

# A Genetically Optimized Graph-based People extraction method for embedded transportation systems real conditions

Christophe Coniglio, Cyril Meurie, Olivier Lézoray, and Marion Berbineau

**Abstract**—In this paper, we present a new method for people extraction in complex transport environments. Many background subtraction methods exist in the literature but don't give satisfactory results on complex images acquired in moving trains that include several locks such as fast brightness changes, noise, shadow, scrolling background, etc. To tackle this problem, a new method for people extraction in images is proposed. It is based on an image superpixel segmentation coupled with graph cuts binary clustering, initialized by a state-of-the-art foreground detection method. The proposed strategy is composed of four blocks. A pre-processing block that uses filters and colorimetric invariants to limit the presence of artifacts in images. A foreground detection block that enables to locate moving people in images. A post-treatment block that removes shadow regions of no-interest. A people extraction block that segments the image into SLIC superpixels and performs a graph cut binary clustering to precisely extract people. Tests are realized on a real database of the BOSS European project and are evaluated with the standard F-measure criteria. Since many state-of-the-art methods can be considered in our three first blocks along with many associated parameters, a genetic algorithm is used to automatically find the best methods and parameters of the proposed approach.

## I. INTRODUCTION

Nowadays, video surveillance systems are widespread in transport applications. They can be used to detect potentially dangerous situations and to track moving objects or people in order to enhance the comfort and security of the infrastructure, as well as that of people. Many authors are therefore concerned with the development of efficient methods for people detection, re-identification, tracking, counting, and action recognition. The common feature of these researches is the detection of moving objects. That is why many algorithms of background subtraction (extracting moving objects by comparison between a current frame with a learned reference background) can be found in the state-of-the-art (see [1] for a review). The literature is large but the existing methods that generally offer very good results in simple controlled environments, have much poorer outcomes in the case of complex transport environments (e.g., embedded systems). In our context of optimizing the management of passenger flows in a multimodal context, people extraction is useful to follow the trajectory of people to re-identify (anonymously) persons in a cameras' networks.

Christophe Coniglio, Cyril Meurie, and Marion Berbineau are with Univ Lille Nord de France, F-59000 Lille, IFSTTAR, LEOST, F59650, Villeneuve d'Ascq, France {christophe.coniglio, cyril.meurie, marion.berbineau}@ifsttar.fr

Olivier Lézoray is with Normandie Univ., UNICAEN, ENSICAEN, GREYC UMR CNRS 6072, Caen, France, olivier.lezoray@unicaen.fr

People extraction is crucial and this step has to be as fine as possible even if the scene contains many scientific locks (such as fast brightness changes, noise, shadow, scrolling background, etc.). Indeed, these latter problems often occur when the images' acquisitions are made in real conditions (e.g., in moving trains). Within this framework, we propose in this paper a robust method for people extraction relying on an image superpixel segmentation coupled with graph cuts binary clustering, initialized by a state-of-the-art foreground detection method.

## II. PROPOSED APPROACH

In most cases, state-of-the-art foreground/background segmentation methods [1] (such as Fuzzy Sugeno Integral, Fuzzy Gaussian, Gaussian Mixture Model, Multi-Layer BGS, Adaptative Som, VuMeter) extract moving objects with good accuracy. However, their performance drops down when they are used in complex environments (e.g., in embedded environments with the scientific locks we previously mentioned). For such difficult real cases, new robust people extraction methods are needed. To that aim we propose a graph-based robust people extraction method that goes beyond the actual state-of-the-art and we devise a general strategy to solve that latter problem. This complete strategy is illustrated in Figure 1 and is composed of four blocks detailed below. Each block can be used as an entry for another block.

- Image pre-processing. The pre-processing block consists in using filters and colorimetric invariants in order to reduce the effects of fast brightness changes and noise in the image. Since different pre-processing can be needed for each of the next blocks, we have chosen to consider many different filters (and associated filters' settings), and different colorimetric invariants. It might be possible that no pre-processing is needed to obtain the best performances, and we also consider that possibility.
- Foreground detection. This block consists in retrieving the first useful information to detect moving people in a complex environment: foreground detection. To that aim, we have considered many state-of-the-art algorithms of the literature.
- Post-treatment. This block is used to remove shadows classified as foreground. Again, we have considered several state-of-the-art algorithms to do so. Finally, a mathematical morphology step is used in order to delete small regions of no-interest.
- People extraction. In this block, a bounding box is automatically delineated around people using the infor-

mation provided by background extraction. The image is segmented with a superpixel method and converted into a superpixel region adjacency graph. A binary clustering is performed with graph cuts on that graph in order to extract people.

