Anomaly Detection Based on Locality Sensitive Hashing with Genetic Algorithm

Tao Zhang1,*, Zijuan Fan2 and Xiangyu Yu2

1School of Electronic and Communication Engineering, Guizhou University, China
2School of Electronic and Information Engineering, South China University of Technology, China

*Corresponding author

Abstract. With the increase of public safety awareness, video anomaly detection has attracted researchers' attention. In the paper, a novel approach is proposed to detect anomalies in the video. It is based on Locality Sensitive Hashing (LSH), which maps similar data to the same bucket with high probabilities, and non-similar data is mapped to the same bucket with a low probability to detect abnormal videos that are not similar to normal videos. In order to improve the probability of similar data mapping into the same bucket, the Genetic Algorithm (GA) is used to optimize the entire hash function group while maintaining the diversity of the hash function group. The algorithm gets AUC 0.78 on the dataset UCSD ped1 and AUC 0.94 on the dataset UCSD ped2, which confirmed the effectiveness of the algorithm.

Keywords: Anomaly detection; Locality sensitive hashing; Genetic algorithm.

1. Introduction

Intelligent video surveillance, which can provide video analysis like anomaly detection and object tracking, is crucial for the security of society. As an important part of intelligent video surveillance technology, video anomaly detection refers to the identification of events that do not conform to expected behavior[6]. However, unlike other fields in computer vision like object classification, video anomaly detection is challenging for the unbalanced amount of normal events and abnormal events, which is rare compared to the normal events and difficult to collect as training dataset[16]. Given the condition that anomaly events are usually missing from training dataset[23, 22], many researchers focus on defining a model, which would represent the normal events based on the training dataset, and detecting the abnormal events in testing dataset by finding the instance that does not follow the defined model like the normal events do. To learn a set of representative features based on the normal events in training dataset, autoencoder[10], as an efficient unsupervised manner which learns data representation by reconstructing its own input, is widely used in the field of video anomaly detection[27, 26]. For example, in[25, 15, 21], researchers adopt autoencoder to represent normal events and try to detect the unseen abnormal events in testing dataset by following the principle that reconstruction error of abnormal events would be larger than that of normal events. With the development of deep learning, many researches are carried out to build an autoencoder for video anomaly detection with convolutional neural network, like Convolutional AutoEncoder (ConvAE)[15], Stacked Recurrent Neural Network (SRNN)[22] and 3D Convolutional AutoEncoder (3DCAE)[30]. However, reconstruction based approaches may lead to an over complete model, in which not only the normal events but also the abnormal ones are successfully reconstructed[19]. It means that it is
difficult to distinguish abnormal events from normal ones in an over complete reconstruction based model.

Some researchers try to describe normal events with statistical models such as Social Force Model (SFM)[24], Bayesian Network (BN)[20, 3] and Hidden Markov Model (HMM)[1, 18]. Mehran et al.[24] introduce SFM in the field of video anomaly detection to map frames into force flow, which is used to model the normal behavior. In [20], Chen et al. use a Cascade of Dynamic Bayesian Networks (CasDBNs) for modeling the decomposed complex behavior pattern which is build based on temporal characteristics or spatial-temporal visual contexts. In [18], Kratz and Nishino adopt the distribution based HMM and a coupled HMM to capture the temporal relationship between local spatio-temporal motion patterns and the spatial relationship separately. In statistical model based approaches, statistical inference is usually adopted to determine the given instance is normal or not. For instance, Kim et al.[17] use a space-time Markov Random Field to describe the normal events and compute the maximum a posterior probability of the given instance to determine if it is normal or not. Kratz and Nishino[18] compute the confidence measure for the given instance as the maximum likelihood base on the prototypical distributions. However, statistical model based methods require prior knowledge for the target application which means a statistical model for some application would not be applicable to other applications[14].

Distance based approaches which do not require prior knowledge about distribution of target applications are also popular in the field of video anomaly detection. For example, K-Nearest Neighbor (KNN) distance[7, 4] and the sum of KNN distance[2, 28] can be adopted to measure the difference between the given instance and normal instances in training dataset. However the computational cost and memory requirement for distance based approaches increase as the amount of normal events in training dataset increases, which is making it difficult to apply these approaches to a large scale problem. To overcome the difficulty, Locality Sensitive Hashing (LSH), an algorithm for solving the approximate near neighbor search in high dimensional spaces, can be used as efficient alternative for distance based approach. Hachiya and Matsugu[14] made improvements on LSH for video anomaly detection and proposed a new hashing scheme called Normality Sensitive Hashing (NSH), which select a set of hashing functions to define a normal region. Zhang et al.[29] use LSH filters to hash normal regions in training dataset and introduce the evaluation function to evaluate the hash functions and the Particle Swarm Optimization (PSO) method to search for the optimal LSH functions.

