

Multi-view access monitoring and singularization in interlocks

Luigi Di Stefano Federico Tombari Stefano Mattoccia
Matteo Balasso

Department of Electronics Computer Science and Systems (DEIS)
Advanced Research Center on Electronic Systems (ARCES)
University of Bologna

{luigi.distefano, federico.tombari, stefano.mattoccia}@unibo.it
matteo.balasso@studio.unibo.it

<http://vision.deis.unibo.it>

Paper ID 50

Abstract

We present a method aimed at monitoring access to interlocks and secured entrance areas, which deploys two views in order to robustly perform intrusion detection and singularization. The main contributions are represented by an original approach to perform background subtraction, which is particularly robust against sudden illumination changes, shadows and photometric distortions, and by the use of a feature extraction and classification approach which allows to reliably determine an estimation of the number of people currently occupying the monitored area. Our system is designed to operate in very small interlocks and can work in a substantially unstructured environment.

1. Introduction

Monitoring access to interlocks and secured entrance areas, such as revolving doors, is a very important task in many security applications. The aim of the task is twofold: first of all, detecting the presence or absence of people in the monitored interlock, secondly allowing only one person at a time to be present in the interlock (*singularization*). Singularization is needed to avoid two (or more) people simultaneously crossing the gate (*piggybacking*) or an unauthorized person crossing the interlock other than the authorized one (*tailgating*). Most solutions deploy sensors such as weight controllers, light barriers and ultrasonic devices [9, 8] which are mounted inside the interlock walls and floor. However, these systems are generally expensive and not accurate enough. Furthermore, they are often un-

practical for maintenance.

Therefore, the use of simple and cost effective video-based approaches is gaining increasing attention in the security industry. These approaches usually rely on background subtraction to detect the presence and estimate the number of people in the interlock. However, it is difficult to perform reliable and accurate background subtraction in real environments, which are characterized by unstructured backgrounds (typically, the floor of the gate), sudden illumination changes, presence of shadows. Commercial systems, such as *Smart Airlock Control System* (SMACS) by Fastcom Technology, deal with this issue by heavily structuring the working environment, i.e. by deploying markers on both the interlock floor (e.g. by means of a chessboard patterned carpet) and walls (e.g. by means of patterned stripes), and by requiring a specially aligned illumination to avoid shadows (i.e. by means of fluorescent tubes disposed in a rectangular arrangement). Another difficulty arises from the size of the interlock, that can often be very small. This constrains the positioning of cameras and generally forces the monitoring system to work with very small field of views, which often do not allow the person to be seen entirely into one single view.

The proposed approach exploits two views to accurately and robustly perform presence detection and singularization in small interlocks. The only modification to the environment required by the system consists in installation of a tiny stripe of reflective material around the borders of the interlock floor. This is not needed for presence detection, but helps improving the feature extraction process required for singularization.

Our approach is based on two main stages, which represent also the two main contributions of this paper. On one

side, the use of a novel background subtraction approach, which deploys a robust matching measure as well as different background models, allows to accurately segment the foreground from the background even in presence of shadows and strong and sudden illumination changes. On the other side, the deployment of distinctive features extracted by comparing the two views allows our system to work with very small field of views. Experimental results concerning the implementation of the proposed approach on a revolving door demonstrate the effectiveness of our proposal.

2. System setup and preliminary work

Since the specifications and the design of our method are closely related to its practical application, we will refer throughout the paper to a specific model of interlock, which was used for the study and performance evaluation of our approach. Nonetheless, the developed methodology can be easily generalized to different kind of interlocks. The considered interlock is a two-door cylindrical revolving door, typically used for security accesses such as bank entrance areas. As shown in Fig. 1), the entrance and exit doors are placed at a right angle and two cameras are fixed to the ceiling of the revolving door.

The gate setup is particularly challenging for video-based monitoring. The small height of the interlock (approximately 2 meters) typically enables each camera to get only a partial view of the gate occupant. Moreover, the interlock floor is characterized by a black anti-slippery material. The presence of small knobs in this material determines a natural light pattern that, unfortunately, changes notably its appearance under different illumination conditions, as depicted in Fig. 2. Besides, the presence of bullet-proof glass all around the interlock typically creates serious light artifacts on the floor and walls, due to the presence of different illumination conditions inside and outside the building. The presence of illuminators inside the interlock and light sources outside generally produces strong and unpredictable shadows when people occupy the interlock. Illumination changes and movements of people and vehicles outside the interlock also account for additional photometric distortions. Finally, further artifacts comes up in time due to dirtying of the walking surface. Some examples showing different appearances of the gate floor when the revolving door is empty are shown in Fig. 3.

