Recognition of Bullet Holes Based on Video Image Analysis

Zhu Ruolin 1,2, Liu Jianbo 1, Zhang Yuan 1 and Wu Xiaoyu 1

1School of Information Engineering, Communication University of China, Beijing 10024, China
2zhuruolin@cuc.edu.cn

Abstract. The technology of computer vision is used in the training of military shooting. In order to overcome the limitation of the bullet holes recognition using Video Image Analysis that exists over-detection or leak-detection, this paper adopts the support vector machine algorithm and convolutional neural network to extract and recognize Bullet Holes in the digital video and compares their performance. It extracts HOG characteristics of bullet holes and train SVM classifier quickly, though the target is under outdoor environment. Experiments show that support vector machine algorithm used in this paper realize a fast and efficient extraction and recognition of bullet holes, improving the efficiency of shooting training.

1. Introduction

Machine learning is a research focus in the field of AI(artificial intelligence) currently. Video Image Analysis based on machine learning is applied in a variety of fields recently, such as biomedical image analysis, vehicle target recognition, expression recognition and character recognition. This paper is aimed at extracting and recognizing Bullet Holes from videos about the training of military shooting. Video Image Analysis can get the distribution of the point of impact during the shooting training of soldiers, and can remind us to take suitable training means to improve the military shooting level of soldiers.

Recognition of bullet holes in video has not only the advantages of high precision, low cost etc, but also the challenging owing to the background region and the noisy stream[1]. Nowadays the recognition of bullet holes mostly uses shallow machine learning methods, and normally their research objects is Chest round target. Liu Q and Chen Y[2] have calculated the position of bullet holes based on threshold segmentation and Boosted Cascade algorithm. Recently, deep learning become hot topics in Video Image Analysis, because of its strong ability to extract robust features. Deep learning has great potential in the field of image recognition.

This paper studies the image recognition of bullet holes based on shallow machine learning and deep learning method. According to the actual shooting situation, we design two algorithms to recognize bullet holes. The first one is based on rough positioning, Histogram of Oriented Gradients (HOG) and Support vector machine(SVM)[3]. The other applies rough positioning and deep convolution neural network method for comparison experiments. Both algorithms require rough positioning of Bullet holes, which contains extracting frames, Otus algorithms, morphological filtering and region labeling process. The process of bullet holes positioning is shown in Figure 1.
2. Bullet holes rough positioning

2.1. Preprocessing
Before training, we need to preprocess the videos because of the noise interference. In this paper, the experimental data is the long-range shooting video for outdoor training with a variety of noise pollution. The target for long-range shooting is larger than normal target. As a consequence, the pixels of bullet holes in the video frame is much less. And the target is made of a soft cloth, which will be affected by wind and other external conditions readily.

First of all, We need to filter out environmental noise in order to get the position of bullet holes accurately, according to a prior knowledge about bullet holes. The median filter is used to filter out the ambient noise. After that, this paper uses Otsu method morphological filtering and region labeling process to complete the rough positioning of the bullet hole.

2.2. Rough positioning based on Ostu algorithm
Otsu method is based on the single order statistical characteristics of gray histogram, and makes maximizing the variance between classes of target regions and background as threshold selection criterion. In this paper, target regions in shooting video is regions of the bullet holes. Figure 2(a) shows results after the Otus algorithm.

But to shooting video frames, when background is a little complex and the ratio of bullet holes to background is very small, the Otsu method will falsely segment the less area of background into the results. So with considering the real-time processing requirement of the shooting videos, in this paper, Otsu method is only used for rough positioning of the actual bullet holes.

Then a part of areas that is too big or too small is removed by morphological filtering and region labeling process. Finally, record the suspected bullet holes as suspected images for recognition. Figure 2(b) draws suspicious samples on a frame image.
3. The recognition of bullet hole region based on shallow machine learning

3.1. Database construction
Training samples, in a way, impact on the algorithm’s generalization ability. The experimental idea of this paper is to recognize actual bullet holes from suspected images which obtained in the previous step. Considering in the actual shooting videos may have different ambient interference, we recorded videos in actual military shooting practice outdoor taken in natural backgrounds. And then make bullet holes rough positioning to videos that we recorded. After obtaining some suspected bullet holes areas by rough positioning, calculate the center of this connected areas, and cut the 24×24 image proposals on the original video frame with the center coordinates. Finally we obtain a total of 890 samples, of which 442 positive samples, while others are negative. The positive samples of database are shown in Figure 3(b), while Figure 3(a) are negative samples.

3.2. The average gradient map
In order to ensure the distinguishing ability of HOG features, in the experiment, we draw the average gradient image and implements the visualization of the HOG features. When calculating HOG features, each sample produces a gradient image [4]. The gradient images of all positive samples from database are taken in average, and then normalized. The average gradient image of positive samples is shown in Figure 4. As average gradient image show, the characteristics of the bullet holes can be properly described by gradient.
Visualizing HOG features, each direction corresponds to a bin [4], while the length is the gradient projection in this direction. Used to say that overlapping blocks lead to multiple updates of this sum, therefore we keep track how often a cell was updated to compute average gradient strengths. In the experiment, we change the size of cell and value of bin so as to obtain the best dimensional HOG features. Comparing the four pairs of in Figure 5, when size of cell choose 8×8 pixels and bin is 9, the characteristics of the positive and negative samples are the most distinguished. Figure 5(c) illustrates that this gradient projection are evenly distributed in all directions.