In this paper, we propose a LSH based video anomaly detection algorithm where Genetic Algorithm (GA) is adopted to optimize the LSH groups. With the optimized LSH functions, the training data are mapped to a list of buckets. For the test data, find the most similar training bucket, and calculate the anomaly score based on the radius of the bucket and the distance from the test data to the bucket center. The main contribution of the proposed method is that we can use a GA-optimized LSH group to model features which improves the accuracy of unsupervised clustering.

The rest of the paper is organized as follows. The background of relevant knowledge is given in Section 2. In Section 3, we introduce the proposed method which contains the overview of the algorithm and explanation of all modules. The experimental results on the dataset UCSD are provided in Section 4. Finally, the conclusions are drawn in Section 5.

2. Background

2.1. Locality Sensitive Hashing (LSH)

LSH is a fast nearest neighbor search algorithm for massive high-dimensional data. The key idea is to map similar data to the same hash value with a high probability, and the unsimilar data is mapped to the same hash value with a low probability. The LSH family[13] is defined as:

Definition 1. A family $H = \{h : S \rightarrow U\}$ is called $(d_1, d_2, p_1, p_2)$-sensitive for $D$ if for any $x_1, x_2 \in S$

1. if $D(x_1, x_2) \leq d_1$, then $Pr[h(x_1) = h(x_2)] \geq p_1$
2. if $D(x_1, x_2) \geq d_2$, then $Pr[h(x_1) = h(x_2)] \leq p_2$

where $x_1, x_2$ are two multidimensional vectors, and $D(x_1, x_2)$ is the degree of dissimilarity of two vectors.

Currently, anomaly is defined as the pattern that is not similar to normal. Therefore, by using LSH modeling for normal video, it is possible to detect abnormal data that is not similar to normal video. The algorithm design of LSH is different for different similarity measurement methods. For Jaccard coefficient the min-hash is used and for Euclidean distance we adopt P-stable hash. P-stable hash is adopted in the paper.

P-stable hash is based on a p-stable distribution[8]. For a distribution $D$, if $p \geq 0$, for any $v_i \in R$ and any $X_i \in D, i = 1, 2, \ldots, n$, with the same distribution, $D$ is called a p-stable distribution. When $p = 2$, the distribution $D$ is a standard normal distribution. Therefore, the random variable $X_i$ can be randomly selected from the standard normal distribution to estimate the value of $\|v\|_2$. A hash function based on p-stable distribution is defined as

$$h_{a,b}(x) = \left\lfloor \frac{a^T x + b}{r} \right\rfloor$$

where $x$ is a $d$-dimensional vector, $a$ is a $d$-dimensional vector with entries chosen independently from a p-stable distribution, $b \in (0, r)$ is a random number and $r$ is the segment length of the line and $\lfloor \cdot \rfloor$ represents the round down operator. It is equivalent to mapping points in space onto a line and segmenting the line by length $r$. The data of the same segment is mapped to the same hash value. It is not difficult to understand that if two points in space are closer together, the probability that they are mapped to the same segment will be higher.

Generally, in order to improve the robustness, L groups of hash functions are randomly selected, defined as a population of the hash function $P_L = \{g_1, g_2, \ldots, g_L\}$, and each group of hash functions $g$ includes k hash functions $g = \{h_1, h_2, \ldots, h_k\}$. If the two data have exactly one set of hash values, then the two data are judged to be similar. The hash function group $g$ can also be written as

$$g_{A,b}(x) = \left\lfloor \frac{A^T x + b}{r} \right\rfloor$$

where $A = \{a_1, a_2, ..., a_k\} \in R^{d \times k}$ and $b$ is a $k$-dimensional vector with each element chosen uniformly from the range $[0, r]$.

2.2. Genetic Algorithm (GA)

GA[9, 11] is a method of searching for optimal solutions by simulating natural evolutionary processes, first proposed by Professor J. Holland in 1975. Each individual in the population in the genetic algorithm is a feasible solution in the solution space. By simulating the evolution process of the organism, the optimal solution is searched in the solution space. The genetic algorithm calculates the fitness value of the individuals in the population, selects the individuals with high fitness values from the population to carry out the crossover, mutation and replication operations to form the next generation population, and eliminates the individuals with low fitness values. After multiple generations of iteration, the whole population is close to the most excellent solution.