Preliminary work, just sketched here for the sake of brevity, was carried out to simulate a real system and verify the feasibility of a possible solution based on a feature extraction and classification approach. To reach this aim, a 3D graphic model of the interlock and the occupants was developed (see Fig. 4). Then, a study concerning the selection of the features to be used by the system was carried out by rendering the two camera views under many working conditions. In particular, the simulations considered the number

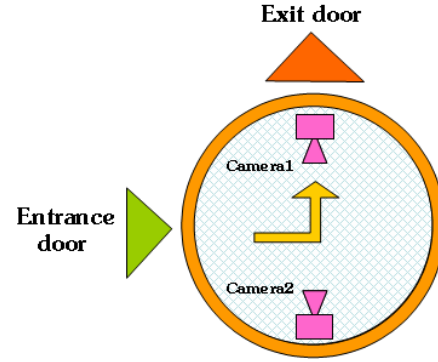


Figure 1. Outlook of the interlock used to test the approach

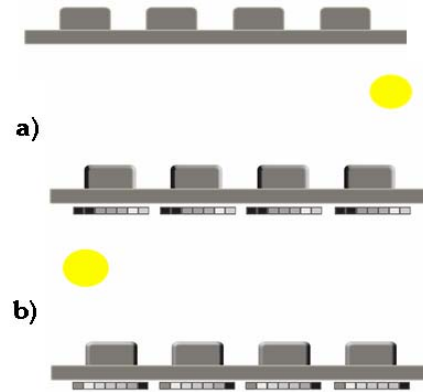


Figure 2. Effect of different illumination conditions (a,b) on the appearance of the gate floor

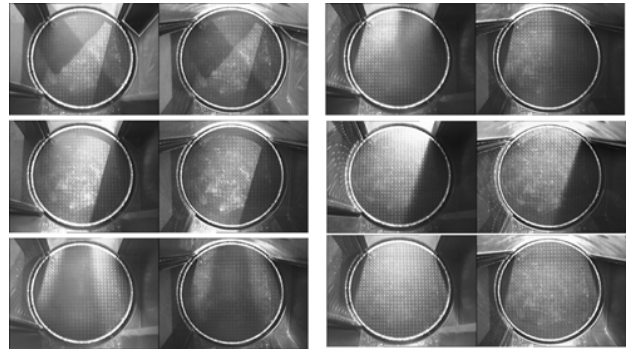


Figure 3. Typical photometric distortions and artifacts affecting the appearance of the gate floor.

of occupants as well as their position, orientation, height and size. Promising results of these simulations pushed forward for a thorough study addressing the real working conditions.

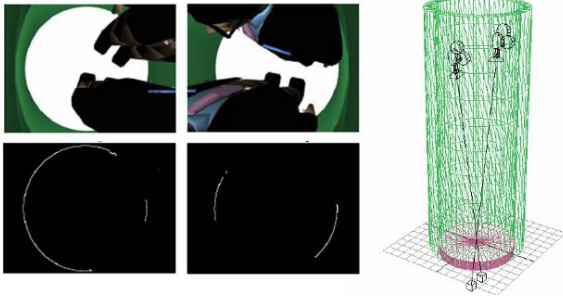


Figure 4. Simulation of the working conditions

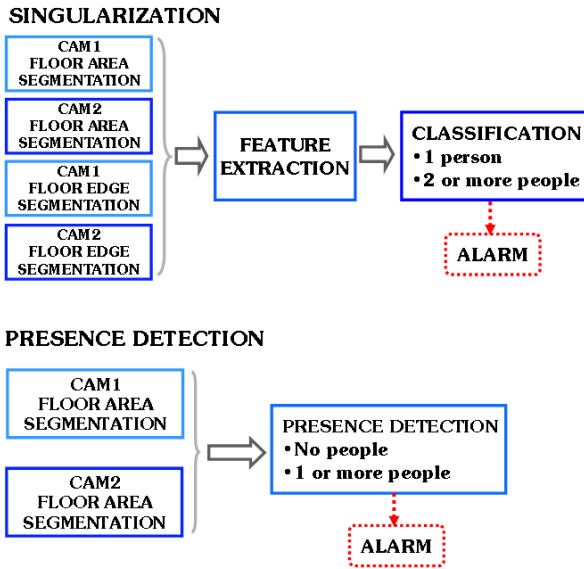


Figure 5. Flow diagrams of singularization and presence detection algorithms

Fig. 5 outlines the proposed approach, which is aimed at both presence detection and singularization. As it can be seen, both processes exploit a common stage concerning segmentation of the floor area visible by both cameras. This stage will be described in Section 3. Then, to carry out presence detection the output of the floor segmentation stage is analysed to decide whether the interlock is currently empty or one or more occupants are present. Differently, singularization exploits also the segmentation of the floor border visible by cameras, so as to extract additional features that allow to discriminate whether the number of occupants is higher or equal than 1. The use of two cameras is mainly motivated by the fact that features are computed based on the output of the segmentation processes associated with the two views. This will be described in Section 5.

