Volume : 2 | Issue : 4 | Apr 2013 • ISSN No 2277 - 8179

Research Paper

Content Based Indexing and Retrieval from Vehicle Surveillance Videos Using Optical Flow Method

Engineering

KEYWORDS : Optical Flow Model, k-means, SOM, Vehicle Surveillance, Vehicle Classificationwww

Shruti V Kamath	Student (B.E), Department of Computer Science and Engineering, R.V College of Engineering
Mayank Darbari	Student (B.E), Department of Computer Science and Engineering, R.V College of Engineering
Dr. Rajashree Shettar	Professor, Department of Computer Science and Engineering,R.V College of Engineering

ABSTRACT

Visual vehicle surveillance is widely used and produces a huge amount of video data which are stored for future or immediate use. To find an interesting vehicle from these videos because of car crashes, illegal U-turns, speeding or anything that the user may be interested in, is a common case. This would be a very labor intensive process, if there is not an effective method for the indexing and retrieval of vehicles.

In the proposed approach, a surveillance video from a highway is taken. Then shots are detected using color histogram method. Key frames are extracted to avoid redundancy in a shot. After key frames are extracted from the input video, vehicles are detected and tracked using Optical Flow. The proposed method extracts vehicles and classifies them based on area and colour. Unsupervised clustering is performed on the vehicles extracted using k-means and SOM (Self Organizing Map) algorithms. Clustering is done based on size (small, medium, large) and color (red, white, grey) of the vehicles. Using the clustering results, a matching matrix is computed for these two clustering algorithms, with respect to the object tracking method. The proposed system gives an accuracy of up to 99% (with respect to size) and 90% (with respect to color).

I. INTRODUCTION

The processing of large data in surveillance videos to be efficient and accurate, a context based indexing and retrieval system is used for this. This paper proposes a system which demonstrates its strong ability in digital surveillance and commonly has applications in police department, traffic control, intelligence bureau and defence. A surveillance video from a highway is taken. Then shots are detected using colour histogram method [1][4], which takes the histogram difference of the frames and computes a histogram. A threshold value is then set. The frames where the value is above the threshold are identified as shots.

To represent a shot succinctly, the redundancies are removed from the video using key frame extraction. After key frames are extracted from the input video, we detect and track objects (moving vehicles) using the Optical Flow Model [2][3]. We have assumed a stationary camera which is fixed at an angle for the input videos. Vehicles identified are classified based on area and colour.

Unsupervised clustering is performed on the objects detected using k-means and SOM [6][7] algorithms. Clustering is done for vehicles under 3 classes namely small, medium and large with respect to area and 3 classes namely, grey, white and red with respect to colour. Then given a threshold value, the extracted vehicles are clustered and the matching matrix is calculated using the results of k-means and SOM, with respect to the Optical Flow method.

The remainder of this paper is organized as follows. Section II is a Literature Survey, citing about the existing algorithms. A brief System Overview is present in Section III. How the feature is extracted and indexed is explained in Section IV. Experiments and Results are discussed in Section V. Finally the conclusion and future work is presented in Section VI.

II. LITERATURE SURVEY

The research on shot boundary detection has a long history, and there exist specific surveys on video shot boundary detection. A shot is a consecutive sequence of frames captured by a camera action that takes place between start and stop operations, which mark the shot boundaries. Generally, shot boundaries are classified as cut in which the transition between successive shots is abrupt and gradual transitions which include dissolve, fade in, fade out, wipe, etc., stretching over a number of frames. Current similarity metrics for extracted feature vectors include the 1-norm cosine dissimilarity, the Euclidean distance, the histogram intersection, and the chi-squared similarity [1][4] is used detect shots.

The key frames of a video reflect the characteristics of the video to some extent. The static key frame features useful for video indexing and retrieval are mainly classified as color-based, texture-based, and shape-based. Color-based features include color histograms, color moments, color correlograms, a mixture of Gaussian models [5], etc. Shape-based features that describe object shapes in the image can be extracted from object contours or regions. The color histogram feature has been used by many researchers for image retrieval. A color histogram is a vector, where each element represents the number of pixels falling in a bin, of an image [10].