### A. Genetic Optimization

For each of the considered block, many different choices are possible. For instance, for foreground detection, up to twenty state-of-the-art algorithms can be considered. In addition, for pre-processing, each denoising filter has several parameters, etc. With this in mind, it is now easy to understand that our proposed method has too many parameters to have them tuned by hand. This is especially true since we consider real and complex images, and a given set of parameters that performs well in a given situation, will not necessarily be efficient in other situations. The search of the values of those parameters is called model selection in machine learning. This problem is very difficult to solve since the set  $\theta$  of parameters to be tuned is very large and it is very hard to determine the set  $\theta^*$  that optimizes a given quality criterion. This problem not being tractable, we have chosen to consider a meta-heuristic with the use of genetic algorithms. This genetic algorithm will be used to determine not only the best parameters of the method involved in the proposed approach, but also to determine the best state-of-the-art approach when several of them are considered. Details on the genetic optimization will be given in the experiments sections. In the sequel we detail each block of our approach. For the first three block, we present all the different state-of-the-art algorithms we considered and for the last block, we present our superpixel graph-based people extraction.

### B. Pre-processing

The pre-processing step consists in using filters and colorimetric invariants in order to reduce the effect of fast brightness changes and noise in the image acquired in transport environment and mobility. The best colorimetric invariants and filters as well as their parameters will be automatically determined by the genetic algorithm among six invariants [2] (greyworld, reduced coordinates,  $l_1l_2l_3$ ,  $m_1m_2m_3$ , affine normalization, and RGB rank) and four filters (blur, gaussian blur, median blur, and bilateral). Since different pre-processings can be used in the different blocks, these pre-processings are optimized independently.

### C. Foreground detection

In foreground detection, the state of art is very large [1], but no method gives very good results (with a short training) in terms of people extraction on our database. Nevertheless, we have used this detection to approximately determine the position of people in images. To that aim, we have considered several state-of-the-art approaches such as a Gaussian mixture model method [3], a fuzzy based method [4] [5] [6], a statistical methods using both color and texture features [7], a neural networks method [8], and a non parametric method [9]. In this paper, we have used the BGS

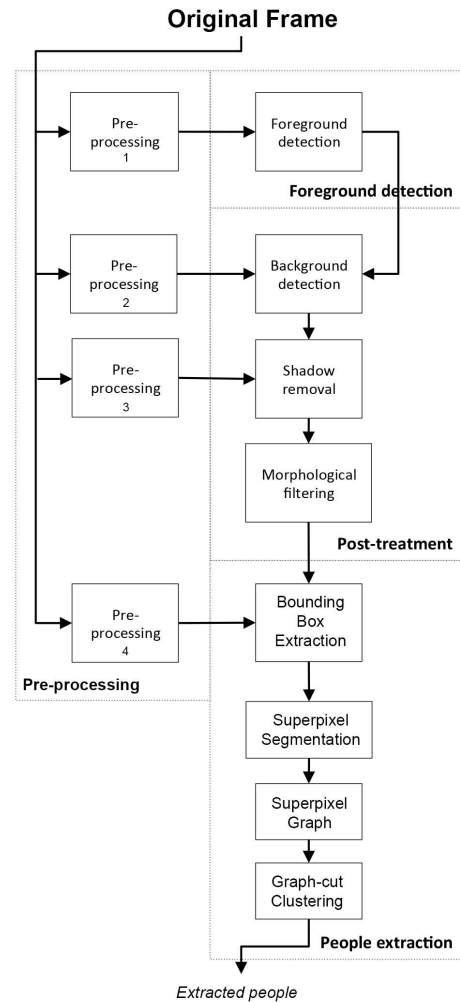


Fig. 1. Synopsis of the proposed method.

library implemented by [10] that provides a unified library for all these methods. The choice of the best foreground detection method will be performed by the genetic algorithm.

### D. Post-treatment

As previously explained, people extraction is a difficult task when it occurs in real transport environments. The pre-processing step is used to limit the brightness changes and noise in the acquired images. However, this does not take into account shadows and other artifacts (errors in foreground detection, or small regions of non interest caused by the scrolling background). That is why, we propose a robust post-treatment composed of two steps: shadow removal and removal of small regions with morphological filtering.