3. The Proposed Method

Our algorithm is based on the paper [29] to make improvements. The overview of our approach is shown in Figure 1. The video is divided into cuboids. The Histogram of Optical Flow (HOF) are
extracted for each cuboid. Then, the optimal LSH function group \( P^* \) is searched by the GA, and training video frames are converted into multiple buckets by \( P^* \). In the test phase, test video cuboids are also converted into test buckets. The most similar training bucket is found. The abnormality score of the test video cuboid is calculated according to the radius of the bucket and the distance from the test data to the bucket center. Finally, the threshold is set to determine the abnormality.

**Figure 1.** Overview of the proposed approach.

### 3.1. Video Feature Extraction

Firstly, the video is divided into non-overlapping spatio-temporal cuboids with size \( T \times H \times W \). Calculate the variance of the time dimension of each cuboid, and set a certain threshold. If it is greater than the threshold, it is judged as the foreground cuboid, otherwise it is the background cuboid. The prospect of extraction is shown in the **Figure 2**.

The optical flow is calculated for each pixel of the foreground cuboids[12], and the HOF of each cuboid is extracted to describe the activity of each cuboid.

**Figure 2.** Foreground map of the video.

### 3.2. Our Algorithm

Assume that the HOF is extracted from the training set is \( X = \{ x_1, x_2, \ldots, x_N \} \in R^{d \times k} \), which can be mapped to a list of buckets \( B = \{ b_1, b_2, \ldots, b_M \} \) through a set of hash functions \( g \). Since the LSH is probabilistic, in order to make similar data have a higher probability of falling into the same bucket, it is necessary to search for a “good” hash function while ensuring the difference between the hash functions of each group. In the paper, the GA is used to search for the optimal hash function population with diversity. We use the evaluation function in the paper[29] to evaluate the merits of the
hash function by the Scatter Within Buckets (SWB) and the Scatter Between Buckets (SBB). A “good” hash function should map data with similar distances into the same bucket, and map data with far distances into different buckets. That is, the SWB is small enough, and the SBB is large enough. So, the evaluation score of \( g \) is defined by

\[
S(g_{A,b}) = \frac{SBB(g_{A,b})}{SWB(g_{A,b})}
\]  

(3)

\[
SWB(g_{A,b}) = \sum_{m=1}^{M} \sum_{x \in B_m} (x - c_m)^T (x - c_m)
\]  

(4)

\[
SBB(g_{A,b}) = \sum_{m=1}^{M} N_m (c_m - c)^T (c_m - c)
\]  

(5)

where \( c_m = \frac{1}{N_m} \sum_{x \in B_m} x \) is the center of the bucket \( B_m \), \( N_m \) is the number of points falling into the bucket \( B_m \) and \( c = \sum_{i=1}^{N} x_i \) is the center of all data.

This evaluation function serves as a fitness function of the GA. The entire population needs to evolve toward a large fitness value, and in order to ensure the diversity of the population, it is also necessary to maintain a large variance of the fitness values of the entire population. Since \( r \) is a parameter according to the training set, \( b \in (0, r) \) is a random number, so one feasible solution of the population is defined as \( A \), and \( A \in \mathbb{R}^{d \times k} \) is reshaped to a one-dimensional vector \( A \). Using the GA to update the population to the optimal solution, the population size is set to \( L \), and the population is expressed as \( P = \{ A_1, A_2, ..., A_L \} \). Population optimization and update process such as Figure 3.
With the iteration of the GA, the fitness value of the individual population will become larger, and the variance of the fitness value of the whole population will also become larger. This means that while the population is evolving in a positive direction, the differences in the population will also increase, thus ensuring diversity in the population.

3.3. Abnormal Evaluation

The output of Algorithm 1 is the LSH population $P^*$. A hash functions group $g_i$ can map the data matrix $X$ into a list of buckets $B_i = \{B_{i1}, B_{i2}, ..., B_{iM}\}$, $i = 1, 2, ..., L$

In the test phase, the test data $x_{test}$ is mapped to the L lists of buckets $B_{test}$ respectively by the L group hash function. For the $i$-th group hash function, find $d_{test i} = \min HD(B_{test}, B_{i})$, and then compare the L distances to obtain the minimum $d_{dis} = \min_{i=1,2, ..., L} d_{test i}$ and the corresponding bucket $B_{i}$.

