Automatic Classification of Abandoned Objects for Surveillance of Public Premises

AHMED FAWZI OTOOM, HATICE GUNES, MASSIMO PICCARDI

Faculty of Information Technology, University of Technology, Sydney (UTS)
Sydney, Australia
E-mail: {afaotoom, haticeg, massimo}@it.uts.edu.au

Abstract

One of the core components of any visual surveillance system is object classification, where detected objects are classified into different categories of interest. Although in airports or train stations, abandoned objects are mainly luggage or trolleys, none of the existing works in the literature have attempted to classify or recognize trolleys. In this paper, we analyzed and classified images of trolley(s), bag(s), single person(s), and group(s) of people by using various shape features with a number of uncluttered and cluttered images and applied multi-frame integration to overcome partial occlusions and obtain better recognition results. We also tested the proposed techniques on data extracted from a well-recognized and recent data set, PETS 2007 benchmark data set [16]. Our experimental results show that the features extracted are invariant to data set and classification scheme chosen. For our four-class object recognition problem, we achieved an average recognition accuracy of 70%.

1. Introduction

Automatic visual surveillance is concerned with obtaining a description of what is happening in a monitored area, and then taking appropriate actions based on that interpretation [4]. The assumptions and requirements of a visual surveillance system may vary depending on which specific area is under surveillance (e.g., airport vs. car park, city centre vs. road etc.) and what is monitored (e.g., an entire scene vs. moving objects).

The main modules in a video surveillance system (VSS, henceforth) involve object detection and tracking, object classification, and activity understanding. Although there exist some similarities, different visual surveillance systems consist of different processing steps. Up to date, different processing steps have been evaluated using different evaluation criteria [6]. While an increasing number of papers have started addressing the issue of how to perform quantitative comparison of existing algorithms, performance evaluation of VSSs is still an unresolved issue (i.e., how to perform objective/comprehensive/comparative evaluation, how to represent the complexity and range of issues handled, etc.).

Depending on which tasks the VSS is handling different evaluation criteria should be used at different stages:

1) Pixel-level evaluation (i.e., segmentation-based: moving objects vs. abandoned objects etc.).
2) Static object-level evaluation (i.e., evaluating per frame objects' features including object type, size etc.).
3) Dynamic object-level evaluation (i.e., evaluating per object the life time features including speed, trajectory etc.).
4) Behavior-level evaluation (i.e., event detection such as a person entering a room etc.).

The rationale behind using different evaluation criterion for each stage is the fact that errors obtained in one stage might well be carried along the successive stages increasing the overall error rate within a VSS. By separating the evaluation, we manage to identify and address limitations within each processing stage

In this paper our focus is on stage 2, where we aim to analyze an abandoned object's type or class. We propose an object classification method that uses information derived from a VSS. Thus, the work presented in this paper aims to become an integral part of a VSS framework that is able to track multiple people and automatically detect abandoned objects for security of crowded public places such as a railway station or an airport terminal. Although our aim is to use the output or results obtained from such a framework to further train an automated system to recognize abandoned objects, modeling the background and extracting objects, tracking them automatically and detecting the abandoned objects is not the focus of this paper.

In general, manual detection of an object provides superior segmentation results to a fully automatic segmentation approach and can be regarded as ground truth. Therefore, by using manually cropped abandoned objects we aim to avoid the detection error induced by the automatic abandoned object detection (i.e., inaccuracy in segmentation etc.). More specifically, our work is based on the assumption that the abandoned object is already detected; its location and size are provided as input. A commercial off-the-shelf technology product (e.g., [14]) can be used for this task.

We also assume that the area of interest comprises a floor area, and the input images are obtained using a camera installed near the ceiling or high on a post with typical tilt angle and resolution. We assume that the area of interest is located within an airport or train station, and the objects of interest consist of trolley(s), bag(s), single person and group(s) of people. Thus, our aim becomes that of classifying an abandoned object into one of the predetermined categories. We introduce and experiment with various features in order to correctly discriminate the aforementioned objects in a robust manner. A classifier is then built based on these features and the output and results of the classification are presented and discussed.

The remainder of this paper is organized as follows. Section 2 describes the related work in the area of object classification for visual surveillance applications and section 3 focuses on the methodology, by presenting the feature extractors utilized, the features extracted and the rationale behind these procedures. Section 4 describes the experiments performed and the classification process. Finally, section 5 concludes the paper by outlining the future work.