3. Segmentation processes

3.1. Segmentation of the visible floor area

A background subtraction approach is deployed in order to segment in each frame of both views the visible area of the interlock floor. Yet, we have to deal with all the disturbance factors briefly described in previous section: presence of sudden local illumination changes and shadows, mutable aspect of the background depending on the position of the light sources. Under these circumstances, most background subtraction algorithm relying on image intensities are prone to fail. Hence, we deploy a recently proposed Matching Function [10], hereinafter referred to as MF, that has showed good robustness with regards to notable photometric distortions. This measure is based on the assumption that distortions between two homologous image patches do not violate the intensity order between neighbouring pixel, which is a rather common assumption when one wants to deal with strong photometric distortions (e.g. [12, 1, 7, 5, 11]). Let B_{c1} and F_{c1} be respectively the background model and the current frame for camera 1. For each pair of points at coordinates (x, y) in B_{c1} and F_{c1} , MF computes a score measuring how well the order is maintained between neighbouring pixel by considering a square patch of side r centered on $B_{c1}(x, y)$ and $F_{c1}(x, y)$. In particular, for each point of the patch MF cross-correlates the horizontal and vertical central differences and then normalizes the correlation term:

$$MF_{B_{c1}, F_{c1}}(x, y) = \frac{\mathbf{B}_{c1}^h(\mathbf{x}, \mathbf{y}) \circ \mathbf{F}_{c1}^h(\mathbf{x}, \mathbf{y}) + \mathbf{B}_{c1}^v(\mathbf{x}, \mathbf{y}) \circ \mathbf{F}_{c1}^v(\mathbf{x}, \mathbf{y})}{N_{F_{c1}}(x, y) \cdot N_{B_{c1}}(x, y)} \quad (1)$$

where \circ denotes the dot product, $\mathbf{F}_{c1}^h(\mathbf{x}, \mathbf{y})$, $\mathbf{F}_{c1}^v(\mathbf{x}, \mathbf{y})$ represent two vectors containing respectively the horizontal and vertical central differences of $F_{c1}(x, y)$ and, analogously, $\mathbf{B}_{c1}^h(\mathbf{x}, \mathbf{y})$, $\mathbf{B}_{c1}^v(\mathbf{x}, \mathbf{y})$ for $B_{c1}(x, y)$. Finally, the two terms at the denominator are given by:

$$N_{B_{c1}}(x, y) = \sqrt{\|\mathbf{B}_{c1}^h(\mathbf{x}, \mathbf{y})\|_2^2 + \|\mathbf{B}_{c1}^v(\mathbf{x}, \mathbf{y})\|_2^2} \quad (2)$$

$$N_{F_{c1}}(x, y) = \sqrt{\|\mathbf{F}_{c1}^h(\mathbf{x}, \mathbf{y})\|_2^2 + \|\mathbf{F}_{c1}^v(\mathbf{x}, \mathbf{y})\|_2^2} \quad (3)$$

It is worth noting that, due to the normalization step, the score provided by MF ranges between -1 (complete dissimilarity) and 1 (complete similarity). Besides, the patch side, r , is a parameter of this method: the higher r , the more robust the measure, the less accurate the segmentation along foreground borders. In our experiments r was tuned to 11. From a computational point of view, MF can be calculated very efficiently by means of incremental techniques [6, 3] such as shown in [4]. To perform background subtraction on the current frame, $MF_{B_{c1}, F_{c1}}(x, y)$ is computed at each point (x, y) that may belong to the interlock floor, then a

