Detection of Abnormal Objects in a Scene based on Local Features

Junya Kobayashi
Department of Information Engineering
Meijo University, Nagoya 480-1192 Japan
m0830009@ccmailg.meijo-u.ac.jp

Keiichi Yamada
Department of Information Engineering
Meijo University, Nagoya 480-1192 Japan
yamadak@ccmfs.meijo-u.ac.jp

Abstract

A method for detecting an abnormal object in video images is proposed. We define abnormal object as an object that does not appear in usual scenes. The detection system is trained with video images of usual scenes, and then detects abnormal objects in new input video images. The proposed method detects an abnormal object by assuming an object as a set of local features and ignoring the positional relationship between local features. SIFT algorithm is used for detecting and describing local features. The proposed method does not need segmentation of the region of an object and expected to be robust to occlusion and cluttered background.

1 Introduction

For realizing safe society, expectation on video monitoring technology using image recognition is high and such technology is considered to be useful in the field such as for accident prevention, crime prevention and healthcare. The purpose of our research is to develop a technology for detecting abnormal objects in a scene from images captured with a video camera. An application example of abnormal object detection is to detect the object in a road environment to which a driver should be careful or pay attention such as a pedestrian, a bicycle and unknown obstacles.

Detection of abnormal object is not a typical classification problem into pre-defined classes, because it is impossible to itemize all the abnormal objects beforehand explicitly and then training samples of the class of abnormal objects cannot be prepared.

Several methods for detecting abnormal objects have been proposed. There are mainly two kinds of approaches. One approach is using saliency for detecting abnormal objects in a scene [1, 2]. The methods using saliency are a bottom-up approach for detecting abnormal objects without using prior knowledge on the objects. A problem of these methods based on saliency is that, in general, a salient object is not necessary an abnormal object and a non-salient objet is not necessary a usual object.

Another approach for detecting abnormal objects is a top-down approach using knowledge on usual objects. In this approach, a detection system is trained with usual scenes beforehand and then detects the object, from an input image, that has not appeared in the training data as an abnormal object. Boiman et al. [3] proposed a method for detecting irregularities in images. They tried to compose a new observed image region using chunks of data extracted from training data. Regions in the observed image which cannot be composed from the training data were regarded as unlikely and suspicious. An issue of this method is partial occlusion of the object. Sato et al. [4] proposed a method for detecting suspicious object based on appearance frequency of a segmented object region in an observed image. An issue of this method is an error on the segmentation stage results in an error of the detection.

On the other hand, for general object recognition, an approach based on local feature vector [5,6,7] has been proposed. This approach does not based on segmentation of a region of an object. Local features have proved to be effective for matching and recognition tasks, as they are robust to occlusion and cluttered background. In [5], objects were recognized by a probabilistic approach in which objects were modeled as flexible constellations of parts. In [6], a visual categorization method was proposed based on a bag of keypoints approach based on local features. The main advantages of this method are its simplicity, its computational efficiency and its invariance to affine transformations, as well as occlusion, lighting and intra-class variations.

This paper proposed a method for detecting an abnormal object in video images by a top-down approach based on local features. The proposed method detects an abnormal object by assuming an object as a set of local features and ignoring the positional relationship between local features on an image. The method does not need segmentation of the region of an object and expected to be robust to partial occlusion and cluttered background. In addition, the computational cost of the method is lower because the method does not refer the positional relationship or co-occurrence information of the local features. This paper presents the proposing method and then shows the performance of the method by an experiment using a dataset of road scene video images.

2 Method

2.1 Approach

We define that the object to be detected, *i.e.* abnormal object, is an object that does not appear in usual scenes. In our approach, the detection system is trained with video images of usual scenes beforehand and then detects an abnormal object that did not appeared in the training data from new input video images. The proposed method is based on the part-based approach based on the local feature as described in section 1. While various methods have been proposed as a method for detecting local feature points and describing the features, we employ SIFT (Scale-invariant feature transform) algorithm [7]. The reason why we use SIFT is that SIFT feature is invariant to the change in scale and rotation, and that is

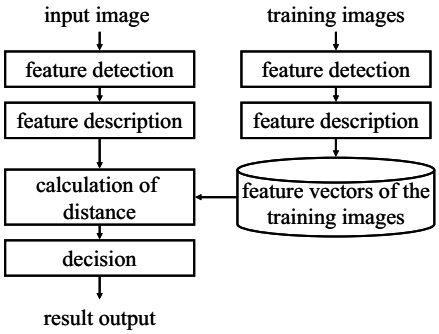


Figure 1: Outline of the proposed method.

strong for an illumination change, and its high performance is known from former researches. Note that, one may be able to use other method than SIFT for detecting local feature points and describing the features. The proposed method also uses information on absolute position of feature points on an image, while the method avoids the use of information on relative position between feature points on an image since the computational cost is considered to be very high.

