

International Joint Conference on Computer Vision and Computer Graphics Theory and Applications. 22-25 January 2008, Funchal, Madeira - Portugal.

FEATURE SETS FOR PEOPLE AND LUGGAGE RECOGNITION IN AIRPORT SURVEILLANCE UNDER REAL-TIME CONSTRAINTS.

J. Rosell-Ortega,
G. Andreu-García,
A. Rodas-Jordá,
V. Atienza-Vanacloig,
J. Valiente-González

DISCA. Universidad Politécnica de Valencia. Camino de Vera s/n Valencia. Spain
jarosell@doctor.upv.es, gandreu, arodas, vatiienza, jvalient{@disca.upv.es}

Abstract

We study two different sets of features with the aim of classifying objects from videos taken in an airport. Objects are classified into three different classes: single person, group of people, and luggage. We have used two different feature sets, one set based on classical geometric features, and another based on average density of foreground pictures in areas of the blobs. In both cases, easily computed features were selected because our system must run under real-time constraints. During the development of the algorithms, we also studied if shadows affect the classification rate of objects. We achieved this by applying two shadow removal algorithms to estimate the usefulness of such techniques under real-time constraints.

Acknowledgments

This work has been partially supported by SENSE project (Specific Targeted Research Project within the Thematic priority IST 2.5.3 of the 6th Framework Program of the European Commission: IST Project 033279), by the Spanish Government by the complementary funding TIN2007-30367-E, and Conselleria d'Empresa, Universitat i Ciència of the Regional Government under the grant ACOMP/2007/205.

SENSE



FEATURE SETS FOR PEOPLE AND LUGGAGE RECOGNITION IN AIRPORT SURVEILLANCE UNDER REAL-TIME CONSTRAINTS.

J. Rosell-Ortega, G. Andreu-García, A. Rodas-Jordà, V. Atienza-Vanacloig, J. Valiente-González
DISCA. Universidad Politécnica de Valencia. camino de Vera s/n Valencia. Spain
jarosell@doctor.upv.es, gandreu, arodas, vatiienza, jvalient@disca.upv.es

Keywords: surveillance, object classification, object recognition, shadow removal, feature sets

Abstract: We study two different sets of features with the aim of classifying objects from videos taken in an airport. Objects are classified into three different classes: *single person*, *group of people*, and *luggage*. We have used two different feature sets, one set based on classical geometric features, and another based on average density of foreground pictures in areas of the blobs. In both cases, easily computed features were selected because our system must run under real-time constraints. During the development of the algorithms, we also studied if shadows affect the classification rate of objects. We achieved this by applying two shadow removal algorithms to estimate the usefulness of such techniques under real-time constraints.

1 INTRODUCTION ¹

In this paper we present a comparative study regarding classification within the framework of visual surveillance. Our aim is to classify each moving object in the input video as a single person, a group of people or unattended luggage.

Visual surveillance, either indoors or outdoors, is an active research topic in computer vision and various surveillance systems have been proposed in recent years: (Haritaoglu et al., 2000), (W. Hu et al., 2004), (Wren et al., 1997). The visual surveillance process may be divided into the following steps: environment modelling, motion detection, object classification, tracking, behaviour understanding, human identification and data fusion.

In our application, airport surveillance, attention must be paid to individuals standing alone, groups of people, and luggage. The system consists of a set of cameras with a DSP running low-level vision algorithms, covering the complete airport. One of their tasks is classifying objects in each frame by considering just the information gathered through the camera

and a limited local history, if any. This classification is then fed to a higher level which controls several cameras and fuses all the incoming data.

It is important that the features used to give this initial classification are quickly computed, and crucial to have a high local rate of successful classification; despite higher levels may correct local classification errors. We used two different feature sets. Geometric features, which are mentioned in many papers discussing surveillance systems and another set of features based on the average density of foreground pixels in areas of the blobs. Both sets of features attempt to extract the essence of object silhouettes. Unfortunately, shadows can connect separate objects and deform shapes. We investigate the suitability, or not, of using shadow removal algorithms for obtaining higher classification success rates in real-time systems.

In section 2 we explain how we extracted figures from videos and converted them into blobs; shadow removal algorithms are introduced in 3, the feature sets we used are explained in section 4. Section 5 shows the experiments we made and the obtained results. Finally, 6 present our conclusions and future works.

¹Acknowledgments: This work is supported partially by the sixth framework programme priority IST 2.5.3 Embedded systems. Project 033279.

2 BLOB SEGMENTATION

Some considerations must be borne in mind regarding blob extraction:

- One blob-one object: each blob must contain a complete object and each complete object must be included into a blob.
- Different objects cannot be in the same blob, if they are separate in the frame.
- Different objects can be included in the same blob if there is occlusion or overlapping between them.

Detecting blobs involves learning an adaptive background model. Various background learning techniques may be found in (L. Wang and Tan, 2003). These background model will be used as a reference in background substraction. The scheme for detecting motion also involves temporal differences between three successive frames. Only those blobs whose mass exceeds a threshold are considered to be objects, discarding the rest. Blobs are joined according to proximity rules. For each blob, the smallest box enclosing it (called bounding box, BB) is calculated. Some samples may be seen on figure 1.