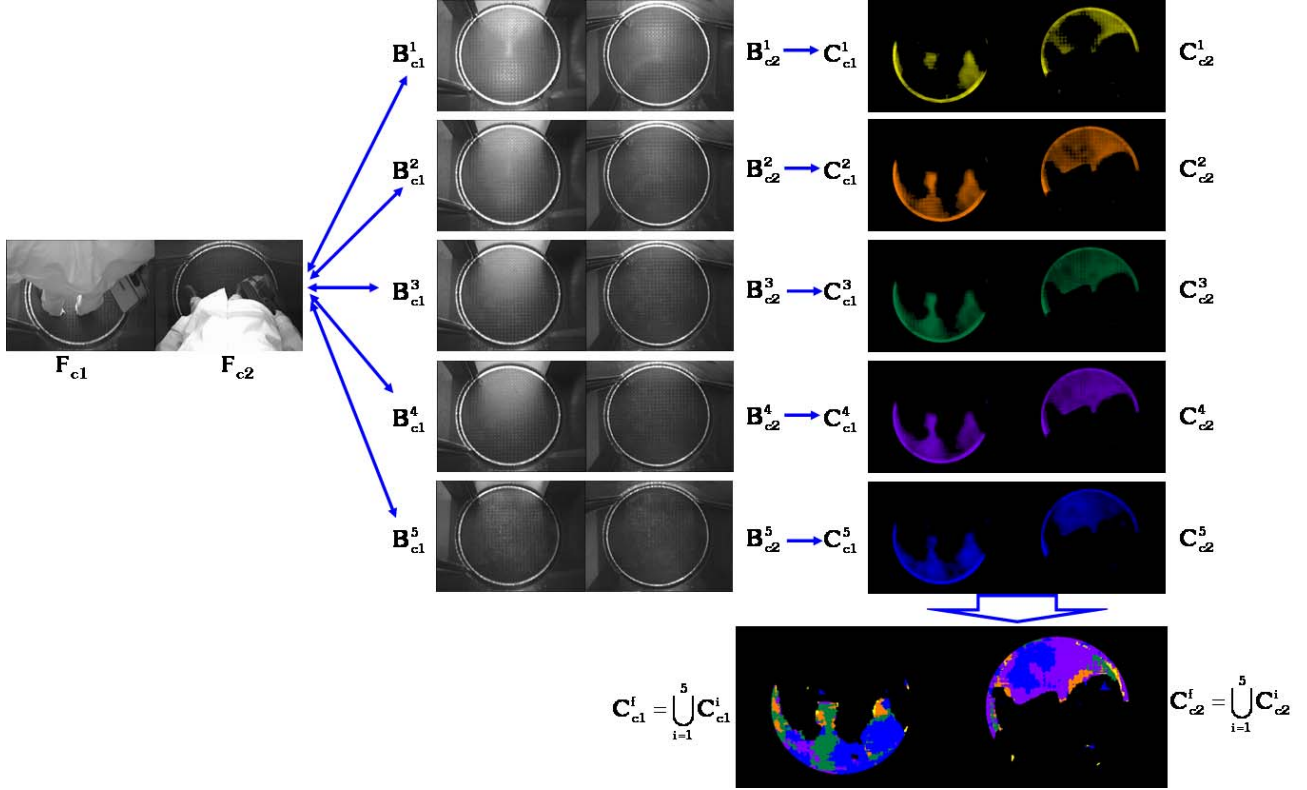


Figure 6. Segmentation of the floor area by means of different background models.

threshold is used to discriminate between floor points (high score) and non-floor points (low score). A pre-computed binary mask is deployed to eliminate the points of the images lying outside the floor area.

Preliminary results dealing with the use of MF showed that this was able to determine a rough estimation of the visible floor area and was robust to strong photometric distortions and artifacts. Anyway, due to the mutable aspect of the floor (as shown in Fig. 2), accuracy along the segmented borders was sometimes not precise enough. Hence, we devised an improvement to the basic approach. In particular, several background models are computed rather than just one, each one encoding the appearance of the gate floor under different static illumination conditions. To this purpose, 5 different background models for each view are built, e.g. for camera 1 $B_{c1}^1, \dots, B_{c1}^5$, with illumination conditions varying from very dark (B_{c1}^1) to very bright (B_{c1}^5). Each background model is obtained by varying the rheostat of the illuminator present inside the interlock. Under each illumination condition the actual background model is attained by averaging the values assumed by each point along an initialization sequence of 30 frames, during which the interlock is empty. This can be regarded as an initialization stage of our system. Then, at run-time, for each background

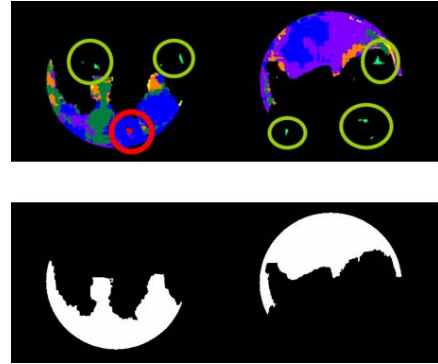


Figure 7. Morphological post-processing of the final floor mask.

model B_{c1}^i , a corresponding floor mask C_{c1}^i is computed by thresholding the MF score between F_{c1} and B_{c1}^i :

$$C_{c1}^i(x, y) = TH \left(MF_{B_{c1}^i, F_{c1}}(x, y) \right) \quad (4)$$

Throughout all of our experiments, the threshold was set to 0.215. The outcome of this operation on each background model votes for the similarity between the current frame and the background. To classify a point as floor, the portion of the current frame must match at least one background