Figure 1 shows an outline of the proposed method. The processing flow of the method is as follows. Feature points (keypoints) in the training images are detected by a SIFT algorithm and then the feature vectors are calculated for each feature point. In the same way, feature points in a new input image are detected and the feature vectors are calculated. Then, for each of the feature points in the input image, the similarity between the feature point in the input image and the feature points in the training images are calculated. After that, for each small region in the input image, whether the region includes abnormal object or not is determined based on the similarities of the feature points in the small region. For the calculation of the similarity of the feature points between the input image and that in the training images, ANN (Approximate Nearest Neighbor) search algorithm [8] is used for searching a nearest feature vector to lower the calculation cost. The detail of the proposed method is described in 2.2 and 2.3.

2.2 Feature vector

The proposed method uses a 133-dimensional feature vector **V** for representing the feature of a feature point. As shown in the following equation, the feature vector **V** for a feature point is composed of the 128-dimensional SIFT descriptor (v_1,v_2, ... ,v_{128}), the orientation θ, the scale size s, and the absolute location (x, y) on the image.

$$\boldsymbol{V} = (v_1, v_2, ..., v_{128},\ c_\theta \cos\theta,\ c_\theta \sin\theta, c_s s,\ c_x x,\ c_y y)$$

Where, c_θ, c_s, c_x and c_y are weight coefficients for the orientation θ, the scale size s and the location (x, y). These coefficients are the parameters that have to be determined experimentally depending on its application using a cross validation for instance. The reason why the scale information is included in the feature vector is to exclude the feature point whose size is extremely different, and why the location is included is to use the information of the location of the object that usually appears.

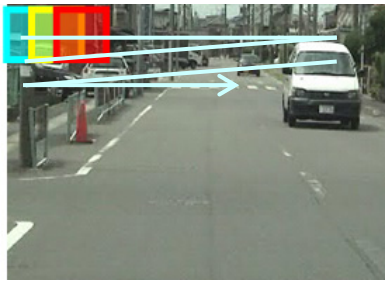


Figure 2: By scanning the input image with a small window, whether the region of the window contains a part of an abnormal object or not is judged.

2.3 Detection

By scanning an input image with a small window, as shown in Figure 2, whether the region of the window contains an abnormal object or not is judged as follows. If many feature vectors of the feature points in the region are differ from any feature vectors of the feature points in the training images, it is judged that the region contains an abnormal object. While there are several options to evaluate the difference between a feature vector of a feature point in the input image and the feature vectors in the training images, we used the distance between the feature vector of the feature point in the input image and the nearest feature vector in the training images. One may be able to use mean distance between the feature vector in the input image and the nearest k feature vectors in the training images to increase the robustness. As described before, ANN search algorithm is used to search the nearest feature vector in the training images for a feature vector of a feature point in the region in the input image in order to lower the computational cost.

The proposed method calculates the degree q_j the j-th window region a_j contains an abnormal object by the following equation.

$$q_j = \frac{n}{m_j} \sum_{i=1}^{m_j} d_i \qquad (1)$$

where, d_i is Euclid distance or Mahalanobis distance between the feature vector of the i-th feature point in the region a_j and its nearest feature vector in the training images. Also, m_j is the number of the feature points in the region a_j and n is the number of the pixels of the region. When the degree q_j the region a_j contains an abnormal object is more than a specific value, it is judged that the region is a part of an abnormal object.

