

Detection and Monitoring of Passengers on a Bus by Video Surveillance

Boon Chong Chee, Mihai Lazarescu and Tele Tan
Curtin University of Technology, Western Australia
boonchong.chee@postgrad.curtin.edu.au
{m.lazarescu, t.tan}@curtin.edu.au

Abstract

This paper presents a method to detect passengers on-board public transport vehicles with the ultimate aim of monitoring their behaviours under suspicious circumstances. The method comprises first an elliptical head detection algorithm using the curvature profile of the human head as a cue. This is followed by applying the geometric blur features which are consistent to affine distortion of the image to keep track of the movement of the head within the vehicle. The profile of the moving heads with respect to each other within a length of time can then be used as indicative features to detect the advent of suspicious behaviour of the passengers.

1. Introduction

Vandalism on public transports is a perennial problem to transit authorities. Most public transport buses in countries such as UK, Canada and Australia have CCTVs installed on-board. Repairing vandalised properties and removing graffiti is costly. Measures to impede such unnecessary expenditure is imperative.

In response to increased vandalism on public transport systems especially on buses and trains, a great deal of money and efforts are being invested to heighten security in these areas. This can be realised by using strategically installed close circuit television (CCTV) cameras to monitor and track commuters' activities and interactions from the time of boarding to departure. While such technology is not new, the increased need and urgency for crime fighting measures has undoubtedly emphasised the importance of such cameras in public transports.

In such perspective, the merits of video surveillance systems on public transports include its use for (1) vandalism deterrence and (2) as evidential record for vandalism [6].

Vandalism is usually conducted under situations when opportunities present itself. Although performed discretely, there are several tell-tale behavioural signs prior to the act

of vandalism. Generally, passengers tend to participate in active movements such as switching of seats and large body motion gestures. These are exploitable cues that can be detected to raise an awareness to the situation. Unfortunately such psychologically motivated indications are not the focus of public transport-related surveillance studies.

There are major problems in the operation of video surveillance systems on buses. Due to limited concentration and awareness abilities, monitoring multiple long running video sequences by human operators is often expensive, tedious, error prone and unproductive. Furthermore, the video acquisitions are not processed until the buses have returned to the bus depot. As a result, no immediate or precautionary actions can be taken immediately after the events of vandalism or abnormal human behaviour have occurred. In the face of such challenges, the innovative use of automated and intelligent agents are advantageous in public transport surveillance technology.

In this paper, we present *an implementation of a video-based surveillance system to detect passenger movements on-board based on the psychological patterns of the passengers.*

This paper is organised as follows: In Section 2, we present a preliminary introduction to related contributions in human detection and tracking. This is followed by an explanation on the adopted method of approach in Section 3. The system evaluation and experimental results is discussed in Section 4. Subsequently, the project conclusion and possible future developments will be presented in Section 5.

2. Background

Specific to a bus scenario, stereotypical activities that can occur are (a) *seat switching* and a (b) *variety of posture transitions such as from sitting to standing, and vice versa*. These are two activities that are of particular interests in this project. Several situational and environmental constraints are involved in implementing a bus surveillance system to monitor passenger activities. Firstly, (1) video surveillance is operated on a constantly moving platform

as opposed to typical surveillances on static grounds, contributing to the (2) nondeterministic lighting and shadow pattern. (3) Passenger movements are usually short and restricted. (4) Passengers are often occluded behind an overcrowded bus and on-board furnitures. These commonalities bring forth the open research problems of background inconsistency, non-trivial object occlusion, drastic lighting and shadow issues which generally cannot be resolved by present methods. Considering these problems, the following paragraphs present image processing methods that have been contemplated for the project objectives.

2.1. Head detection

From observation of a typical bus footage, passengers are commonly occluded by seats and other passengers. While occlusion handling is the highlight of recent publications, human body detections is difficult and not viable in such circumstances especially when passengers' movements are both short and limited. This has motivated the project to focus on head detection techniques instead.