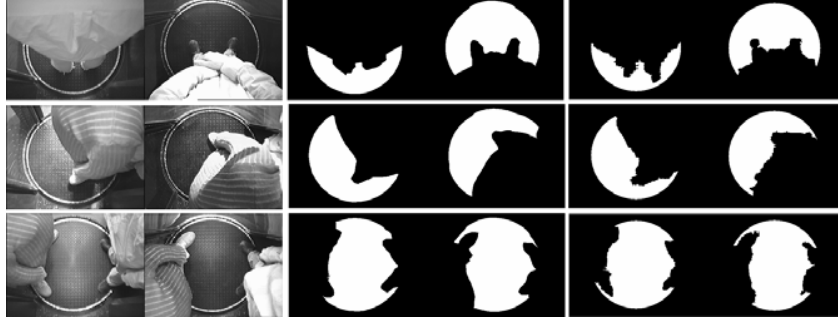


Figure 8. Qualitative comparison between ground-truth (centre) and output of the floor area segmentation algorithm (right).

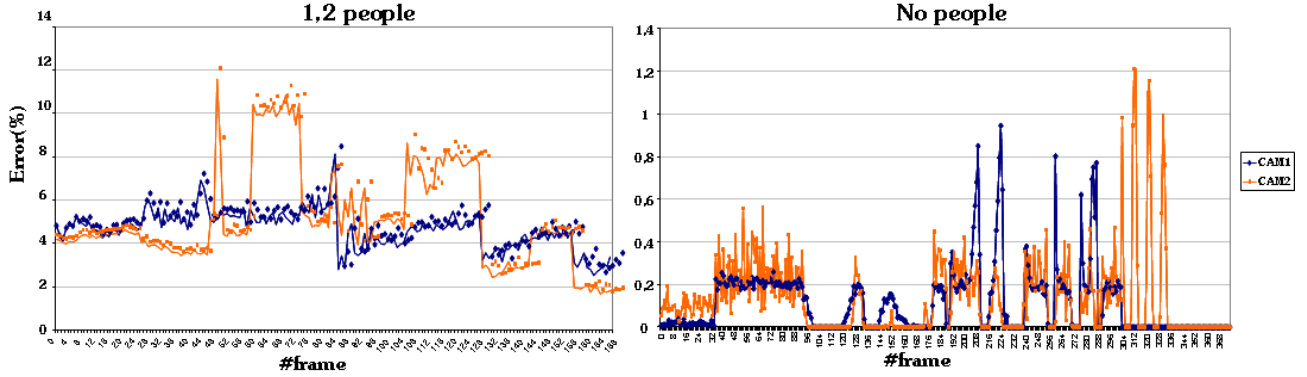


Figure 9. Quantitative assessment of the errors yielded by the floor area segmentation algorithm.

model, i.e. the matching score for at least one background model must be above threshold. Hence, the final floor mask is obtained as the union between each floor mask:

$$C_{c1}^f(x, y) = \cup_{i=1}^5 C_{c1}^i \quad (5)$$

Given the 5 background models $B_{c2}^1, \dots, B_{c2}^5$ and the current frame F_{c2} , the same processing stages provide the final floor mask for the second view, C_{c2}^f . An example of how the approach works can be seen in Fig. 6. Here, the two current frames F_{c1}, F_{c2} (left column) are matched with each of the corresponding 5 background models (central column), yielding the floor masks shown in the right column. In the final floor masks, C_{c1}^f, C_{c2}^f , the contributions of the 5 floor masks are depicted with different colours, so as to point out how the deployment of different background models allows to increase the accuracy of this process.

A further step based on area-closing and area-opening morphology is applied to improve the output of the background subtraction stage by reducing false positives and false negatives from the final floor mask. An example is shown in Fig. 7: red and green circles (above) indicate the connected components filtered out from the floor masks, while the final binary output for both views is shown below.

To conclude this sub-section we show some experimen-

tal results regarding the floor area segmentation process. Footage has been acquired concerning scenes with different numbers of people inside the gate (0, 1, and 2) and notably varied illumination conditions obtained by the use of light sources placed outside the revolving door (as in Fig. 3). In particular, the reported segmentation results refer to 546 frames taken from 18 real sequences for each view. Of these frames, for those concerning 1 or 2 people (170 frames taken from 12 sequences) the ground-truth was obtained by hand-segmentation of the visible floor area and border. Fig. 8 shows samples of frames with 1 and 2 people occupying the interlock and allows for qualitative comparison between the ground-truth (centre) and the output obtained by the proposed floor segmentation approach. Besides, Fig. 9 reports the segmentation error (sum of false positives and false negatives normalized by number of points within the floor area) along the footage. These results demonstrate that the proposed approach provide significant accuracy, with a very small mean error around 5% when the interlock is occupied, and a negligible error when the interlock is empty.