2. Background and related work

The main structure of any VSS involves detection, tracking, classification and recognition of objects. The objects involved in a VSS can be either moving (e.g., people, groups of people etc.) or abandoned objects (e.g., suitcases, trolleys etc.).

The first component of a VSS is detection, where the regions belonging to the object are detected using various methods such as background subtraction techniques or frame differencing methods [5]. After detection, objects are tracked from frame to frame using distinctive features associated with these objects.

Classification is another important component of any VSS where objects are classified into different categories of interest. Generally, there are three main approaches to classification: shape-based classification (e.g., [1, 8-10, 12]), motion-based classification (e.g., [3]) or combined shape-motion classification (e.g., [2, 6, 13]).

To date, in the visual surveillance research area, the literature reports attempts to analyze four main categories of objects, namely, person, vehicle, group of people, and package (e.g., [1, 2, 6,8-10, 12-13]).

For instance, in [8, 9, 12], moving objects are classified into either a person or a vehicle according to their shape-based features. In [8], the authors use the dispersion and area features as a metric for classification. They assume that the dispersion value for a person is generally higher than a vehicle. The authors classify the object multiple times in successive frames before having the final decision of the object category. This method is expected to be useful in partially overcoming the occlusion problem. However, at times, group of people are misclassified as a vehicle as they may have the same dispersion value. In [12], the authors propose the use of two features for classifying objects as a person or a vehicle. These features are the (height/width) ratio and the number of corners. This method works only when objects are well separated from each other and hence misclassification can occur if there is a group of people with people occluding each other. In [9], two more features are added to the (height/width) ratio used by [12]; these two features focus on the relative relationship between the size of the object and the size of the bounding box. The authors further classify the object labeled as 'person' into one person, two persons, or three persons categories. A similar strategy is used for the classification of object labeled as 'vehicle'. They build one classifier based on Bayesian inference and another one based on neural networks. When the results are compared, the neural network seems to outperform the Bayesian classifier in terms of classification accuracy. However, the results could have been improved if the authors considered a higher number of motion features similar to these proposed in [2], where the reported classification accuracy, for a two-class problem of person and vehicle, is 92% without normalizing, and 97% with normalization. The high accuracy achieved is due to the combination of motion related features including magnitude of velocity, direction of motion, and average recurrent motion image (RMI). In [6] the authors use shape features combined with the RMI motion feature for classifying objects in a hierarchical manner. In the first part of the experiment, objects are classified into a (single person/groups) category or vehicles category and a classification accuracy of 100% is obtained. In the second part of the experiment, objects that were classified as single person/groups are further classified into a single person or a group of people. The single person category is recognized with 100% accuracy and groups of people with 87.5% accuracy. Thus, in each part of the experiment, the authors target a two-class problem.

In [13], detected moving objects are classified into four categories: a person, a group of people, a vehicle, or a bicycle using both shape and motion features. This problem is more challenging than a two-class classification problem because the new classes are more likely to lay in between the other two classes in the feature space. Two main features are used for classification, namely, the variation of motion and the variation in compactness. The first feature seems good in distinguishing a person from a vehicle while the second can discriminate a person from a group of people. The authors claim that the results of the classification are high; however, the visualization results of the data in the feature spaces do not show good discrimination between the bicycle, group of people, and vehicle classes.

In [1, 10], the focus is on the classification of abandoned objects. In order to classify abandoned objects, only shape features can be used because the objects are static and no motion occurs. In [10], two important classes are studied: person and abandoned packages. Two main features are used for this classification: the x-elongation and the y-elongation of an object. This is based on the idea that an abandoned package may have similar elongation values whereas a person exhibits different elongation values. These features are then fed into a neural network and a classifier is built based on them. In [1], the authors classify abandoned objects into three main categories: person, package or unknown category. Similar to the work introduced in [8], the area and compactness features are used again for classification. The area feature represents the number of pixels belonging to the object and the compactness feature represents how the shape of the object is stretched out. A Bayesian classifier is then built based on these features. In both works (i.e., [1] and [10]) the same challenging problem may arise where classification accuracy decreases in the case of occlusion or when there are multiple objects in the same scene. Generally speaking, the feature sets used account for relatively simple features. It has to be kept in mind that the low spatial resolution typical of surveillance frames prevents the realistic extraction of highly detailed features. As for the categories, although in airports or train stations abandoned objects are mainly luggage or trolleys, none of these previous works have attempted to classify or recognize trolleys. Compared to these previous works, in this paper, we: (i) analyze and classify the images of trolley(s), bag(s), single person(s), and group(s) of people, (ii) analyze and experiment with various features and define which one(s) are more significant than the rest, and (iii) train and test classifiers both on uncluttered (images with clean background) and cluttered (images segmented out from the background in real videos) data and compare their classification accuracy.