Methods such as template matching based on image correlation, active shape models and snakes are available. Template matching exhaustively scans for an object give. Unlike template matching, active shape models [7, 13] generates a parametric model of a shape based on the principle components of the average shape of an object. Matching is then based on estimating legal parameters constrained by the model. This allows a more robust shape match. On the other hand, snakes perform localization by forming a contour around the edges of the object based on 'energy' models that controls smoothness, elasticity and external sensitivities. Elliptical matching is another popular approach in head detection. The uncanny similarity of a human head contour to an ellipse has inspired several works [3, 11]. To achieve better robustness and accuracy, several techniques have included the study of color models characterising skin colors of diverse ethnicities as part of the detection process [3, 11].

2.2. Feature tracking

Human tracking and motion modelling is perhaps the critical task in a visual surveillance system. Algorithms such as condensation, particle filter and Kalman filter can be used to both track and predict the human motion based on predefined prediction and sampling models. These algorithms can be seen in the works of [9, 5, 12]. Stable tracking features such as scale invariant feature transform (SIFT) [14] have been introduced. SIFT generates descriptors from oriented filter responses within a window patch. The resultant descriptor is designed to be insensitive to scale, orientation and affine transformations. Several other tracking fea-

tures and their variants include Fourier descriptors, shape signatures and robust edge features.

From a higher level surveillance viewpoint, motion detection techniques such as optical flow and image differencing [8] are used in event detection and recognition. To discriminate between abnormal and normal events, classifiers that can be trained with machine learning algorithms like neural networks and Support Vector Machines (SVM) are used [15].

Despite the growing demand and clear benefits of automated video surveillance on public transports, inadequate work is performed with scenarios on-board bus in particular. Broadly, attempts of video surveillance on buses are limited to boarding passengers when buses are stationary at bus stops and lack 'in-journey' surveillance [1]. Other transport related surveillance tasks includes estimating geometrical positioning of passengers' head [10] and monitoring of crowd and individuals in public transportation areas [4, 15]. New methods to address the specific issues faced here must be developed.

3. Methodology

The proposed method of approach is illustrated in Figure 1. For a video sequence of f frames, each raw image frame at time $\mathcal{T} \in \{0, \dots, f\}$ is extracted for preprocessing following a head detection process that highlights head candidate regions. Subsequently, geometric blur descriptors [2] are obtained for each head candidate regions. These affine distortion tolerant descriptors are the key features for measuring and associating correspondence between head candidate regions that appear in other frames. Overtime, the dynamic evolution of the passengers' motion trajectories can be described. The details of the modules are presented in the following paragraphs.

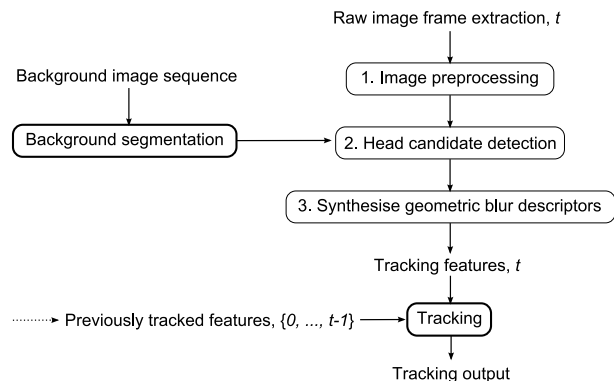


Figure 1. An overview of the proposed bus surveillance system.

3.1. Background segmentation

Background segmentation is performed on an empty bus scene to demarcate an area of interest for head detection in a reduced local search space. We assume that the location of a detected head on the seat and the pathway to be highly improbable. Hence, a region of seat and the pathway is segmented, leaving the rest of the background as a region of interest by using a standard Expectation-Maximization (EM) algorithm.

An average sequence of background images is convolved with a Gaussian smoothing kernel to produce a normalized background image. The EM algorithm is then performed on the intensity histogram of the normalized background image. Uniformly textured regions suggest the global grey-level distribution of the normalized background image to adhere a multivariate Gaussian mixture model (GMM), suitable for the application of EM to estimate the hidden model parameters. For the purpose of this segmentation, the EM algorithm is constrained to converge on a bimodal distribution density, with each Gaussian mixture component representing a region's intensity distribution. The converged GMM parameters by the E-M stepping are used to classify each pixel on the background image as either 'interest' region or 'seat and pathway' region. Figure 2a) shows the intensity mesh plot of the normalized background image and Figure 2b) shows the resulting segmented image.