3.2. Segmentation of the visible floor border

In order to estimate the percentage of visible floor border, as illustrated in Fig. 10, an approach similar to the floor

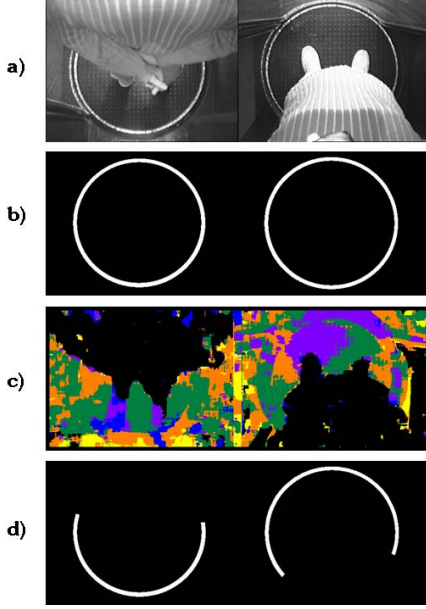


Figure 10. Floor border segmentation: a) input frames, b) binary mask, c) background subtraction results, d) final segmentation

area segmentation algorithm is adopted. A binary mask is used to filter out the output of the background subtraction algorithm for those pixels not belonging to the border area (Fig. 10, b)). Even though the algorithm worked rather well by exploiting the natural edge of the floor border, we tested extensively our approach by applying a reflective tape on the floor border. In fact, this solution is minimally invasive and allows for a significant increase in the robustness of the border segmentation stage under strongly varying illumination conditions. This reflective tape is the only synthetic fiducial used in the whole system.

For what regards quantitative results, Fig. 11 reports the segmentation error computed along the footage by comparing the hand-segmented ground-truth with the output of the proposed floor border segmentation algorithm. When the interlock is occupied (1 or 2 people) the segmentation error is typically below 10%, while in case the interlock is empty the error is always negligible (never higher than 0.3%).

4. Presence detection

The algorithm enabling presence detection aims at raising an alarm when one or more occupants are present within the monitored interlock. This allows, for instance, to avoid performing singularization when the gate is empty, or to alert security personnel when the gate is occupied. The idea is to deploy the segmentation of the visible floor of the gate from the two views to estimate whether the interlock is empty or not. In particular, denoting as A_{c1} and A_{c2} the estimated percentage of visible floor area respectively in view

1 and 2, the presence detection rule is defined as follows:

$$presence = \begin{cases} true & (A_{c1} < Th_p) \text{ or } (A_{c2} < Th_p) \\ false & otherwise \end{cases} \quad (6)$$

In other words, presence is detected when the floor is not almost completely visible from at least one of the two cameras. Typically, Th_p is set to 80%. Obviously, the reliability of the presence detection algorithm heavily depends on the outcome of the floor area segmentation stage described in Sec. 3. Some experimental results are shown Fig. 12, where the estimated percentage of visible floor area in the two views is plotted along the different frames of a video sequence. In this sequence, a person enters the revolving door, then the entrance door is automatically closed. The most important events are chronologically marked along the frames, that is: the person's first foot is visible, his second foot is visible, the person is completely inside the interlock, the entrance door starts to close, the entrance door is completely closed. In this sequence, the ground-truth is available only for certain frames and is indicated by blue dots. As it can be seen by comparison with the available ground-truth, the accuracy of the segmented floor area is very high along the frames of the sequence and for both views, despite the notable photometric distortions, shadows and artifacts appearing during the entrance of the person and the entrance door closing time. Furthermore, the use of the presence detection rule in (6) allows to reliably detect that the gate is not empty just after the first foot of the person is placed into the interlock.

5. Singularization by means of feature extraction and classification

In order to perform fast and reliable singularization we propose an approach based on feature extraction and classification. The key point for such an approach is to rely on a small number of distinctive features: the small number allows for a reduced computational cost, while the distinctiveness is necessary for reliability of singularization.

A thorough study concerning the possible features to be adopted in the singularization stage lead us to the determination of two simple features to be computed once the two views have been registered. The first feature, referred to as *Area OR*, is given by the percentage of the floor area which is visible in at least one of the views. This is motivated by the fact that the presence of a single person which stands around the center of the interlock allows a large portion of floor area being seen by at least one of the two cameras (ideally, the whole area except for the portion of the person touching the floor, i.e. his feet). Conversely, the presence of two or more people would force a notable part of the floor to be not visible in both views. Similarly, the second feature is given by the percentage of the floor border which is visible