3. Methodology

Our aim is to use the output or results obtained from a visual surveillance framework to further train an automated system to recognize abandoned objects. Our work is based on the assumption that the abandoned object is already detected; its location and size are extracted previously and passed onto our object recognition machine. The object recognition component presented in this paper consists of feature extraction, training classifier(s), testing and evaluation. These steps are explained in detail in the following subsections and in section 4.

3.1. Feature extraction

In general, human beings distinguish objects from each other by taking into account many varied criteria. We follow this rationale and choose to use simple yet efficient features in order to obtain a discriminative feature set that will help differentiate between the predetermined categories (trolley(s), bag(s), single person(s), and group(s) of people). There are many features that can contribute to a greater or lesser extent to the recognition of these objects. However, we aim to detect features that are relatively stable, consistent across different categories of objects, easy to extract and calculate, and useful for the classification process with some discrimination content. To this aim, we performed analysis of 124 input images (31 for each category) obtained from various sites on the Internet.

Trolley: Experimentally we found out that the images of trolleys are characterized by containing a relatively high number of relatively closely packed straight lines (vertical, horizontal and/or diagonal lines). Therefore, we use a line edge detector to detect lines within the region of interest. We calculate the number of weak, intermediate and strong lines detected. A trolley usually has higher number of strong lines compared to a bag, person or group. Moreover, for the analysis of the trolleys, we found out that trolleys have higher number of corners compared to person or bag categories. Therefore, we detect corners and calculate their spatial distribution over the input image in terms of ratios (e.g., top vs. bottom half of the image, left vs. right half of the image etc.). We also calculate the standard deviation over the distribution of the corners both horizontally and vertically. Trolleys also contain circles in the lower half of the image (close to the location of the wheels). Therefore, we use a circle detector to detect the circles within the region of interest. We also calculate ratios of the numbers of circles located in different parts of the image. Moreover, we calculate how the radius of such circles deviate from each other to study the uniformity of their size.

Single Person: For detection of *person* we found out that the person category in general has intermediate number of corners around the head, hands and arms, the center of the body, legs and toes. The person category also has a limited number of vertical lines depending on the posture (e.g., if (s)he is standing with open legs, up to two vertical/diagonal lines; if (s)he is standing with closed legs one strong line crossing the centroid of the body etc.). Moreover, the person category in general has one circle on the upper-half of the input image where the head is located.

Group of people: For detection of groups of people, we found out that the number of heads detected relatively close to each other in the specified region of interest can prove useful. As our system detects multiple persons forming a group if they are relatively close to each other and connected as components, the number of circles detected and the standard deviation of these circles can be used as significant features for recognition.

Bag: For the bag case, we found out that bags have higher number of corners around the handles, zippers and wheels (if there are any). Moreover, a roller bag may contain a small number of circles in correspondence with the handles' shape and the small wheels in the bottom part. They also have fewer lines around the edges or boundaries. Lines detected in a bag image are directly related to the positioning of the bag (e.g., bag standing upright, bag tilted to the right etc.). For instance, if the number of vertical lines is higher than any other types of lines, then the probability that the bag is standing vertically becomes high.

The results obtained from an initial experiment show that for trolleys, bags and humans, the detected circles will not differ significantly in terms of size. In other words, in an input image we expect to find circles of similar size, without much variation in terms of size. We use this feature to constraint the size of the circles detected in an input image. Larger, spurious circles can thus be eliminated by calculating the standard deviation of the circles and removing those that are far from the value of the standard deviation (a simple yet effective outlier elimination technique).

Table 1. List of all features extracted

Features
Corners: - Number of corners. - Percentages and ratios of corners in different parts of the image. - Horizontal and vertical standard deviations of the corners.
Lines: - Number of lines (strong, intermediate, and weak). - Number of horizontal, vertical, and diagonal lines, and the ratios between them.
Circles: - Number of circles. - Percentages and ratios of circles in different parts of the image. - Horizontal and vertical standard deviations of the circles.
Compactness: the compactness value is calculated as the perimeter^2/ area. This value is calculated relative to the image size.
Height/Width ratio.