1) *shadow removal*: The presence of shadows is a well-known problem in the literature dedicated to foreground detection. Indeed, these methods often detect shadows as foreground. The impact of shadows can be crucial for the foreground detection precision. This can cause moving objects to merge and distorts their size and shape. This can be explained by the fact that shadows share the same movement patterns and have similar magnitude of intensity changes as

that of the foreground object of interest. In many cases, these errors can much decrease the performance of people detection. In our method, this can strongly affect the outcomes of the graph cut binary clustering for people extraction since it is initialized with the estimated foreground detection. So, confusion between pixels belonging to background shadow pixels and foreground pixels has to be reduced to ensure a good extraction of people. References [11], [12] provide a good survey of shadow detection methods. Regarding their conclusions, we have considered the most prominent methods in each category they exhibited: a chromaticity based method [13], a physical method [14], a geometry based method [15], and a texture based method [16], [17]. Many authors seem to agree that these methods provide better results when the background is known. In this paper, before using the shadow removal methods [11], we propose to optimize the pre-processing and dissociate the algorithm used to detect the background and the foreground. Again, the choice of the best shadow removal method will be performed by the genetic algorithm.

2) *Morphological filtering*: Once background detection and shadow removal are performed, some small regions of no-interest can remain. They can be easily deleted by a few operations of mathematical morphology. This consists in applying series of erosions and dilatations. As before, the parameters of this morphological filtering are genetically tuned.

### E. People extraction

At the end of the post-treatment block, we obtain a rough foreground detection (in some cases some people are not even completely extracted). Therefore, we propose to refine this first extraction through the people extraction block illustrated in Figure 1. This block is composed of two steps. In a first step, a superpixel graph is build, based on a delimitation of a bounding-box followed by a superpixel segmentation and a conversion of the segmentation into a region adjacency graph. The second step consists in extracting precisely the people with a graph cut binary clustering.

1) *Superpixel Graph*: To reduce the processing time of the graph cut clustering algorithm as well as increasing the smoothness of the final graph clustering in the people extraction, we delimitate a bounding-box (a rectangle wherein the foreground information can be inserted) that is used as a mask image  $I_{MASK}$  in the segmentation step in order to limit the number of pixels to be considered. Then, the information contained in the bounding-box is segmented with a SLIC superpixel method in order to obtain several homogeneous regions without losing the edge information [18]. Unlike others steps described in this paper, no optimization is performed on the superpixel segmentation. Finally, we convert the segmented image into a region adjacency graph in the form of the pair  $G = (V, E)$  where  $V$  is the set of nodes (regions of the segmented image, that we represent with a color histogram) and  $E$  the set of undirected edges (connections between adjacent regions). Figure 2 shows the results of these two steps one image of the database.

2) *Graph Cut clustering*: Graph cuts [19] are a powerful segmentation algorithm that enables binary clustering of a graph representing, for instance, an image as a grid graph of vertices (corresponding to pixels). It consists in formulating the clustering problem as an energy minimization in the form of a labeling problem. In this paper, we have used the min-cut/max-flow implementation of [20]. As a result, the superpixel graph  $G = (V, E)$  of the segmented image is classified into two classes starting with an initialization of vertices in sources (i.e., foreground) and sinks (i.e., background),  $l = \{\text{source, sink}\}$  given the result of background extraction (i.e., each node  $R_i \in V$  is assigned with a binary label  $l_i$ ). To that aim, we assign a capacity to each node of  $V$  for these two classes and a similarity for each edge of  $E$ . So, the minimum of the energy  $\hat{i}$  shown below corresponds to the best segmentation among the set  $F$  of all possible labeling solutions:

$$\hat{i} = \operatorname{argmin}_{l \in F} \left( \sum_{R_i \in V} W^{l_i}(R_i) + t \sum_{R_i \in V} \sum_{R_j \in N_{R_i}} S(R_i, R_j) \cdot \delta_{l_i \neq l_j} \right)$$

where  $S(R_i, R_j)$  is the similarity between two regions,  $W^{l_i}(R_i)$  is the capacity of a node, the term  $\delta_{l_i \neq l_j}$  in the second sum is the Potts prior that encourages piecewise-constant labeling, and  $N_{R_i}$  is the set of regions adjacent to the region  $R_i$ .

