Abstract

We focuse on a privacy-preserving video system and try to obtain the internal information of video with image processing techniques. Using optical flow features, we try to estimate and segment inhomogeneous crowds composed of pedestrians that travel in different directions. With these features, we apply several classfying algorithms to tell the normal behaviors from the abnormal. Then we compare the accuracy of these algorithms and aim to make some improvement on the algorithms. We will try to validate both the crowd segmentation algorithm, and the crowd counting system, on a large pedestrian dataset (2000 frames of video, containing 49,885 total pedestrian instances). This project will help monitering abnormal behaved objects in public environment with vision information.

Key words: image processing, optical flow, classification algorithms

Background

There is currently a great interest in vision technology for monitoring all types of environments. And one of the most important goals of monitoring environments is to find abnormal behaviors, which requires the identification and understanding of human behavior. The existing behavior recognition and understanding technology can be divided into visual-based methods and non-visual methods based on ways of information collection. With the development of computer vision technology, vision-based human behavior recognition and identification has attracted increasing attention. Its application to intelligent control, medical diagnosis, perception interface, human-computer interaction, motion analysis has inspired wide interest among the majority of researchers.

Yet, the deployment of vision technology is invariably met with skepticism by society at large, given the perception that it could be used to infringe on the individual’s privacy rights. So Antoni B.Chan proposed a new formulation for surveillance technology by detecting people in the scene and tracking over time, which is averse to individual tracking. But he didn’t put forward any practical methods for telling the abnormal behaviors from the normal ones. As machine learning methods are widely used, this project try to apply this technique for solving the detection goal.

In addition, images could be added an amount of noise in the process of formation, transmission, reception. The noise hinders human visual perception, prevents the system sensors from correctly understanding or analyzing the image information, resulting in deviations between the actual value and the detected value. Hence, it’s necessary to use appropriate methods to reduce noise as pre-processing before further application of image fusion, image reconstruction, image enhancement, image segmentation, feature extraction and analysis.

Literature review

The abnormal behavior detection consists of three paradigms: 1)crowd detection, 2)visual feature extraction and clustering, 3) feature-based regression. Crowd detection algorithms are based on boosting appearance and motion features[1]. Because there exists significant occupation among crowd, the detection algorithms may performs badly, which can be partly solved by part-based detectors [2,3]. The second paradigm plays a fundamental role by identifying and tracking visual features over time. Based on the features clustering, normal behaviors are gathered. Examples of this formulation include [4], which uses the KLT tracker and agglomerative clustering, and [5], which takes an unsupervised Bayesian approach. These methods typically work by: 1) subtracting the background; 2) measuring various features of the foreground pixels, such as total area [6,7,8], edge count [6,7,8], or texture [9].

With regard to clustering algorithms, they attempt to classify elements into categories based on their similarity. Several different clustering strategies have been proposed [10]. In K-means, clusters are generated by data characterized by a small distance to the cluster center. An objective function, typically the sum of the distance to a set of putative cluster centers, is optimized [11] until the best cluster centers candidates are found. In distribution-based algorithms, one attempts to reproduce the observed realization of data points as a mix of predefined probability distribution functions; the accuracy of such methods depends on the capability of the trial probability to represent the data

Objectives

This project aims to detect and mark the abnormal behaviors of people in crowded public environment, such as runners (fast motion) among walking pedestrians (slow motion) in a park. At the end of project, it is supposed to generate a robust algorithm that with a video clip given, it could spot the abnormal behaved persons and track these persons with highlighted mark. The project should employ multiple classifiers and find the optimal one with certain evaluation metrics.

Project Overview:

The project contains 4 phases. In the data collection phase, a video sample from database of UCSD is obtained as the training and testing data. It is a 2000 frames (200 seconds) video collected from a stationary digital camcorder overlooking a park with pedestrian walkway at UCSD. The original video was captured at 30fps with a frame size of 740×480, and was later downsampled to 238 × 158 and 10 fps [\*]. In pre-processing and feature extraction phase, it applies Canny edge detection, Hough transformation, morphological filter and optical flow techniques to detect and segment the moving objects from each frame, and then extract and normalize their optical flow and RGB features. Next, in the training phase, three different models, fisher, clustering, and CNN are trained based on the training data from the features extracted in previous phase to obtain the optimal parameters. Last, in the testing phase, use the rest of video frames as test data and compute the accuracy for each model. The best model that should be used to develop the program is determined from the largest area under the curve of ROC.

The program is coded in Python with 3 dependencies

* OpenCV for image pre-processing and feature extraction
* Keras (with Tensor-flow backend) for CNN modelling, training and testing
* Sciki-Learn for
  + Linear and Clustering modelling, training and testing.
  + ROC metric plot and AUC calculation.