The meaning of calculating the degree the region includes an abnormal object by the equation (1) can be explained as follows. Now, we assume that the probability p_{usual} that a feature point is a part of a usual object is proportional to exponential of the distance d_i as the following equation.

$$p_{\text{usual}}(d_i) = \exp(-c \cdot d_i) \qquad (2)$$

where, c is a coefficient. Taking logarithms of both side of the equation (2),

$$\ln p_{\text{usual}}(d_i) = \ln \exp(-c \cdot d_i) = -c \cdot d_i \qquad (3)$$

The probability $P_{\text{usual},j}$ that a region a_j is a part of a usual object is thought to be the product of the probabilities that each pixel is a part of a usual object. When the

probability p_{usual} for a pixel that does not have a feature point is assumed to be proportional allotment of the probabilities of the feature points in the region, the logarithmic probability $\ln P_{\text{usual},j}$ that the region a_j is a part of a usual object becomes

$$\ln P_{\text{usual},j} = \ln \prod_{i=1}^{m_j} p_{\text{usual}}(d_i)^{\frac{n}{m_j}} = \frac{n}{m_j} \sum_{i=1}^{m_j} \ln p_{\text{usual}}(d_i) \quad (4)$$

Substitution of equation (3) into equation (4) yields

$$\ln P_{\text{usual},j} = -c \frac{n}{m_j} \sum_{i=1}^{mj} d_i \quad (5)$$

The degree q_j the j-th region a_j contains an abnormal object can be represented as adding a minus sign to the logarithmic probability that the region a_j is a part of a usual object. Then, from (5), the degree q_j the j-th region a_j contains an abnormal object is represented as

$$q_j = -\ln P_{\text{usual},j} = c \frac{n}{m_j} \sum_{i=1}^{m_j} d_i \quad (6)$$

The proportion coefficient c in (5) can be assumed to be 1 without loss of generality. Then the equation (1) is obtained.

3 Experiment

3.1 Method

Performance of the proposed method was evaluated by the following experiment. The dataset used for the evaluation was video images that were taken with a video camera installed in a car in a forward-looking manner. The dataset was composed of three sequences of video images each of which was taken on a road whose length was 0.8 km. One sequence of that included pedestrians and bicycles and other two sequences included no pedestrian nor bicycle. The latter two sequences were used for training and the former one sequence was used for test. The pedestrians and the bicycles, that did not appeared in the training data, were assumed to be abnormal objects in this experiment. The proposed method was trained and tested using these dataset and the detection performance of the abnormal objects was evaluated.

(a) examples of the training data

(b) examples of the test data

Figure 3: Examples of the training data (a) and the test data (b).

The resolution of the video images was 320 × 240 pixels, and the frame rate was 10 frames per second. Then, the number of the images of the training data was 1140 frames and that of the test data was 228 frames. Also, the number of the frames, in the test data, that included a pedestrian or a bicycle was 42. Examples of the training data and the test data are shown in Figure 3(a) and (b), respectively. Note that, in (b), left side image includes an abnormal object, *i.e.* a bicycle, and right side image does not include any abnormal objects.

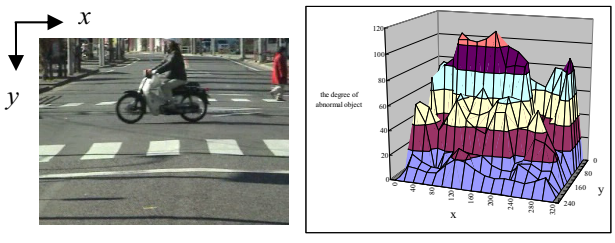
In the experiment, the size of the window a_j was 40 × 40 pixels and the window was scanned by a step of 20 pixels horizontally and vertically on an input image. The parameter ε in ANN search algorithm was 5. Note that, those parameters were determined empirically. A pedestrian or a bicycle whose size is smaller than 20 × 20 pixels were exclude from the evaluation, since it was apparently difficult to detect in such lower resolutions.

3.2 Results

Figure 4 shows an example of the test results. In this figure, (a) is a frame of the test images and (b) is the obtained degree of abnormal object for the frame. In the figure (b), the bottom surface of the 3-D graph corresponds to the two-dimensional plane of the test image, the vertical axis shows the degree q_j that the pixel of the image is a part of an abnormal object. From this figure, it can be seen that the probability at the bicycle and the pedestrian is higher. Figure 5 shows another example of the test result in which a pedestrian was in a cluttered background. In the figure, (a) is a test image and (b) is the detection result with a decision threshold. In (b), the red region shows the detected abnormal object. From this figure, the robustness of the method to a cluttered background can be seen.

Figure 6 shows ROC (Receiver Operating Characteristic) curves of the proposed method when the decision threshold of the judgment was varied for several cases of the weight coefficients. In this figure, the detection rate is defined as the ratio of the number of frames in which an abnormal object is correctly detected to the number of frames that include abnormal objects. The false detection rate is defined as the ratio of the number of the frames in which a region is judged as abnormal while the region is not abnormal object to the number of the frames in which no abnormal object is included. Each curve in this figure shows the result for the case in which the weight coefficients in the feature vector **V** were (c_θ, c_s, c_x, c_y).