The abnormal score of the test data $x_{test}$ is calculated as

$$A(x_{test}) = \frac{1}{1 + \exp(-\frac{\|x_{test} - c_i\|^2}{r_{ij}} - 1))}$$

where $r_{ij} = \max_{x \in B_{ij}} \|x - c_{ij}\|^2$ is the the radius of the bucket $B_{ij}$ and $c_{ij}$ is the center of the bucket $B_{ij}$.

When the distance from the data to the center of the bucket is less than the radius of the bucket, the
anomaly score is less than 0.5. On the contrary, the data is far away from the bucket and the abnormal score is high.

For a frame of test video, the maximum anomaly score of the video cuboid in the frame is calculated as the anomaly score of the frame. The threshold is set to determine the anomaly of the frame of the video according to the abnormal score. If the abnormal score is greater than the threshold, the frame of the video is judged as anomaly.

4. Experiments

4.1. Dataset

The UCSD dataset[23] is a public dataset that is often used to evaluate the performance of video anomaly detection algorithms, containing two subsets: pred1 and pred2. The image resolution of ped1 is 238 × 158 pixels and the image resolution of ped2 is 360 × 240 pixels. Ped1 has 34 training images and 16 test image sequences, which contains 3400 anomalous and 5500 normal frames. On the ped2, there has 16 training images and 12 test image sequences with 1652 anomalous frames and 346 normal frames. The main anomalies contain cycling on the sidewalk, skateboarding, and driving.

4.2. Evaluation Criter

This paper focuses on frame-level anomaly detection. One frame is an abnormal frame that is considered “positive”, otherwise it is considered “negative”. By taking different values for the threshold, the ROC curve can be obtained, the abscissa is FPR, and the ordinate is TPR.

\[
TPR = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad (7)
\]

\[
FPR = \frac{\text{false positive}}{\text{true negative} + \text{false positive}} \quad (8)
\]

Based on the ROC curve, the Area Under Curve (AUC) and Equal Error Rate (EER) are taken as the evaluation criteria. EER is, the value of FPR when FPR = 1 – TPR.

4.3. Performance

In order to prove the advantage of using GA to optimize the hash function group, Table 1 is the result of randomly generated hash function groups and optimized hash function groups. It can be seen that there is a certain degree of improvement on pred1 and pred2. This is mainly because the optimization of the GA improves the fitness value of the population as a whole, so that similar data points can be mapped to a bucket with higher probability, which improves the reliability of finding similar buckets.

In Table 2 we compare evaluation results of the proposed approach with some methods on the UCSD Ped1 and Ped2 dataset, containing AUC and EER. From Table 2 we can see that the proposed approach performs better than other methods in pred2. But it is inferior to other method in pred1. I think, it is because the spatio-temporal features extracted by other methods are more detailed for the description of the video, and the camera of pred1 has the different angle with the pred2. In pred1, the filming angle and the road are inclined, and the description of the optical flow is not accurate. These factors affect the performance of the algorithm.

Table 1. The optimal projection compared with the random projection on the UCSD dataset.

| Method    | UCSD ped1 | UCSD ped2 |
|-----------|-----------|-----------|
|           | AUC | EER | AUC | EER |
| Random    | 0.74 | 0.33 | 0.93 | 0.13 |
| Optimal   | **0.78** | **0.29** | **0.94** | **0.12** |

Table 2. Comparison of the proposed method with the other methods on the UCSD dataset.
5. Conclusion
In the field of video anomaly detection, the algorithms are mainly divided into three steps. Firstly, extracting video features from the video. Secondly, modeling the extracted features, and finally, the anomaly evaluation method is proposed to calculate the abnormal scores of the video frames. In the paper, we propose to use GA to optimize the LSH groups, so that similar data can be mapped to the same bucket with higher probability. The training data are mapped to a list of buckets. For the new test data, find the most similar training bucket, and calculate the anomaly score based on the radius of the bucket and the distance from the test data to the bucket center. Finally, for a frame of test video, the maximum anomaly score of the video cuboid in the frame is calculated as the anomaly score of the frame. The main contribution of the proposed method is that we can use a GA-optimized LSH group to model features which improves the accuracy of unsupervised clustering. Experimental results have justified the effectiveness of the proposed algorithm.

Acknowledgments
This work was supported by the Foundation (GYU—KYZ(2019~2020)DT—02).