Although we do calculate the width and height of the object, we assume that classification based upon this measurement alone might not work, as in real-life scenarios the distance between the camera and the object of interest varies for different cases. In other words, a trolley that was detected as a small object can be detected with much larger height and width in another sequence, thus not proving to be useful features on their own. However, we calculated the height/width ratio which proves useful in recognition.

The compactness feature of an object shows how much an object is elongated. In our case, the compactness value for each object can be calculated as the ratio of (perimeter^2/ area). This value is calculated relative to the size of the image as images are not re-scaled to have the same size.

After identifying a set of justifiable features over all four object categories, features such as lines, circles, and corners are extracted using various functions available within the Open Source Computer Vision Library (OpenCV) that is freely available for research purposes [15].

Lines are extracted by applying an edge detector. The straight lines are then detected using the Hough transform for lines. Circles are also detected using a Hough transform function. All other related statistical features are then calculated to form the final feature vector. A complete list of all the features extracted is illustrated in Table 1. Moreover, Figure 1 shows examples of how lines, circles, and corners are detected in a number of input images.

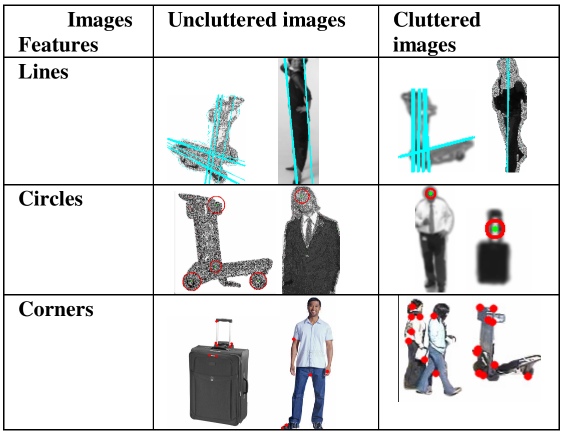


Figure 1. Examples of features detected in both uncluttered images from the internet and cluttered images from videos

3.2. Classification

This section is concerned with the classification and performance evaluation procedures utilized in our system.

3.2.1. Classifiers. The classifiers that have been used for the classification experiments in our system are the Bayesian-based classifier BayesNet, C4.5 or Decision Trees, and the Sequential Minimal Optimization (SMO) algorithm [11]. BayesNet is the Bayesian Network classifier available as part of the WEKA package, a publicly available toolbox for automatic classification [11]. BayesNet enables the use of a Bayesian Network learning using various search algorithms and quality measures. It provides data structures (network structure, conditional probability distributions, etc.). Various estimator algorithms for finding the conditional probability tables of the Bayesian Network can be used, namely, the SimpleEstimator, BMAEstimator etc. C4.5 is a class for generating a pruned or unpruned C4.5 decision tree. C4.5 is a supervised symbolic classifier based on the notion of entropy since its output - a decision tree - can be easily understood and interpreted by humans. SMO class in WEKA implements Platt's sequential minimal optimization algorithm for training a support vector classifier. This implementation transforms nominal attributes into binary ones and normalizes all attributes by default. Multi-class problems are solved using pair wise classification.

3.2.2. Performance Evaluation. The performance of the classifier is evaluated in terms of classification accuracy that is calculated as the proportion of the number of objects correctly detected against the total number of objects. Both N-fold cross-validation and holdout estimate are performed. In N-fold cross-validation the data are divided into N subsets of (approximately) equal size. The classifier is then trained N times (N=10), each time leaving out one of the subsets from training, but using only the omitted subset to compute the error criterion in question. The holdout method splits the data into two mutually exclusive sets, the training and test sets. The classifier is designed using the training set and performance evaluated on the independent test set. A number of experiments are then conducted and the results are illustrated in section 4.