3 SHADOW REMOVAL

Several shadow removal algorithms have been developed (Rosin and Ellis, 1995), (Javed, 2002), (Onoguchi, 1998). Removing shadows is important in order to improve object disambiguation and classification. Shadows can be of two types:

- Self shadows: these are part of the object which are not illuminated by direct light. They will not affect the shape or silhouette of an object, so we will ignore them.
- Cast shadows: these are the area in the background projected by the object in the direction of light rays. They affect seriously the shape of blobs by enlarging them or joining several blobs together.

Algorithms introduced in (Rosin and Ellis, 1995) for grayscale images and in (Xu et al., 2005) for RGB space were used, testing the gain of the RGB space algorithm over grayscale approach.

4 FEATURE SETS

We used two different feature sets: *geometric features* and *foreground pixel density*, introduced in the following sections.



Figure 1: Captured images of different instances of *single person*, *group of people* and *luggage*, with their associated blob and the blob after removing shadows in gray levels and RGB space.

Geometric features are usually discussed when dealing with classification aspects; we chose the most efficient in classification success rate and computational cost to meet the real-time response constraint. We used the following:

- Aspect ratio: given by $\frac{BB_{height}}{BB_{width}}$.
- Dispersedness: given by $\frac{Perimeter^2}{Area}$.
- Solidity: computed as $\frac{Area}{ConvexArea}$.
- Foreground ratio: $\frac{foreground\ pixels}{Total\ amount\ of\ pixels}$

Objects of different classes show different pattern of occupancy, some samples are shown in figure 2.

We divide a bounding box into a grid of $n \times m$ cells; introducing this way the requirement of scale invariance. For each cell, the amount of foreground pixels (P_i) and background pixels (B_i) is calculated and the result of the division $\frac{P_i}{P_i+B_i}$ is stored.



Figure 2: Sample of person, group of people, and luggage with a 4×4 grid.

5 EXPERIMENTS AND RESULTS

Off-line experiments were made with data obtained from indoor video sequences taken on different days with different light conditions, scene activity and camera location; including people with luggage, projected shadows and lighting changes. Feature sets are evaluated focusing on the recognition rate and the improvement introduced by shadow removal algorithms.

Data consists of blobs extracted from videos of different lengths recorded by ourselves, together with a bitmap crop of its bounding box from the real video in RGB space and in grey levels. We extracted 4659 blobs, which were filtered by size (minimum size 250 pixels) removing up to 49% of the total. We then manually labelled the remaining blobs, corresponding to 1371 images of class *single people*, 367 of class *group of people* and 631 images of class *luggage*.

The criteria we followed to label blobs was:

- Every blob representing a person with or without luggage is labelled as *single person*.
- Blobs representing an object are classified according to the object as *single person*, *group of people*, and *luggage*.
- At least 2/3 of the figure must be inside the bounding box.
- A blob representing more than two people is considered to be a *group of people*, provided that at least 2/3 of two people are visible. No matter how many objects occluded may occur nor which kind of objects are present.

Any other blob was not considered for further processing.

For each image, we kept the result of applying to it the shadow removal algorithms introduced in section 3 and calculated both feature sets discussed in section 4 to the original image and the resulting images, yielding 6 different data sets of 2369 images.

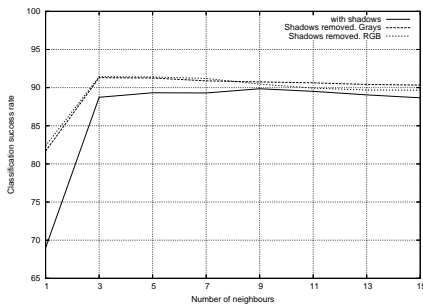


Figure 3: Classification rate of images with and without shadows using geometric features.

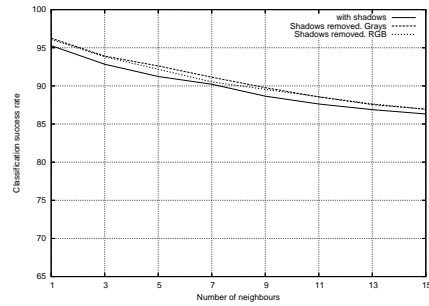


Figure 4: Classification rate of images with, and without shadows, using the matrix of foreground pixels density. Grid size is 4×4 .

We trained a k -nn classifier with different randomly chosen images; the training set was constructed using 80% of the database and the test set was composed of the other 20%. The experiments were repeated 100 times to ensure the statistical independence of the selected samples.

The optimal number of regions to divide each blob into was found by dividing the blobs into different sets of 2×2 , ..., 10×10 regions and classifying objects using a k -nn classifier. A grid of size 4×4 was chosen because it is the smallest with a good classification rate.