Methodologies

1. Data collection
   1. Download Pedestrian Traffic Database from Statistic Visual Computing Lab of UCSD.
   2. Clip the first 2000 frames (200 seconds)
   3. Image down sampled to 238\*158 and 10 fps.
2. Image Preprocessing and feature extraction

2.1 Foreground segmentation

2.1.1 Apply optical flow method to find the moving pixels

2.1.2 Apply morphological filter on the optical flow images

2.1.3 Segment each moving object on the optical flow images

2.2 Computing weight matrix for normalization

2.2.1 Convert RGB images to grey-scale images.

2.2.2 Apply Gaussian blur on the images to smoothen and reduce the noise

2.2.3 Use Canny edge detector to generate the edge-only-images

2.2.4 Use Hough transformation to find lines in edge-only-images

2.2.5 Compute the weight matrix for each row pixels referred to the length of lateral line generated by road shoulder lines which is found in 2.2.4; Computer the weight matrix for each column pixels referred to the height of persons on the line

2.3 Feature extraction and normalization

2.3.1 Obtain the optical flow features for each object from 2.1.3

2.3.2 Normalize the optical flow values according to matrix from 2.2.5

2.3.3 Using the mean of optical flow values to obtain 2-dimension features of each object

1. Model training and testing
   1. Fisher classifier

We consider objects from first 50 pictures as training samples (including features, labels). The goal of fisher classifier is to maximize distance between projected class means and minimize projected class variance. We consider objects from pricture120 to picture180 as testing samples. Compute the testing error and plot roc curve. Mark the abnormal behaviors in testing pictures.

* 1. Clustering by fast search and find of density peaks

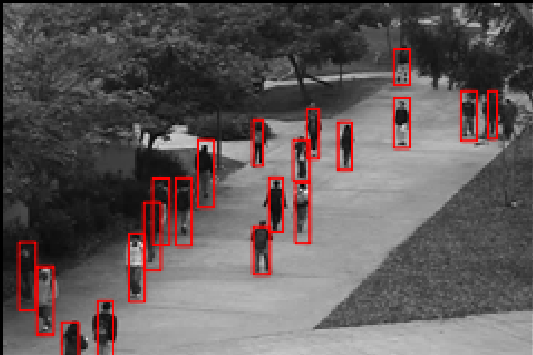
We consider optical flow features of objects from first 100 pictures as training data and apply clustering method proposed by [11]. Then we consider objects from picture 120 to picture 180 as testing samples. Compute the testing error and mark the abnormal behaviors in testing pictures.

1. Model analysis and comparison

use the rest of video frames as test data and compute the accuracy for each model. The best model that should be used to develop the program is determined from the largest area under the curve of ROC.

Expected results:

After preprocessing, we should be able to track each person with highlighted red square box and obtain the optical flow features:



(Draw manually)

Train the model with data features and use the model to label the abnormally behaved person:



(person on bike, fast motion)

Reference

[1] P. Viola, M. Jones, and D. Snow, “Detecting pedestrians using patterns of motion and appearance,” IJCV, vol. 63(2), pp. 153–61, 2005.

[2] B. Wu and R. Nevatia, “Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors,” in *ICCV*, vol. 1, 2005, pp. 90–7.

[3] S.-F. Lin, J.-Y. Chen, and H.-X. Chao, “Estimation of number of people in crowded scenes using perspective transformation,” *IEEE Trans. System, Man, and Cybernetics*, vol. 31, no. 6, 2001.

[4] V. Rabaud and S. J. Belongie, “Counting crowded moving objects,” in *CVPR*, 2006.

[5] G. J. Brostow and R. Cipolla, “Unsupervised bayesian detection of independent motion in crowds,” in *CVPR*, vol. 1, 2006, pp. 594–601.

[6] N. Paragios and V. Ramesh, “A mrf-based approach for real-time subway monitoring,” in *CVPR*, vol. 1, 2001, pp. 1034–40.

[7] S.-Y. Cho, T. W. S. Chow, and C.-T. Leung, “A neural-based crowd estimation by hybrid global learning algorithm,” *IEEE Trans. Syst, Man, Cybern.*, vol. 29, pp. 535–41, 1999.

[8] A. C. Davies, J. H. Yin, and S. A. Velastin, “Crowd monitoring using image processing,” *Electron. Comm. Eng. J.*, vol. 7, pp. 37–47, 1995.

[9] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, “Dynamic textures,” *Intl. J. Computer Vision*, vol. 51, no. 2, pp. 91–109, 2003.

[10]1. R. Xu, D. Wunsch 2nd, IEEE Trans. Neural Netw. 16, 645–678 ,2005.

[11] Rodriguez A, Laio A. Clustering by fast search and find of density peaks[J]. Science, 2014, 344(6191):1492.