The detection of moving objects in video sequences is an important and challenging task in multimedia technologies. To segment moving foreground objects from the background a pure background image has to be estimated. This reference background image is then subtracted from each frame and binary masks with the moving foreground objects are obtained by thresholding the resulting difference images [1].

Machine learning-based approach uses labeled samples with low-level features to train a classifier or a set of classifiers for videos.Vehicle detection and vehicle classification using neural network (NN), can be achieved by video monitoring systems [7].

III. SYSTEM OVERVIEW

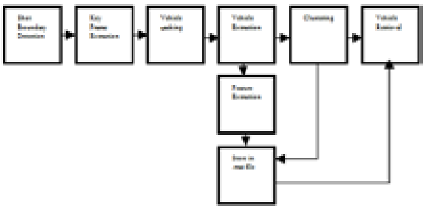

Fig. 1 shows the architecture of the proposed framework.

As shown in Fig. 1, the raw surveillance video is captured and the vehicles are extracted. During object tracking, on one hand the information of vehicles is extracted and saved as .mat file.

Volume : 2 | Issue : 4 | Apr 2013 • ISSN No 2277 - 8179

The color feature is picked up from the vehicles and stored in .mat file. When the user selects a color or/and a type provided by the system, the retrieval results will be returned.

VI. VIDEO PROCESSING & OBJECT TRACKING AND INDEXING

A. SHOT BOUNDARY DETECTION

A shot is a consecutive sequence of frames captured by a camera action that takes place between start and stop operations, which mark the shot boundaries. There are strong content correlations between frames in a shot. In the surveillance videos we used, abrupt transitions were present, where transitions between successive shots are abrupt.

In this paper, threshold-based approach is implemented [1]. The threshold-based approach detects shot boundaries by comparing the measured pair-wise similarities between frames with a predefined threshold: When a similarity is less than the threshold, a boundary is detected. In the input videos we collected consisted of abrupt transitions. One of the effective ways is intensity histogram. According to NTSC standard, the intensity for a RGB frame can be calculated as,

$$I = 0.299R + 0.587G + 0.114B \quad (1)$$

where R, G and B are Red, Green and Blue channel of the pixel. The intensity histogram difference [9], is expressed as,

$$SDi = \sum_{j=1}^{G} |Hi(j) - hi + 1(j)| \quad (2)$$

where Hi(j) is the histogram value for ith frame at level j. G denotes the total number of levels for the histogram.

In a continuous video frame sequence, the histogram difference is small, whereas for abrupt transition detection, the intensity histogram difference spikes. The threshold value to determine whether the intensity histogram difference indicates an abrupt transition can be set to,

$$Tb = \mu + \alpha\sigma \quad (3)$$

where and are the mean value and standard deviation of the intensity histogram difference. The value of typically varies from 3 to 6.

Sometimes gradual transitions can have spikes that even higher than those in abrupt transitions. In order to differentiate the abrupt transitions from gradual transitions, the neighboring frames of a detected spike are also tested, if there're multiple spikes nearby, the transition is more likely to be gradual transition, and we simply drop this detection.

B. KEY FRAME EXTRACTION

There are great redundancies among the frames in the same shot; therefore, certain frames that best reflect the shot contents are selected as key frames to succinctly represent the shot. Sequential comparison between frames is used in our approach [1]. A threshold value is appropriately set such that only salient information of the video is intact.

The Algorithm:

1. Choose the first frame as the standard frame that is used to compare with the following frames.
2. Get the corresponding pixel value in both frames one by one, and computing their difference respectively.
3. After finishing 2, add the results in 2 altogether. The sum will be the difference between these two frames.
4. Finally, if sum is larger than a threshold we set, select frame (1+i) as a key-frame, then frame (1+i) becomes the standard frame. Redo 1 to 4 until there is no frame to be captured.

C. OBJECT TRACKING USING OPTICAL FLOW

Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image. Optical flow can arise from relative motion of objects and the view. Discontinuities in the optical flow can help in the segmentation of images into regions that correspond to different objects. Horn-Schunck Method is used here [2]. By assuming that the optical flow is smooth over the entire image, the Horn-Schunck method computes an estimate of the velocity field.

