. L. S. Kumar, and R. Shettigar, "Comparative study on SIFT and SURF face feature descriptors," in *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2017, pp. 200–205.

[5] J. Farooq, "Object detection and identification using SURF and BoW model," in *2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, 2016, pp. 318–323.

[6] D. F. Llorca, R. Arroyo, and M. A. Sotelo, "Vehicle logo recognition in traffic images using HOG features and SVM," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2013, pp. 2229–2234.

[7] Q.-Q. Tao, S. Zhan, X.-H. Li, and T. Kurihara, "Robust face detection using local CNN and SVM based on kernel combination," *Neurocomputing*, vol. 211, pp. 98–105, Oct. 2016.

[8] X.-X. Niu and C. Y. Suen, "A novel hybrid CNN–SVM classifier for recognizing handwritten digits," *Pattern Recognit.*, vol. 45, no. 4, pp. 1318–1325, Apr. 2012.

[9] Jing Sun, Xibiao Cai, Fuming Sun, and J. Zhang, "Scene image classification method based on Alex-Net model," in *2016 3rd International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS)*, 2016, pp. 363–367.

[10] "Image Classification with Bag of Visual Words - MATLAB &amp; Simulink - MathWorks United Kingdom." [Online]. Available: https://uk.mathworks.com/help/vision/ug/image-classification-with-bag-of-visual-words.html. [Accessed: 19-Aug-2017].

[11] Fei-Fei Li and P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 524–531.

[12] "Create a Custom Feature Extractor - MATLAB &amp; Simulink - MathWorks United Kingdom." [Online]. Available: http://uk.mathworks.com/help/vision/ug/create-a-custom-feature-extractor.html. [Accessed: 19-Aug-2017].

[13] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[14] "Introduction to Support Vector Machines — OpenCV 2.4.13.3 documentation." [Online]. Available: http://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html. [Accessed: 19-Aug-2017].

[15] "Support vector machines: The linearly separable case." [Online]. Available: https://nlp.stanford.edu/IR-book/html/htmledition/support-vector-machines-the-linearly-separable-case-1.html. [Accessed: 19-Aug-2017].

[16] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using Histograms of Oriented Gradients," *Pattern Recognit. Lett.*, vol. 32, no. 12, pp. 1598–1603, Sep. 2011.

[17] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using HOG–EBGM," *Pattern Recognit. Lett.*, vol. 29, no. 10, pp. 1537–1543, Jul. 2008.

[18] "Convolutional Neural Network - MATLAB &amp; Simulink." [Online]. Available: https://uk.mathworks.com/discovery/convolutional-neural-network.html. [Accessed: 20-Aug-2017].

[19] D. GUPTA, "Transfer learning &amp; The art of using Pre-trained Models in Deep Learning," *JUNE 1*, 2017. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/06/transfer-learning-the-art-of-fine-tuning-a-pre-trained-model/. [Accessed: 20-Aug-2017].

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *NIPS'12 Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, pp. 1097–1105, 2012.

[21] H. Lee and H. Kwon, "Going Deeper With Contextual CNN for Hyperspectral Image Classification," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4843–4855, Oct. 2017.

[22] "MATLAB - MathWorks." [Online]. Available: https://uk.mathworks.com/products/matlab.html. [Accessed: 19-Aug-2017].

[23] "Knives Images Database," Katedra Telekomunikacji AGH. [Online]. Available: http://kt.agh.edu.pl/matiolanski/KnivesImagesDatabase/. [Accessed: 19-Sep-2017].