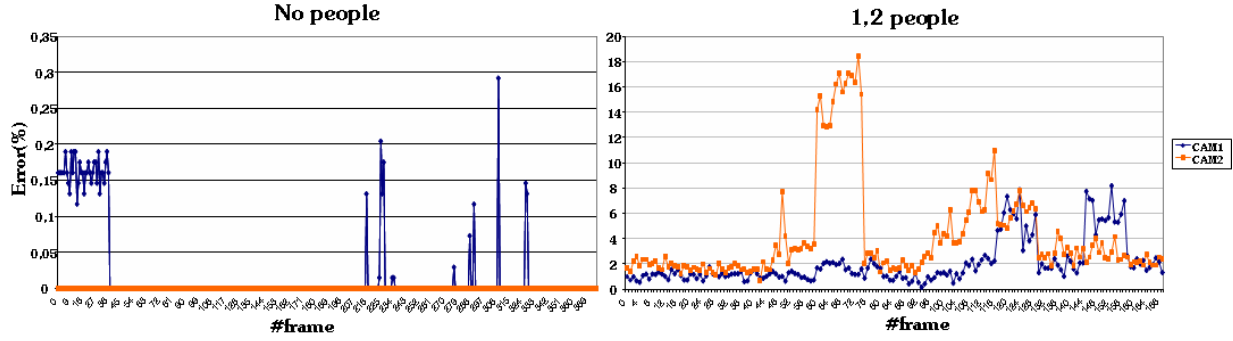


Figure 11. Error concerning the floor border segmentation stage in the two views compared to ground-truth.

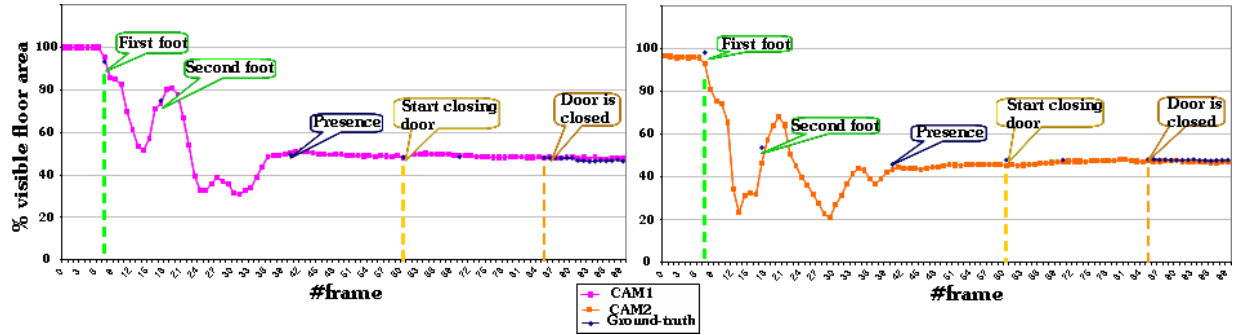


Figure 12. Experimental results comparing the percentage of visible floor area in the two views with ground-truth along a test sequence.

in at least one of the views (referred to as *Edge OR*). This is also motivated by the fact that the presence of a single person standing around the center of the interlock generally allows all floor border points to be seen in at least one of the two views, which is not the case when two or more people are present. An example of how the Area OR and Edge OR features appear in two cases (one person and two people) is shown in Fig. 13.

Then, the output of the feature extraction step is provided to a trained classifier which discriminates between the two cases: "one person" or "two or more people". The two features can be rapidly computed since the homographies between the two planar views with respect to the floor are fixed and precomputed at startup. During operating mode, singularization is applied only when the two doors of the interlock are closed. This reduces false alarms due to the single user's movements while entering the revolving door. If singularization detects piggybacking, then an alarm is raised, the exit door stays closed and the entrance door is opened once again to allow the occupants to leave the interlock.

Experimental results are shown here by means of iterated 2-fold cross-validation. In particular, the frames of the available footage sequences concerning the presence of 1 or 2 people (overall 170 frames taken from 12 real sequences) were randomly subdivided into two groups, A and B. As

#	% E_A	% M_A	% F_A	% E_B	% M_B	% F_B
1	2	0	2	0	0	0
2	0	0	0	0	0	0
3	0	0	0	2	0	2
4	2	0	2	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	0	3	0	3
8	0	0	0	3	0	3
9	0	0	0	0	0	0
10	0	0	0	0	0	0

Table 1. Cross-validation of SVM-based singularization: percentages of error, missed detection, false alarm yielded for different subdivision of data into two subsets (A and B).

shown in Tab. 1, this subdivision has been done 10 times by means of different shuffles. For each case, two SVM classifiers [2] are trained on the two sets, then each set is used as testing sequence for the other. An example of two classifiers obtained with this kind of evaluation is depicted in Fig. 14, which shows that training has been done toward generalization and to avoid data over-fitting. Tab. 1 shows the percentages of, respectively, total errors (E), missed detection (M , i.e. 'two or more people' being recognized as