As shown in this figure, the proposed method detected 86% of the abnormal objects (*i.e.*, false negative rate was 14%) when the false detection (false positive) rate was 28%.



(a) test image (b) obtained result

Figure 4: An example of the test results: a test image (a) and the obtained distribution of the degree of abnormal object (b).

(a) test image (b) detection result

Figure 5: A test result in which a pedestrian was in a cluttered background. In (b), the red region shows a detected abnormal object.

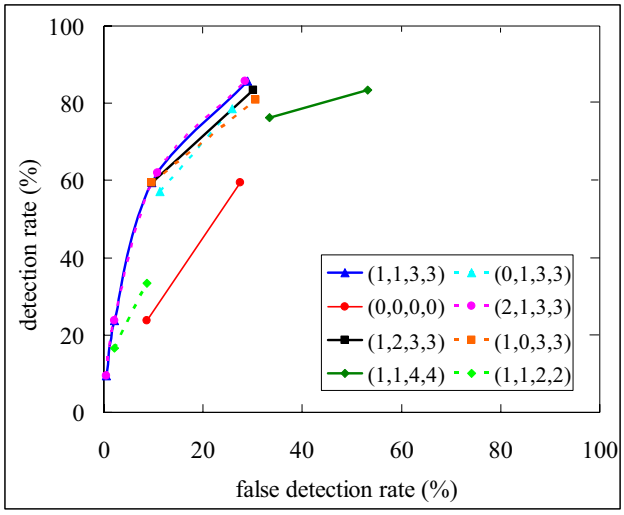


Figure 6: ROC curves of the proposed method. Each curve shows the evaluation result for the case in which the weight coefficients in the feature vector **V** were (c_θ, c_s, c_x, c_y).

(a) detection miss

(b) false detection

Figure 7: Typical cases of detection miss (a) and false detection (b). For each case, left side image is test image and right side image shows the detection result.

Figure 7 shows typical cases of detection miss and false detection. Detection misses were occurred mainly due to the blur of the image caused by a rapid motion of the object on the image plane. Detection misses were also occurred when the contrast of the object on the image was low. A solution of the low-contrast problem is thought to be using color SIFT descriptors [9] as the feature point detection and description. A cause of false detections was that some objects that did not appear in the training data, except for pedestrians and bicycles, appeared in the test data. This problem is thought to be able to be solved by increasing the amount of training data.

4 Summary

A method for detecting an abnormal object in video images based on local features was proposed. The method detects an abnormal object by assuming an object as a set of local features and ignoring the positional relationship between the local features. The method does not need segmentation of an object and expected to be robust to partial occlusion and cluttered background. The performance of the method was evaluated by an experiment and the effectiveness of the method was shown.

References

[1] T. Asami, et al.: "A Saliency Based Abnormality Detection Algorithm for Visual Inspection," *Meeting on Image Recognition and Understanding (MIRU)*, pp. 1400-1407, 2008. (in Japanese).

[2] Z. Wang, B. Li: "A Two-stage Approach to Saliency Detection in Images," *IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 965–968, 2008.

[3] O. Boiman, M. Irani: "Detecting Irregularities in Images and in Video", *Proc.ICCV'05*, Vol.1, pp.462-469, 2005.

[4] K. Sato, N. Nitta, Y, Ito, N. Babaguchi: "Suspicious Object Detection Based on Appearance Frequency for Video Surveillance," *Meeting on Image Recognition and Understanding (MIRU)*, pp. 404-409, 2008. (in Japanese)

[5] R. Fergus, P. Perona, and A. Zisserman: "Object Class Recognition by Unsupervised Scale-invariant Learning," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 264-271, 2003.

[6] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, C. Bray: "Visual Categorization with Bags of Keypoints," *Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1-22, 2004.

[7] D. G. Lowe: "Distinctive Image Features from Scale-invariant Keypoints", *Journal of Computer Vision*, vol. 60, no. 2, pp.91-110, 2004.

[8] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu: "An Optimal Algorithm for Approximate Nearest Neighbor Searching in Fixed Dimensions," *Journal of the ACM*, vol. 45, no. 6, pp. 891-923, 1998.

[9] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek: "Evaluation of Color Descriptors for Object and Scene Recognition," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2008.