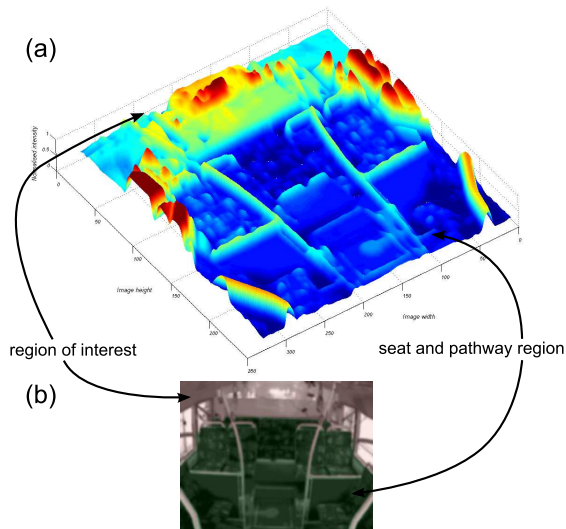


Figure 2. (a) 3D mesh plot of normalized background image. (b) Result of background segmentation using EM.

3.2. Elliptical head detection

As mentioned earlier, full body detection techniques for the purpose of tracking is not practical in a bus scenario. For example, only heads and shoulders are well within the camera's field of view when passengers are seated. A human head retains an elliptical shape contour under a variety of orientations offering itself a suitably good feature for elliptical matching. The head detection technique that is implemented in the bus surveillance system is a variant to that in [3] using an ellipse as a two-dimensional matching model.

Algorithms based solely on grey-level pixel intensities are not robust enough against illumination variations. Hence, an edge detector is applied on an input image to obtain object boundaries used for head detection. In [3], the measure for the goodness of a head match, based on cosine law, takes both intensity gradient orientation and magnitude into consideration:

$$\phi = \frac{1}{N} \sum_{i=1}^N |\mathbf{g}_i \cdot \hat{\mathbf{n}}_i| \quad (1)$$

where the score, ϕ , is calculated from the average of N absolute dot products of the unnormalized gradient orientation at pixel i , \mathbf{g}_i , and its corresponding unit normal vector of the matching ellipse, $\hat{\mathbf{n}}_i$. N is the total number of edge pixels along the perimeter of a matching ellipse. The best match for an input image is found by iterating through a search window under various ellipse sizes for a maximally valued score.

Modifications to the head detection method are necessary for its application in the bus surveillance system. Multiple smaller sized heads are required to be detected in a cluttered environment as opposed to detecting a single head as in [3]. Furthermore, full elliptical matching is more difficult where boundaries of a human head contour that appears in an edge image are generally shorter and drastically discontinuous, due to partial occlusion and poor edge detection results.

Consequently, an initial simple correlation template matching is performed for locating potential heads using a template of a typical size of a passenger's head. For each matched region, a vertically rotated duplicate is appended under to fabricate a pseudo-artificial ellipse. A least-squares ellipse fitting algorithm is subsequently applied on the fabricated ellipse. This approach allows a robust head fitting since no constraints on the elliptical parameters is administered during fitting. Finally, only the fitted ellipse contour residing in the original portion of the fabricate is retained. With the elliptical arc, a top-hemispherical head detection can be performed using Equation 1. In addition, fitted elliptical arcs with neighbouring local maxima are merged assuming adjacent arcs are associated to the same head.

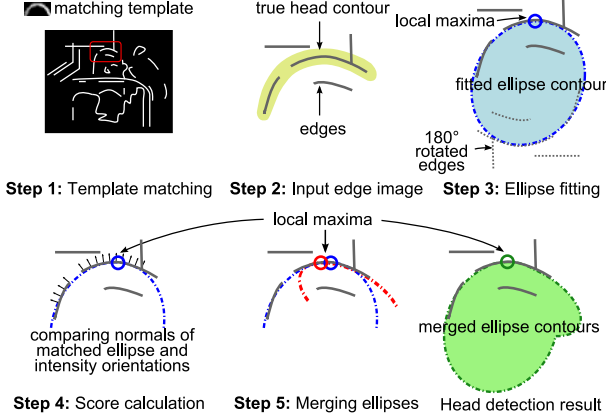


Figure 3. Procedural steps of the head detection module.