Blob analysis, or Blob detection is to detect and analysis connected region in a frame and the vehicles are tracked. The optical flow vectors are stored as complex numbers and their magnitude squared is computed, which is used for thresholding the frames. Median filtering is applied to remove speckle noise. Morphological operations are done to remove small objects and holes. The bounding box is drawn for blobs with extent ratio above 0.4 and those of suitable size. The motion vectors are drawn for these detected vehicles as well.

D. FEATURE EXTRACTION

The vehicles are extracted from the video at a point where they are prominent. In the proposed method, a range value is specified within which vehicles are clearly visible. The row value of centroid of the bounding box (i.e. position) is considered for the comparison with the above range. Vehicles within the bounding box are extracted and stored in a .mat file along with the area and frame number.

The Algorithm:

1. A range is set within which the bounding box of the vehicles are extracted, when they come within range. The range is set such that view of the camera is uniform.
2. Vehicles do not move uniformly, the same vehicle may be detected more than once.
3. An array is set which holds the vehicles of the previous 8-10 vehicles extracted. We have observed through experimental analysis that a max of the previous 10 vehicles, the object may be extracted more than once for our dataset.
4. An object detected is thus compared to previously extracted vehicles by using Euclidean distance norm formula.
5. The area (number of pixels within the blob) and the image of cropped portion of car are stored in a .mat file.

Here, the colors red, grey and white are considered as the color category. In order to find out which of these is a dominant color in the image, the cropped image of the vehicle is divided into 8x8 blocks and the centroid pixel's HSL color value is extracted. This is similar to extraction of color in [8]. Low values of luminance are closer to black and higher values are closer to white. An approximate hue and saturation value range is associated for the color required. Starting with lightness value of the centroid pixel of the blocks, a histogram is computed for the values using the approximate ranges. The values which satisfy a particular range of luminance for a color are considered for comparing with saturation range, and those satisfying the saturation range of the same color, are considered for comparison with hue range of that particular color. The count of a color if significant in an image is considered as the block's dominant color. Thus using the block's centroid color value, we obtain dominant color of an image.

E. CLUSTERING

Clustering is performed on the samples gathered based on size of vehicle and color. k-means and SOM algorithms are used.

SOMs are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space. This makes SOMs useful for visualizing low-dimensional views of high-dimensional data, akin to multidimensional scaling. 200 epoches are used.

k-means clustering is a method of cluster analysis which aims to partition 'n' observations into 'k' clusters in which each observation belongs to the cluster with the nearest mean using Euclidean distance formula.

A threshold is set and the actual classification into small, medium and large is determined. Similarly, this approach is done for colour as well. The matching matrix is then calculated. The results of the clustering are stored in .mat files, which are used for retrieval.

F. VEHICLE RETRIVAL

The user is provided with a graphical interface. The user selection is matched with values stored in the mat files. The output is displayed to the user.

V. EXPERIMENTAL ANALYSIS AND RESULTS

MATLAB is used for our implementation. The test dataset consists of a video having 7142 frames. After shots boundaries have been detected and key frames have been extracted from the video, the number of frames is brought down to 5793. Using Optical Flow the vehicles detected are extracted and stored. Up to 100 vehicles were extracted. The results are shown in Table 1, 2 and 3.

A threshold is set and the actual classification into small, medium and large is determined. The approach used for calculating the actual class of the dataset is as follows; the median value of the area is taken as M. Small sized vehicles fall below the range of 1.4*M (=a), medium sized between 1.4*M and 3.5*M (=b) and the rest is large sized vehicles. The difference between 'a' and 'b' value must at least be 2.5 for sample size (number of vehicles) of 50 and above. For the 100 sample, actual classes were manually assigned.

PERFORMANCE MEASURES

ACCURACY

Accuracy is the overall correctness of the model and is calculated as the sum of correct classifications divided by the total number of classifications.

Accuracy = P/C (4)

Where P is the total no. of predicted classifications, and C is the total no. of actual classifications

PRECISION

Precision is a measure of the accuracy provided that a specific class has been predicted. It is defined by:

Precision = tp/(tp + fp) (5)

where tp and fp are the numbers of true positive and false positive predictions for the considered class.