4. Experiments

Our system has been tested on two data sets: the PETS 2007 data set (with different situations, diverse number of people and different types of luggage [16]), and a mixed data set consisting of 184 images of empty trolleys, bags, persons and groups of people. The PETS 2007 benchmark data set was chosen as it is a recently released and well-recognized data set within the visual surveillance community. In the context of PETS 2007, *abandoned object* is defined as items of luggage that have been abandoned by their owner (i.e., handbag, carry-on case, 70 litre backpack and ski gear carrier). However, the PETS 2007 benchmark data contains only two scenarios of abandoned objects (i.e., sequences S7 and S8). Sequence S7 contains a single person with two bags. The individual enters the scene, stops in the middle of the scene, before walking away whilst accidentally leaving one bag on the ground. The bag owner then returns to the scene to retrieve the bag. Sequence S8 contains an individual who enters the scene carrying a large bag, which is placed on the ground. The owner then walks away from the bag before retrieving it, and leaving the scene.

Although persons, groups of people, empty or loaded trolleys are not within the scenarios or definition of PETS 2007 left-luggage category, we believe the classification of these is as important as bags from the point of view of a risk assessment procedure. A scenario with these objects would be very similar to the scenarios of S7 and S8, and therefore, classification of these objects could be well utilized in (real-world) public places to raise appropriate alarms.

Due to the aforementioned reasons we extend the problem of left-luggage classification into four categories: trolleys, bags, persons and groups of people. As there exist insufficient number of sequences (only S7 and S8) in the context of left luggage, we use other existing sequences in order to obtain a sufficient number of training and testing samples for these four categories.

After obtaining the feature vectors for all input images, we perform a set of experiments in order to see how classification was affected under various criteria.

4.1. Experiment 1: Invariance to data set and classification technique

This first experiment aims to explore the following issues that are reinforced by the features chosen and extracted:

- *Invariance to data set:* the invariance of our feature extractors and feature set to different data set(s) is tested. In other words, we aim to find out whether *not* tailoring our feature extractors to a specific data set affects the classification results for different data sets.
- *Invariance to classification technique:* the invariance of our feature extractors and feature set to different classifiers is tested. In other words, we aim to find out whether using different classification algorithms significantly affects the classification accuracy for different data sets.

To this aim, we used two data sets, various classifiers and 10-fold cross-validation. The first data set consisted of 125 images of empty and loaded trolleys, bags, persons and groups of people from PETS 2007 benchmark data set varying in size, type and view angle. Please note that we did not use multiple images of a single object (i.e., the same object appearing in consecutive frames) as they would have likely appeared in both the training and test folds, thus leading to overly optimistic estimates of the error rates. Instead we intentionally collected a single image for each different physical object (from S1-S8 in the PETS 2007 data set). The second data set consisted of 184 images of (mostly) empty trolleys, bags, persons and groups of people. 124 of these were uncluttered images collected downloaded from the WWW and 60 were cluttered images that were clipped from real videos taken in an airport and provided by the industrial partner of the project.

As a first step we experimented with three classifiers, namely BayesNet, Decision Trees (C4.5) and SMO and only the PETS data set. We then experimented with the same three classifiers using 309 images obtained by mixing the PETS 2007 data set and our own data set. Note that the inclusion of the PETS 2007 data in the data set generally increases the within-class variance within the feature set. The comparative results are presented in Table 2.

By looking at Table 2, we observe that the average classification accuracy obtained for the two data sets is approximately the same (70% vs. 69%). Therefore, it is possible to conclude that the features we have chosen are sufficiently robust to variation in illumination, type, size and view angle. Moreover, the accuracy achieved based on this feature set seems to be almost invariant to the different classification algorithms tested. This in turn implies that we can apply our method to various environments and conditions without the need to re-tailor it.

Table 2. Classification results on two data sets using various classifiers with 10-fold cross-validation.

	Classification Accuracy (%)	
Classifier type	PETS	Mixed
BayesNet	70.4	67.9
C4.5	72	66.9
SMO	68.8	73.1
Average	70.4	69.3

4.2. Experiment 2: Handling occlusions

In this section, we focus on abandoned objects that are subject to temporal and/or spatial occlusion and present results of two experiments.