References
[1] Andrade, E., Blunsden, S., Fisher, R.: Modelling crowd scenes for event detection. vol. 1, pp. 175 - 178 (01 2006)
[2] Angiulli, F., Pizzuti, C.: Fast outlier detection in high dimensional spaces. vol. 2431, pp. 15 – 26 (08 2002)
[3] Antic, B., Ommer, B.: Video parsing for abnormality detection. pp. 2415 – 2422 (11 2011)
[4] Bhaduri, K., L. Matthews, B., Giannella, C.: Algorithms for speeding up distancebased outlier detection. pp. 859 – 867 (08 2011)
[5] Biswas, S., Babu, R.V.: Anomaly detection via short local trajectories. Neurocomputing 242, 63 – 72 (2017)
[6] Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Comput. Surv. 41 (07 2009)
[7] D. Bay, S., Schwabacher, M.: Mining distance-based outliers in near linear time with randomization and a simple pruning rule. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (04 2003)
[8] Datar, M., Immorlica, N., Indyk, P., Mirrokni, V.S.: Locality-sensitive hashing scheme based on p-stable distributions pp. 253 – 262 (2004)
[9] E. Goldberg, D.: Genetic algorithm in search, optimization, and machine learning. Addison-Wesley, Reading, Massachusetts xiii (01 1989)
[10] E. Rumelhart, D., E. Hinton, G., J. Williams, R.: Learning representations by back propagating errors. Nature 323, 533 – 536 (10 1986). https://doi.org/10.1038/323533a0
[11] Davis, L.: Handbook of genetic algorithms. Van Nostrand Reinhold, New York 115 (01 1991)
[12] Farneback, G.: Two-frame motion estimation based on polynomial expansion 2749, 363 – 370 (2003)
[13] Gionis, A., Indyk, P., Motwani, R.: Similarity search in high dimensions via hashing. pp. 518 – 529 (1999)
[14] Hachiya, H., Matsugu, M.: Nsh: Normality sensitive hashing for anomaly detection. pp. 795 – 802 (12 2013)
[15] Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A., Davis, L.: Learning temporal regularity in video sequences. pp. 733 – 742 (06 2016)
[16] Jiang, F., Yuan, J., Tsafkaris, S., Katsaggelos, A.: Anomalous video event detection using spatiotemporal context. Computer Vision and Image Understanding 115, 323 – 333 (03 2011)
[17] Kim, J., Grauman, K.: Observe locally, infer globally: A space-time mrf for detecting abnormal activities with incremental updates. pp. 2921 – 2928 (06 2009)
[18] Kratz, L., Nishino, K.: Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. pp. 1446 – 1453 (06 2009)
[19] Liu, W., Luo, W., Lian, D., Gao, S.: Future frame prediction for anomaly detection – a new baseline. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
[20] Loy, C.C., Xiang, T., Gong, S.: Detecting and discriminating behavioural anomalies. Pattern Recognition 44, 117 – 132 (01 2011)
[21] Lu, W., Cheng, Y., Xiao, C., Chang, S., Huang, S., Liang, B., Huang, T.: Unsupervised sequential outlier detection with deep architectures. IEEE Transactions on Image Processing PP, 1 – 1 (06 2017)
[22] Luo, W., Wen, L., Gao, S.: A revisit of sparse coding based anomaly detection in stacked rnn framework. In: IEEE International Conference on Computer Vision (2017)
[23] Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes pp. 1975 – 1981 (2010)
[24] Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. pp. 935 – 942 (06 2009)
[25] Sabokrou, M., Fathy, M., Mojtaba, H.: Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder. Electronics Letters 52 (04 2016)
[26] Sabokrou, M., Fathy, M., Mojtaba, H., Klette, R.: Real-time anomaly detection and localization in crowded scenes (06 2015)
[27] Xu, D., Ricci, E., Yan, Y., Song, J., Sebe, N.: Learning deep representations of appearance and motion for anomalous event detection. CoRR abs/1510.01553(2015)
[28] Zhang, J., Wang, H.: Detecting outlying subspaces for high-dimensional data: The new task, algorithms, and performance. Knowl. Inf. Syst. 10, 333 – 355 (10 2006)
[29] Zhang, Y., Lu, H., Zhang, L., Ruan, X., Sakai, S.: Video anomaly detection based on locality sensitive hashing filters. Pattern Recognition 59, 302 – 311 (2016)
[30] Zhao, Y., Deng, B., Shen, C., Liu, Y., Lu, H., Hua, X.S.: Spatio-temporal autoencoder for video anomaly detection. In: Proceedings of the 25th ACM International Conference on Multimedia. pp. 1933 – 1941. MM’17, ACM, New York, NY, USA(2017)