In figure 3 and 4 we show the global results of object classification for different values of k and for each feature set. In figure 3, we show the results using geometric features with shadows left and with shadows removed. In figure 4, we show the same results for also show results for foreground pixel density with shadows and with shadows removed, using a grid size of 4×4 . It can be seen for geometric features that for $k = 1$ the classification rate is 68%. For higher values of k , the classification rate rises and stays between 87% and 90%. While foreground pixel density behaves completely differently and shows a good success rate (95% – 92%) for values of k between 1 and 5 and then decreases as k grows.

k	1	3	5	7	9	11	13
P geo.	36.25	48.39	48.39	50.73	51.69	54.51	54.51
L geo.	0.22	0.09	0.02	0.17	0.12	0.08	0.09
P den.	0.48	0.89	11.06	13.79	15.76	16.33	15.93
L den.	0.02	0.02	0.01	0.04	0.14	0.27	0.56

Table 1: Comparison of the confusion rates between groups of people, person (P), or luggage (L) classes. Top rows correspond to geometric features and bottom rows correspond to foreground pixel density.

Figures 5 and 6 show performance of both feature sets for each class. Class *person* and *luggage* have a good classification success rate in both cases;

but *group of people* class shows low values, worse for geometric features than for foreground pixel density. Analyzing confusion between classes, we saw that classes *person* and *group of people* are easily confused, other interclass confusions are low. Both feature sets are valid for classifying *person* and *luggage* class with a good degree of accuracy. In table 1 rows indicate the percentage of confusion of samples of class *group of people* classified as either *person* or *luggage*; top rows correspond to geometric features and bottom rows to foreground pixel density.

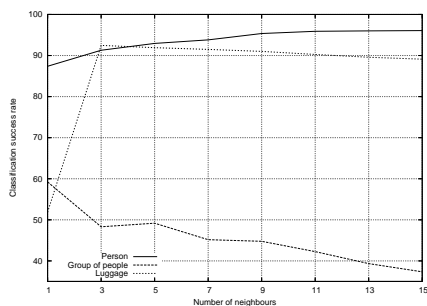


Figure 5: Classification success rates using geometric features, each line correspond to results for one class.

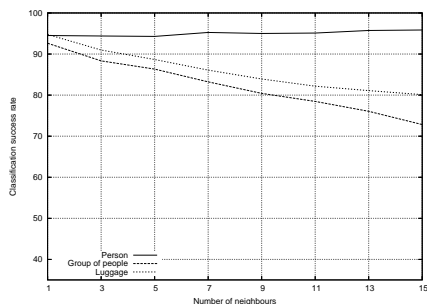


Figure 6: Classification success rates using foreground pixels density features, each line corresponds to results for one class.

Removing shadows increases the classification success rate; although the gain may be low under certain constraints. For instance, for a movie in which the complete tracking loop takes up to 0.91 seconds, and there are an average of 1.8 objects per frame, the grayscale approach takes up to 14.03 seconds on average per frame; and the RGB space approach around 30.10 seconds on average per frame.

6 CONCLUSIONS

We tested two different sets of features: geometric features and foreground pixel density. The for-

mer shows a poorer global performance due to the fact that the *group of people* class is often mis-classified as *person*, although *luggage* and *person* classes are classified satisfactorily. Foreground pixel density shows a better performance with less interclass confusion, although performance decreases when k increases, leading to a poor performance in both *groups of people* and *luggage* classes. Geometric features seem to be more stable than the foreground pixel density. Results show that, independently of the features we use, removing shadows increases the performance of the classifier although the difference may not be worth the effort in cases in which time performance is important.

Currently a head detection algorithm is under development to improve classification of *group of people* class. A method based on people symmetry axis detection is expected to be useful for detecting the number of people represented in a blob. Experiments were carried out off-line, and after these conclusions, we will adapt this classification scheme for on-line operation and apply it to real-time video.

REFERENCES

- Haritaoglu, I., Harwood, D., and Davis, L. S. (2000). W4: Real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 809 – 830.
- Javed, O. (May, 2002). Tracking and object classification for automated surveillance. *The seventh European Conference on Computer Vision (ECCV 2002), Copenhagen*.
- L. Wang, W. H. and Tan, T. (2003). Recent developments in human motion analysis. *Pattern Recognition*, pages 585 – 601.
- Onoguchi, K. (1998). Shadow elimination method for moving object detection. *Proceedings of the 14th International Conference on Pattern Recognition*, page 583.
- Rosin, P. and Ellis, T. (1995). Image difference threshold strategies and shadow detection. *Proceedings of the 6th British Machine Vision Conference. BMVA Press.*, pages 347 – 356.
- W. Hu, T. T., Wang, L., and Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems Man, and Cybernetics-part C: Applications and Reviews*, pages 334 – 351.
- Wren, C. R., Azarbayejani, A., Darrell, T., and Pentland, A. P. (1997). Pfunder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 780 – 785.
- Xu, L., Landabaso, J. L., and Pardàs, M. (2005). Shadow removal with blob-based morphological reconstruction for error correction. *ICASSP 2005. USA*, pages 729 – 732.