The head detection process is illustrated in Figure 3 and the sample detection results in Figure 8. Spurious detections may be ignored using basic shape descriptors such as elliptical roundness, aspect ratio, area and perimeter.

3.3. Geometric blur features

[2] describes geometric blur feature as a discriminative descriptor that averages geometrical transformations of a signal in a spatial domain. It is suitable for matching signals that have an affine relationship. The geometric blur feature is formed on a sparse signal (e.g. image’s first derivative) by means of spatially varying Gaussian kernel convolutions. The non-uniform kernel dimension relatively increases along with the sampling euclidean distances from a point of interest. Geometric blur features can be constructed on images with impoverished interest operators and hence, more applicable and flexible over feature descriptors such as SIFT [14].

After a head detection process, geometric blur features are used for establishing point correspondences for head candidate within a search window in other frame. Given a search window, a template geometric blur feature is first constructed around the local maxima of a head contour. This template feature is matched with other geometric features that are constructed around each edge pixel within a target frame. The confidence of each correspondence is determined by a match between two geometric blur features using the L_2 normalized correlation technique.

In constructing a geometric blur feature, an image window around a point of interest is convolved with a vertical, horizontal and cross oriented operators each producing an oriented edge filter response. A total of six sparsely signalled half-wave rectified channels (see Figure 4b)) are obtained by difference of Gaussian on individual oriented

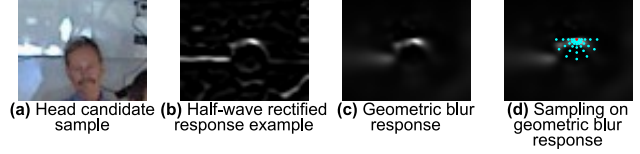


Figure 4. Snapshots of the geometric blur template matching process

response. Subsequently, each channel is geometrically blurred around a point of interest. Unlike [2], geometric blur feature matrix follows a restricted sub-sampling pattern shown in Figure 4d). The set of sampling points, \mathcal{S}_{x_0} , justifies an interest area under the local maxima of a contour and disregards the rest which possibly contain irrelevant background features. Furthermore, the matching is performed sequentially over each of three color channels instead of single grey-level channel. The best matching corresponding point is derived from the highest average score over three channels (refer to Algorithm 1). Following [2], the geometric blur descriptor around interest point x_0 of image I can be defined as:

$$G_{I_{x_0}}(x) = I * B_{\alpha|x_0-x|+\beta} \Big|_x \quad (2)$$

where $x \in \mathcal{S}_{x_0}$ and $B_{\alpha|x_0-x|+\beta}$ is a symmetric Gaussian kernel with α and β smoothing parameters.

Algorithm 1: Color template matching

input : template image, target image, x_0
output: point correspondent
foreach color channel $c \in \{\mathcal{H}, \mathcal{S}, \mathcal{V}\}$ **do**
 $templ \leftarrow c$ component of template image
 $target \leftarrow c$ component of target image
 initialise zero matrix R_c
 compute $G_{templ_{x_0}}$
 foreach sampled edge pixels at (i, j) in target **do**
 compute $G_{target_{i,j}}$
 $R_{c,i,j} \leftarrow \text{correlation}(G_{templ_{x_0}}, G_{target_{i,j}})$
 end
end
return $\arg \max_{i,j} (R_{\mathcal{H}_{i,j}} + R_{\mathcal{S}_{i,j}} + R_{\mathcal{V}_{i,j}})$

3.4. Motion detection

When the correspondent of a point is located, a track profile of a passenger’s motion trajectory can be established from the point to the nearest euclidean distanced head candidate from the correspondent. This is illustrated in Figure 5. Overtime, the entire motion trajectory of each passenger may be observed. As part of the surveillance system,

both trajectory lengths and displacements are monitored for activity detection.