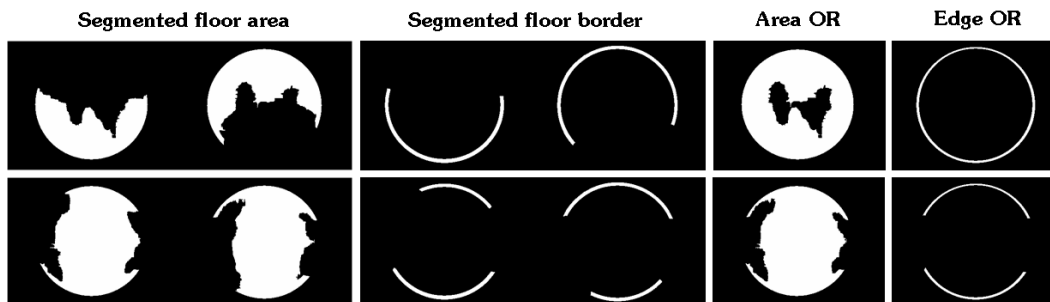


Figure 13. Extracted features "Area OR" and "Edge OR" in case of one person (above) and two people (below).

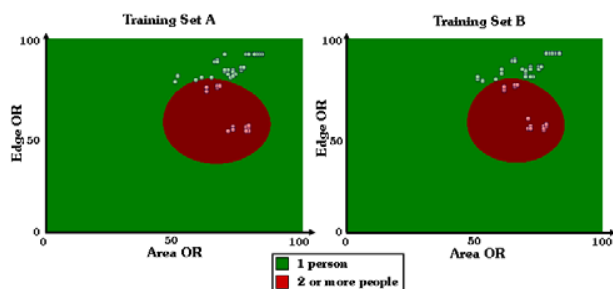


Figure 14. An example of SVM classification of two sets A, B.

'one person') and false alarms (F , i.e. 'one person' being recognized as 'two or more people') for both sets A and B along the various subdivisions. It can be noted that the total error percentages are extremely low in all cases, being often 0% and anyway never higher than 3%. This demonstrates that the two classes are distinguishable by means of the chosen features. Furthermore, in all cases the errors are represented by false alarms only, with no missed detection being actually yielded by our system: this is very important for practical applications, since the most important aspect in anti-piggybacking and anti-tailgating is to avoid misclassifying two or more people as one, while a few false alarms can be usually easily dealt with.

6. Conclusions

An effective method to perform presence monitoring and singularization in interlocks has been proposed. The method relies on an original background subtraction stage, which has proved to be robust and accurate in segmenting the visible area and border of the interlock floor. Experimental results showed significant capabilities of our system to perform fast and reliable presence detection in the considered interlock, as well as singularization by means of a feature extraction and classification approach. It is worth pointing out that the proposed method including all processing stages is very efficient: in our implementation, which

does not contemplate any hardware or SIMD optimization, response time for each new frame is 140 ms (i.e. ≈ 7 fps) on a 2.14 GHz Intel Core Duo.

7. Acknowledgments

This research has been supported by Plexa (www.plexa.com). We wish to thank them for the grant to publish the results presented in this article.

References

- [1] D. Bhat and S. Nayar. Ordinal measures for image correspondence. *IEEE Trans. Pattern Recognition and Machine Intelligence*, 20(4):415–423, April 1998. 3
- [2] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. 7
- [3] F. Crow. Summed-area tables for texture mapping. *Computer Graphics*, 18(3):207–212, 1984. 3
- [4] L. Di Stefano, F. Tombari, S. Mattoccia, and E. De Lisi. Robust and accurate change detection under sudden illumination variations. In *Proc. ACCV'07 Workshop on Multi-dimensional and Multi-view Image Processing*, 2007. 3
- [5] A. Lanza and L. Di Stefano. Detecting changes in grey level sequences by ml isotonic regression. In *Proc. AVSS*, 2006. 3
- [6] M. Mc Donnell. Box-filtering techniques. *Computer Graphics and Image Processing*, 17:65–70, 1981. 3
- [7] A. Mittal and V. Ramesh. An intensity-augmented ordinal measure for visual correspondence. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 849–856, 2006. 3
- [8] M. Schwarz. Revolving door with metal detection security. US Patent 946867, June 2004. 1
- [9] M. Schwarz and R. Mayer. Anti-piggybacking: sensor system for security door to detect two individuals in one compartment. US Patent 5201906, April 1993. 1
- [10] F. Tombari, L. Di Stefano, and S. Mattoccia. A robust measure for visual correspondence. In *Proc. ICIAP*, 2007. 3
- [11] B. Xie, V. Ramesh, and T. Boult. Sudden illumination change detection using order consistency. *Image and Vision Computing*, 2(2):117–125, 2004. 3
- [12] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proc. European Conf. Computer Vision*, pages 151–158, 1994. 3