4.2.1. Temporal occlusion. An abandoned object undergoes temporal occlusion when another object such as a person or person with a bag/trolley moves in front of it for a certain period in time (short term occlusion). The main assumption here is that the occluding object is not stationary and moves along with approximately linear speed. This enables the partial or full observation of the abandoned object at least at some stage. The purpose then becomes that of classifying the abandoned object correctly despite the occlusion assuming that the correct class is the most frequently recognized one over a number of frames. To this aim we choose to experiment with a multi-frame integration scheme and sequence S8 from the PETS 2007 data set. The final decision is made based on a multi-frame integration approach, where single frame recognition results are combined by first calculating the total number of recognized frames for each class and then choosing the class with the maximum value as the final decision. Let x be the class of an abandoned object at frame i and $d(x \mid f_i)$ be the binary decision $(0 \mid 1)$ for frame i given feature vector f_i. Since x is one of the four classes (bag, trolley, single person or group of people) then $d(x \mid f_i) = 1$ for only one class and 0 for all the others. In general, the number of frames to be integrated will depend on the frame rate. In our case we just choose an arbitrary number of frames for experimental purposes denoted as T. For each class the multiple decisions are added up as: $D(x \mid f_1..f_T) = \sum_{1}^{T} d(x \mid f_i)$. The multi-frame integration approach can then be described simply as:

$$x^* = \underset{x}{arg\ max}\left(D(x \mid f_1..f_T)\right) \qquad (1)$$

For this experiment, the abandoned object was manually extracted from 111 consecutive frames in sequence S8. Starting from frame no. 1330, the object was clipped from every second frame resulting in 56 images of the abandoned object. Within these 56 frames it was occluded partially or fully for 32 frames. Using the methods introduced in the previous sections we extracted the feature vectors for each frame, trained BayesNet with the mixed data set of 309 images and tested it on the aforementioned 56 frames (i.e., Holdout validation). We obtained an overall recognition accuracy of 94%, 53 out of 56 frames were correctly classified as 'bag'. Rather surprisingly, the classifier was able to correctly classify the abandoned object as 'bag' even when it underwent significant occlusions. This might be explained by the fact that the occluding object was mostly a person carrying another bag, thus the object was still containing typical features of class 'bag'. Figure 2 illustrates some examples for the abandoned object undergoing different types of temporal occlusions and the classification results obtained for each frame.

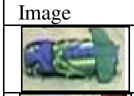
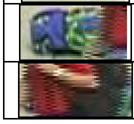
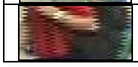
Image	Occlusion type	Class
	No occlusion	Bag
	Partial	Person
	Full	Bag

Figure 2. Example images for the abandoned object undergoing different types of occlusions and the classification results.

4.2.2. Spatial occlusion. An abandoned object undergoes spatial occlusion when another object such as a person or person with a bag/trolley moves in front of it for unknown periods of time (short or long term occlusion). The main assumption here is that the occluding objects might be of moving or stationary nature, thus the abandoned object might be partially or fully occluded. The purpose then becomes that of classifying the abandoned object correctly for each frame and identifying how much occlusion affects the classification drastically.

In order to test the robustness of the classifier at correctly classifying the abandoned object under different types of occlusions, and how spatial occlusions affect classification in each frame, we again used frames for the abandoned object from sequence S8. This time, however, we manually overlayed an occluding object in front of the abandoned bag. In other words, a small part of a trolley from PETS 2007 was extracted and manually placed in front of the abandoned bag to occlude it, in different positions and for a number of frames. Again using the methods introduced in the previous sections, we extracted the feature vectors for 30 frames (where the object was occluded for 9 consecutive frames), trained BayesNet with the mixed data set of 309 images and tested it on the aforementioned 30 frames. We obtained an overall recognition accuracy of 90%, with 27 out of 30 frames correctly classified as 'bag'. The classifier was able to correctly classify the abandoned object as 'bag' as long as it did not undergo extensive occlusions. Under such occlusions the abandoned object was classified as either a 'trolley' or 'groups of people'. Figure 3 illustrates some examples for the abandoned object undergoing different types of artificial occlusions and the classification results obtained for each frame.

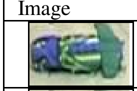
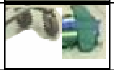
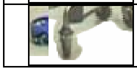
Image	Occlusion type	class
	No occlusion	Bag
	Partial	Bag
	Full	Trolley

Figure 3. Example images for the abandoned object undergoing different types of artificial occlusions and the classification results.

As illustrated in Figure 3, the abandoned object is classified correctly despite being partially occluded. However, when the object is occluded for more than 2/3, as the features are not accurately extracted the classifier outputs incorrect results. By applying a multi-frame integration approach (described in Eq. 1) the spatial occlusion problem can be handled and the final result obtained from the classification is the class 'bag'.