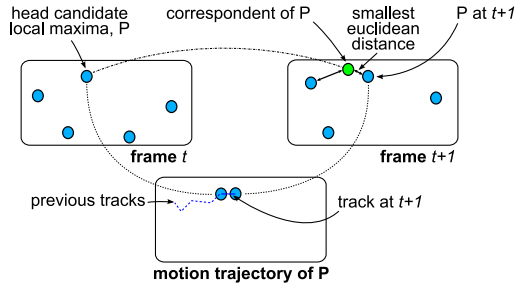


Figure 5. Snapshot of a tracking process

The bus surveillance system is designed to actively track motion while the bus is moving based on the displacement of an object. An object that has moved over a considerable distance is marked with motion streak as shown in Figure 7. The motivation of such a system is based on the psychological behaviour of passengers in a bus journey. During periodical stops, the system is set to a dormant mode while passengers are free to board and leave the bus. While a bus is travelling, passengers are typically well seated or standing still. During this period, detecting large motion would indicate abnormality.

4. Experiments and Results

To demonstrate the robustness and accuracy of the bus surveillance system, the implementation was tested over a variety of bus scenarios and alternative system setup. Three scenario enactments were used in the experiments for testing each with display rates of 25 *fps*. Video 1 (790 frames) shows a cluttered scenario of six passengers featuring passenger 1A switching seat locations. Video 2 (865 frames) shows another crowd of five passengers without major activity. Video 3 (835 frames) shows three passengers displaying abnormal behaviour (refer to Figure 6). The test video ground truths indicating true positions of passenger heads and appearance times were manually extracted. It was compared with the experimental results to measure the system's detection performance. The experiments were measured against 'correct' and 'incorrect' evaluation metrics. Each correctly tracked object was awarded with a 'correct' metric point for each correct track location. Upon occurrence of an incorrect track, the penalty was single one-off increment of the 'incorrect' metric until the object resumed its correct track.

An accuracy test on head detections for each video sequences was conducted based on average percentage of passenger heads correctly detected. Subsequently, true positive detections were brought forward to evaluate the tracking

procedure. The surveillance system was tested using the pioneer implementation of geometric blur feature computation using single grey-level channel and original sampling pattern. This will be compared against the proposed system of employing color template matching process using restricted sampling pattern. Finally, sample tracking results on Figure 7 attempts to demonstrate the ability of the proposed surveillance system to detect motion in a bus. The empirical results of the experiments are tabulated in Table 1. System with ellipse matching acceptance at 1.5 aspect ratio using HSV color template matching and restricted sampling pattern is assumed in the experiments unless otherwise specified.



Figure 6. Snapshots of video enactments

From the results in Table 1, the system shows an overall reasonable head detection accuracy. However, Video 2 has lower detection accuracy than other tests. This is apparently due the weak edge features of passenger 2A and 2E resulting in lower detection frequencies. Furthermore, movements nearer the camera tend to exhibit motion blur trails making edge detection difficult. Figure 8 shows snapshots of the head detection process. Although there are falsely detected heads, their locational persistences do not contribute to the eventual motion detection result. A comparison of the tracking methods has shown that color template matching with HSV model is predominantly better than its counterparts even though it is restricted by sampling pattern. Regardless of tracking methods, the system demonstrated good motion detection capabilities justified by fairly high counts of 'correct' tracks with occasional falsely detected motions. These false motions are often caused by moving shoulders and face regions coincidentally having the same curvatures of a head as shown in Figure 7. Most motions were detected during the tests. Even though the motion tracks are disconnected, the system detected significant motions enough to raise an awareness for Video 1 and Video 3, demonstrating its secondary capability as an alarm operator.