Let us now explain how we define these similarities and capacities. The similarity  $S(R_i, R_j)$  between two regions  $R_i$  and  $R_j$  is given by:

$$S(R_i, R_j) = \exp \left( -\frac{d(R_i, R_j)}{2\theta^2} \right) \cdot \frac{1}{\operatorname{dist}(R_i, R_j)}$$

where  $\operatorname{dist}(R_i, R_j)$  is the Euclidean distance between the barycenters of regions  $R_i$  and  $R_j$ , and  $d(R_i, R_j)$  is a color distance between the two regions color histograms. For this latter, we have tested several metrics:  $L_1$ ,  $L_2$ , and Bhattacharyya distances. The coefficient  $\theta$  is a bandwidth similarity parameter that will be fixed by the genetical algorithm. Finally we define  $K = \max S(R_i, R_j)$ . To compute the capacity  $W^{l_i}(R_i)$ , we consider two possible capacities  $W^{fg}(R_i)$  and  $W^{bg}(R_i)$  regarding if the node corresponds to background or foreground (resp. source or sink):

- The capacity  $W^{fg}(R_i)$  of the source (foreground) class for a region  $R_i$  is given by:

$$W^{fg}(R_i) = \begin{cases} K & \text{if } |R_i \cap I_{FG}| > \alpha \\ 0 & \text{if } R_i \cap I_{MASK} \neq \emptyset \\ \lambda P^{fg}(R_i)K & \text{otherwise} \end{cases}$$

- The capacity  $W^{bg}(R_i)$  of the sink (background) class for a region  $R_i$  is given by:

$$W^{bg}(R_i) = \begin{cases} 0 & \text{if } |R_i \cap I_{FG}| > \alpha \\ K & \text{if } R_i \cap I_{MASK} \neq \emptyset \\ \lambda P^{bg}(R_i)K & \text{otherwise} \end{cases}$$

where  $I_{FG}$  is the image of foreground detection,  $I_{MASK}$  corresponds to the mask,  $\alpha$  and  $\lambda$  are two parameters fixed by the genetic algorithm.

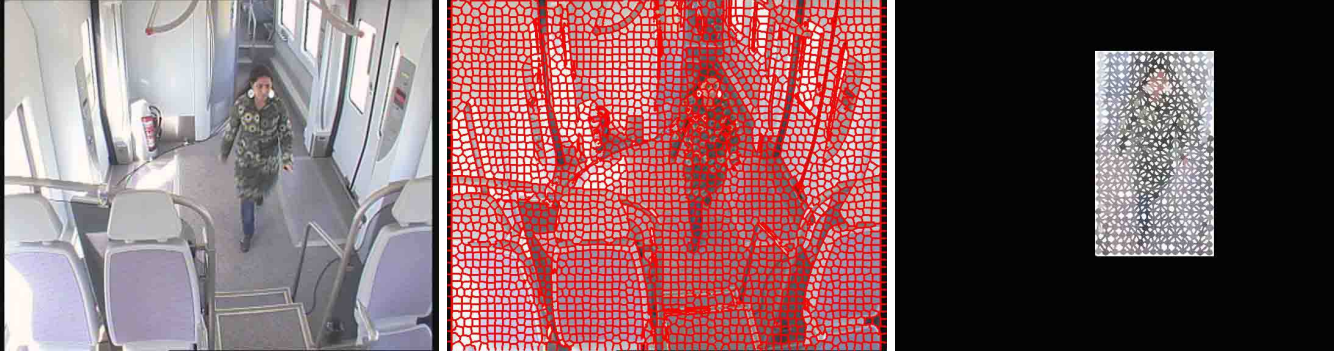


Fig. 2. Superpixel Graph: from left to right, original image, SLIC superpixels segmented image, mask image  $I_{MASK}$  with superpixel graph superimposed.

It is useful to notice that, in the case where regions cannot be classified as foreground or background (the "otherwise" case in the capacities), we must predict, for each region, a probability of belonging either to the source (for foreground)  $P^{fg}(R_i)$  or to the sink (for background)  $P^{bg}(R_i)$ . These quantities are defined by:

$$P^{fg}(R_i) = \frac{\sum_{c=1}^3 \sum_{p_i \in R_i} P_c^{fg}(p_i, R(i))}{3 \times m}$$

$$P^{bg}(R_i) = \frac{\sum_{c=1}^3 \sum_{p_i \in R_i} P_c^{bg}(p_i, R(i))}{3 \times m}$$

where  $p_i$  denotes a pixel of the region  $R_i$ ,  $m$  corresponds to the size of the image, and  $c$  to the RGB color space channel number.

Finally, we define, for a pixel  $p_i$ , and a color space channel number  $c$ , the probability of belonging to these 2 classes (source and sink)  $P_c^*(p_i, R(i))$ . To that aim, we use two color histograms (one for the foreground and another for the background) directly calculated from the foreground detection and the bounding box (as illustrated in Figure II-E.2).