However, we note that the results obtained in these experiments relate to these two cases and cannot be easily generalized to other conditions or types of abandoned objects without extensive experimentation. The recognition accuracy might vary significantly if the abandoned and the occluding objects are of different nature or type.

5. Conclusions and future work

Although in airports or train stations, abandoned objects are of specific, known categories such as luggage or trolleys, none of the existing works in the literature has tried to identify a general and robust feature set allowing the recognition of all such categories with good accuracy. In this paper, we analyzed and classified images of trolleys, bags, persons, and groups of people by using an original and rich combination of shape features and applied multi-frame integration to overcome partial occlusions and obtain improved recognition results. We evaluated the proposed techniques on the PETS 2007 benchmark data set which consisted of eight scenarios, and correctly predicted the class of the objects (assumed to be abandoned) with an overall recognition accuracy of 70%.

The results are encouraging considering that a four-class problem in crowded environments is highly challenging with objects located far from the camera(s) and relatively low image quality. Results are likely to improve with higher image resolution and where multiple views of a single object are available. The categories into which objects are classified can be extended further by incorporating the following criteria: group size (in terms of numbers of people), person with/without a trolley, person with/without a bag, etc. Creating a system wherein the output result is combined with other data relative to the area of interest (e.g., the carrier) in order to enable a richer analysis is the focus of future research.

6. Acknowledgements

This research is supported by the Australian Research Council and iOmniscient Pty Ltd under the ARC Linkage Project Grant Scheme 2006 - LP0668325.

7. References

[1] M.D. Beynon et al., "Detecting abandoned packages in a multi-camera video surveillance system", *Proc. of the IEEE Advanced Video and Signal Based Surveillance Conference*, pp. 221-228, 2003.

[2] L.M. Brown, "View independent vehicle/person classification", *Proc. of ACM 2nd International Workshop on Video surveillance & Sensor Networks*, NewYork, USA, pp. 114-123, 2004.

[3] R. Cutler and L.S. Davis, "Robust real-time periodic motion detection, analysis, and applications", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 22, No. 8, pp. 781-796, 2000.

[4] A. R. Dick and M. J. Brooks, "Issues in automated visual surveillance", *Proc. of 7th Digital Image Computing: Techniques and Applications Conference*, pp. 195-204, 2003.

[5] W. Hu et al., "A survey on visual surveillance of object motion and behaviors", *IEEE Transactions on Systems, Man and Cybernetics-Part C*, Vol 34, No. 3, pp. 334-352, 2004.

[6] O. Javed and M. Shah, "Tracking and object classification for automated surveillance", *Proc. of European Conference on Computer Vision*, pp. 439-443, 2002.

[7] N. Lazarevic-McManus, J. Renno, D. Makris and G.A. Jones, "Designing evaluation methodologies: The case of motion detection", ", *Proc. of IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, pp. 23-30, 2006.

[8] A.J. Lipton, H. Fujiyoshi and R.S. Patil, "Moving target classification and tracking from real time video", *Proc. of IEEE Workshop of Applications on Computer Vision (WACV)*, pp. 8-14, 1998.

[9] R.J. Oliveira et al., "A video system for urban surveillance: Function integration and evaluation", *Proc. of International Workshop on Image Analysis for Multimedia Interactive Systems*, 2004.

[10] E. Stringa and C.S. Regazzoni, "Real-time video-shot detection for scene surveillance applications", *IEEE Transactions on Image Processing*, Vol 9, No. 1, pp. 69-79, 2000.

[11] I. H. Witten and E. Frank, *Data mining: Practical machine learning tools with java implementations*, Morgan Kaufmann, San Francisco, CA, 2000.

[12] Q. Zang and R. Klette, "Object classification and tracking in video surveillance", *Proc. of Computer analysis of Images and Patterns*, Springer Berlin / Heidelberg, pp. 198-205, 2003.

[13] Q.A. Zhou and J.K. Aggarwal, "Tracking and classifying moving objects from video", *Proc. of IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, 2001.

[14] Motion Detection Technology: http://www.iomniscient.com/ products_iq140.htm (Access date: 1 November 2007).

[15] OpenCV: http://sourceforge.net/ projects/ opencvlibrary (Access date: 1 November 2007).

[16] PETS 2007 benchmark data set: http://pets2007.net/ (Access date: 1 November 2007).