Figure 4. The average gradient image of positive samples

(a) The average gradient image of dx

(b) The average gradient image of dy

Figure 5. The visualization of the HOG features

(a) when size of cell is 3×3 pixels and bin is 9, left one is the HOG visualization of positive samples, right one shows negative samples.

(b) when size of cell is 4×4 pixels and bin is 9, left one is the HOG visualization of positive samples, right one shows negative samples.

(c) when size of cell is 8×8 pixels and bin is 9, left one is the HOG visualization of positive samples, right one shows negative samples.

(d) when size of cell is 8×8 pixels and bin is 12, left one is the HOG visualization of positive samples, right one shows negative samples.

(e) when size of cell is 8×8 pixels and bin is 18, left one is the HOG visualization of positive samples, right one shows negative samples.
3.3. The recognition of bullet hole region based on HOG and SVM
As Figure 5 shown, negative samples are similar to bullet holes showing Intense changes in gradients, but differ in that this gradients are evenly distributed in all directions. HOG can properly describe the difference by histogram of oriented gradients. Histogram of oriented gradients contains measurements of direction and magnitude in each direction.

We propose HOG+SVM algorithms. In this experiment, HOG is used to extract gradients features from input images. Set block to 16×16 pixels and the stride to 4×4, so that each cell which is 8×8 resolution can be overlaped by some others. SVM is a typical binary classification algorithm[5], using it to distinguish between the true bullet holes and the others.

4. The recognition of books based on deep learning
Recent studies in image analysis have shown that the performance of convolutional neural networks (CNNs) is superior to that of shallow machine learning[6], owing to its capable of learning the essence of data. The network structure in deep learning is composed of the convolution layers of which the number is normally 3 or 5 or even more, pooling layers and the Fully connected layers. In this paper, we use the Faster R-CNN model that trained on the VOC2007 database[7] and then retrain it with the database we built ourself of bullet holes. Then images can be recognized and classified by the retrained model.

4.1. Training samples
In this experiment, training data choose the all positive samples from database which we built for bullet holes. But the experiments use images of 500×500 resolution for training, because the sample from the database is too small to locate bullet holes accurately. The samples of 24×24 resolution are amplified to images 500×500 resolution. Then label the images, putting them into network. The network will randomly select the size of crop_size × crop_size resolution as input.

4.2. Model Parameters and Experimental Process
Faster R-CNN is composed of two modules. The first module is a deep fully convolutional network called region proposal network(RPN), and the second module is the Fast R-CNN detector that uses the proposed regions. During the initializing phase, we set learning rate of 0.001 for 60 percent mini-batches, and 0.0001 for the others mini-batches on the database, the weight decay of 0.0005 and a momentum of 0.9. The weights are randomly initialized with zero-mean Gaussian function whose standard deviation 0.01. The Faster R-CNN model that adopted in this paper is shown in Figure 6.

![Figure 6. The Faster R-CNN model adopted in the experiment.](image)

5. The comparison and analysis of two recognition methods
The testing samples using in the experiments are different from training database. We record a fresh video for Evaluating the performance of two algorithms. Select 160 frames from the vedio, and make a rough positioning.

This paper respectively adopts HOG+SVM and Faster R-CNN in deep learning algorithms, realizing the video image analysis of bullet holes. The recognition results are shown in Table 1.
Table 1. The comparison of two methods for the recognition of bullet holes

|                | Fp | Tp  | Recall   | Precision |
|----------------|----|-----|----------|-----------|
| HOG+SVM        | 30 | 2081| 0.9294   | 0.9857    |
| Faster R-CNN   | 358| 2085| 0.9312   | 0.8535    |

For the Recall, the deep learning method has shown a good performance, comparing the SVM. But for the Precision, performance of the shallow machine learning method is much better than it that Faster R-CNN shows. It means that features of bullet holes Faster R-CNN learned is not robust enough.

The reasons for this result may be that bullet holes are too small to study the inherent features, and the size of the training sample is inappropriate. To the samples of 24×24 resolutions, convolutional neural network can not get enough negative samples from whole frame images, while negative samples for SVM is different. For Faster R-CNN, to the small target such as bullet holes in this experiment, target image block can be shrunk to a point after several convolution operations. So the last layer of convolution layer has lost some of the features of bullet holes.

6. Conclusion
This paper studies two main methods which are shallow machine learning and deep learning algorithms. According the result of comparison experiments, analyze the recognition results from the aspects of Recall and Precision. In overall consideration, HOG+SVM algorithm is more suitable for the outdoor training of military shooting. The next step is to use information of adjacent frames in the video optimizing the algorithms.

Acknowledgments
The work on this paper was supported by the National Key Technology R&D Program (2015BAK22B02) and the Program (JXJYG1606) of Communication University of China.

References
[1] Zhou J 2014 Study of Military Automatic Target-scoring System Based on Machine Vision Technology
[2] Liu Q and Chen Y 2015 Research on Automatic Target-scoring Method Based on Video Image Analysis Computer Engineering
[3] Luo J, Zhang Z, Zeng G 2016 The study of multi-class classification based on support vector machine algorithm ITOEC 189
[4] Dalal N and Triggs B 2005 Histograms of Oriented Gradients for Human Detection l 1.7w p 886-893
[5] Zhu B, Yang L, Wu X and Guo T 2015 Automatic Recognition of Books Based on Machine Learning International Symposium on Computational & Busi.pp 74-78
[6] Tang F, Wang Z and Chen M 2005 The study of multi-class classification based on support vector machine algorithm Control and Decision Vol.20 No.7 pp 746-754
[7] Ren S, He K, Ross G and Sun J 2016 Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks Microsoft Research arXiv:1506.01497v2 200