No. of Vehicles →	32		54	
Clustering Algorithm	SOM	k-means	SOM	k-means
Accuracy	96.87%	96.87%	100%	98.29%
Precision Small	100%	100%	100%	100%
Precision Medium	100%	100%	100%	90.90%
Precision Large	75%	75%	100%	91.33%

No. of Vehicles →	75		100	
Clustering Algorithm	SOM	k-means	SOM	k-means
Accuracy	100%	100%	99%	98%
Precision Small	100%	100%	98.43%	96.87%
Precision Medium	100%	100%	100%	100%
Precision Large	100%	100%	100%	100%

Table 1: Experimental results

Class after clustering → Actual Class ↓	Small	Medium	Large
Small	63	0	0
Medium	1	25	0
Large	0	0	11

Table 2: Matching Matrix for SOM for a sample size of 100

Class after clustering → Actual Class ↓	Small	Medium	Large
Small	62	0	0
Medium	2	25	0
Large	0	0	11

Table 3: Matching Matrix for K means for a sample size of 100

With regards to colour identification of the vehicle, for a sample of 100 vehicles,
Accuracy = 90/100 = 90%
Precision Red = 6/8 = 75%
Precision White = 34/39 = 87.17%
Precision Grey = 50/53 = 94.33%

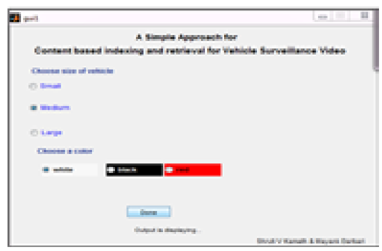

Fig. 2 Size of vehicle is chosen as 'medium' and color as 'white'

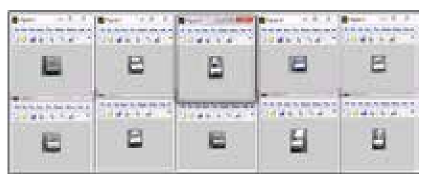

Fig. 3 Output is displayed

VI. CONCLUSION

As shown in the results, the combination of steps involved is unique. The vehicle extraction technique applied is an exclusive method implemented. The discussed method adopts content-based video retrieval for surveillance videos. The results show that it is an efficient, simple and quick method. In the future, we aim to detect different types of objects, incorporating more features for indexing.

REFERENCE

[1] Weiming Hu, Nianhua Xie, Li Li, Xianglin Zeng,, Stephen Maybank "A Survey on Visual Content-Based Video Indexing and Retrieval," IEEE Transactions on System, Man, and Cybernetics—Part C: Applications and Reviews, Vol. 41, No. 6, November 2011. [2] B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artificial Intelligence, vol. 17, pp. 185–203, August 1981. [3] B. K. P. Horn and B. G. Schunck, "Determining optical flow: A retrospective," Artificial Intelligence, vol. 59, pp. 81–87, Feb. 1993. [4] J. S.Boreczky and L.A. Rowe, "Comparison of video shot boundary detection techniques," in Proc, SPIE Conf, on Vision, Communication, and Image Processing. [5] Katharina Quast, Matthias Obermann and Andr´eKaup, "Real-Time Moving Object Detection In Video Sequences Using Gaussian Mixture Models," Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg. [6] Chang Liu, George Chen, Yingdong Ma, Xiankai Chen, Jinwu Wu, "A System for Indexing and Retrieving Vehicle Surveillance Videos," 4th International Congress on Image and Signal Processing, 2011. [7] P.M. Daigavane, Preeti R. Bajaj, M.B. Daigavane, "Vehicle Detection and Neural Network Application for Vehicle Classification," International Conference on Computational Intelligence and Communication Systems, 2011. [8] M.Babu Rao et al., " Content Based Image Retrieval using dominant color, texture and shape", International Journal of Engineering Science and Technology (IJEST), 2011. [9] Zhongwei Guo, Gang Wang, Renan Ma, Yulin Yang, "Battlefield Video Target Mining", 3rd International Congress on Image and Signal Processing, 2010. [10] Manimala Singha, K.Hemachandran, "Content Based Image Retrieval using Color and Texture", Signal & Image Processing: An International Journal (SIPIJ) Vol.3, No.1, February 2012.