5. Discussion and future work

This project laid the basic foundation for a promising video surveillance system that can be operated within a bus. As a future extension, this can provide support for activity recognition that can compliment the current system in targeting the art of vandalism. Furthermore, the functionality

Table 1. Experimental results

Let K be the total number of passengers in a video

| | Results | | |
|----------------------------|--------------------|--------------------|--------------------|
| | Video 1 $K = 6$ | Video 2 $K = 5$ | Video 3 $K = 3$ |
| Head detection acc. | 88% | 68% | 86% |
| Tracking methods * | | | |
| Original ^a | 91% | 94% | 90% |
| Proposed-HSV ^b | 92% | 95% | 91% |
| Proposed-RGB ^c | 91% | 93% | 90% |

* as $\sum_{i=1}^K \frac{correct_i \times 100}{correct_i + incorrect_i} \%$, where $i \in \{1, \dots, K\}$
^a Grey-level processing using original sampling pattern
^b HSV color processing using restricted sampling pattern
^c RGB color processing using restricted sampling pattern

of motion detection can be extended to a full tracking system deployable on other suitable public transports such as trains. In the view of unstable edge features under adverse lighting conditions, it is also our motivation to explore other possible tracking techniques and features such as the KLT feature tracker [16] that are suitable for our bus scenario.

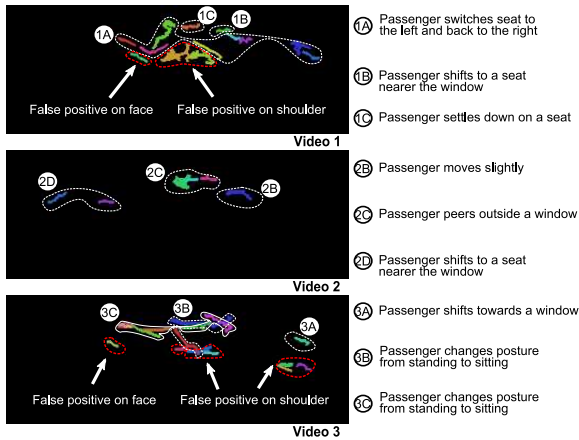


Figure 7. Motion detection results



Figure 8. Snapshots of head detection results

References

- [1] F. Bartolini, V. Cappellini, and A. Mecocci. Counting people getting in and out of a bus by real-time image-sequence processing. *Image and vision computing*, 12(1):36–41, Jan 1994.
- [2] A. C. Berg and J. Malik. Geometric blur in template matching. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 607–614, Kauai, Hawaii, Dec 2001.
- [3] S. Birchfi eld. Elliptical head tracking using intensity gradients and color histograms. In *IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, California, Jun 1998.
- [4] N. D. Bird, O. Masoud, N. P. Papanikolopoulos, and A. Isaacs. Detection of loitering individuals in public transportation areas. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):167–177, Jun 2005.
- [5] R. Bodor, B. Jackson, and N. Papanikolopoulos. Vision-based human tracking and activity recognition. In *Proceedings of the 11th Mediterranean Conferences on Control and Automation*, Rhodes, Greece, Jun 2003.
- [6] N. Brew. An overview of the effectiveness of closed circuit television (cctv) surveillance. *Research Note*, 14, Oct 2005.
- [7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape model - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan 1995.
- [8] J. W. Davis. Hierarchical motion history images for recognizing human motion. In *IEEE Workshop on Detection and Recognition of Events in Video*, pages 39–46, Vancouver, Canada, Jul 2001.
- [9] S. L. Dockstader and A. M. Tekalp. Multiple camera tracking of interacting and occluded human motion. *Proceedings of the IEEE*, 89(10):1441–1455, Oct 2001.
- [10] P. Faber. Image-based passenger detection and localization inside vehicles. In *Proceedings of the 19th International Society for Photogrammetry and Remote Sensing Congress*, pages 230–238, Amsterdam, Jul 2000.
- [11] J. Garcia, N. D. V. Lobo, M. Shah, and J. Feinstein. Automatic detection of heads in colored images. In *Proceedings in the 2nd Canadian Conference on Computer and Robot Vision*, pages 276–281, May 2005.
- [12] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.
- [13] S. Klim, S. Mortensen, B. Bodvarsson, L. Hyldstrup, and H. H. Thodberg. More active shape model. In *Image and Vision Computing*, pages 396–401, New Zealand, Nov 2003.
- [14] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 20:91–110, 2003.
- [15] C. Sacchi, C. Regazzoni, and G. Vernazza. A neural network-based image processing system for detection of vandal acts in unmanned railway environments. In *Proceedings of the 11th International Conference on Image Analysis and Processing*, pages 529–534, Palermo, Italy, Sep 2001.
- [16] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University, Apr 1991.