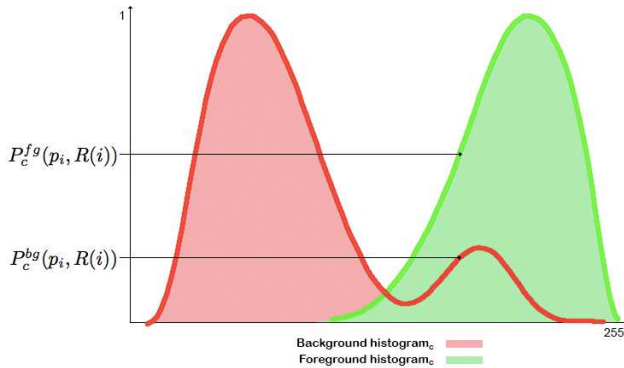


Fig. 3. Computation of the probability of a pixel to belong to background or foreground.

In order to optimize the processing, computations of similarities and capacities are performed during the conversion of the segmented image into a region adjacency graph.

### III. EXPERIMENTAL RESULTS

Our people extraction method has been tested with the BOSS European project database [21]. Video sequences were recorded inside a train in motion during a sunny afternoon and with many people that move in front of the cameras installed in the different wagons of the train. From this database, we have chosen one sequence including many scientific locks such as fast brightness changes, shadows, scrolling background and noise. In this sequence (4258 frames including 1439 frames with different people), we have 12 persons that move in front of a color camera. To evaluate the proposed approach, we have created 4258 reference segmentations corresponding to the passing of the 12 persons. The obtained people extraction results have been compared to the reference segmentation with a F-Measure. For the tests, a genetic algorithm is used. It is applied with a population of 256 chromosomes. Each chromosome corresponds to a complete setting of our proposed people extraction method. It is composed of 30 genes representing methods and parameters. As illustrated below, we have divided a chromosome into 6 parts to represent the different possible configurations to explore:

- 6 genes for the steps of foreground detection and pre-processing 1: 1 to design the chosen filter, 3 for the associated filters' settings, 1 for the chosen colorimetric invariant, 1 for the chosen foreground detection method.
- 6 genes for the steps of background detection and pre-processing 2: 1 for the chosen filter, 3 for the associated filters' settings, 1 for the chosen colorimetric invariant, 1 for the chosen background detection method.
- 6 genes for the steps of shadow removal and pre-processing 3: 1 for the chosen filter, 3 for the associated filters' settings, 1 for the chosen colorimetric invariant, 1 for the chosen shadow removal method.
- 3 genes for the step of morphological filtering: 1 for the chosen morphological method and 2 for the associated morphological settings.
- 6 genes for the steps of bounding box extraction and pre-processing 4: 1 for the chosen filter, 3 for the associated filters' settings, 1 the chosen for colorimetric invariant, 1 for the padding of the bounding box.
- 3 genes for the step of graph-cut clustering: 1 for the



Fig. 4. People extraction: original images (lines 1 and 3), people extraction with the proposed method (lines 2 and 4).

distance metric and 2 for the graph cut settings.

The fitness measure used to perform the genetic optimization is the  $F_1$  score:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

with:

$$\text{precision} = \frac{|\text{well detected foreground pixels}|}{|\text{foreground pixels}|}$$

$$\text{recall} = \frac{|\text{well detected foreground pixels}|}{|\text{detected pixels}|}$$

The genetic algorithm optimization is then used to choose the best combination. The training phase of the algorithm is realized on a cluster of 96 cores (2 Ghz - 768Go) and has been done only on the displacement of the first two persons. This corresponds to a short part of the sequence: 530/4400 frames that refer to the displacement of two persons out of the 12 we considered. Regarding the testing phase, we have used a laptop with one core (2Ghz - 16 Go). The proposed method is implemented with OpenCV and has an average processing time of 8 seconds per image.

After genetic optimization, the best settings of the proposed people extraction method are given below. One can notice that these results correspond to an average, and several differences can exist for a given optimum:

- Steps of foreground detection and pre-processing 1: no filter and color invariants, foreground detection based on fuzzy logic method [5].
- Steps of background detection and pre-processing 2: bi-lateral filter, affine normalization colorimetric invariant, background detection based on fuzzy logic method [4]

- Steps of shadow removal and pre-processing 3: Gaussian blur filter,  $l_1l_2l_3$  colorimetric invariant, shadow removal based on physical methods [14].
- Step of morphological filtering : no filtering is required in average on this database.
- Steps of bounding box extraction and pre-processing 4: Gaussian blur filter,  $m_1m_2m_3$  colorimetric invariant, 60 pixels of padding in average (15% of the image) .
- Step of graph-cut clustering:  $L_1$  norm based on the RGB color information.

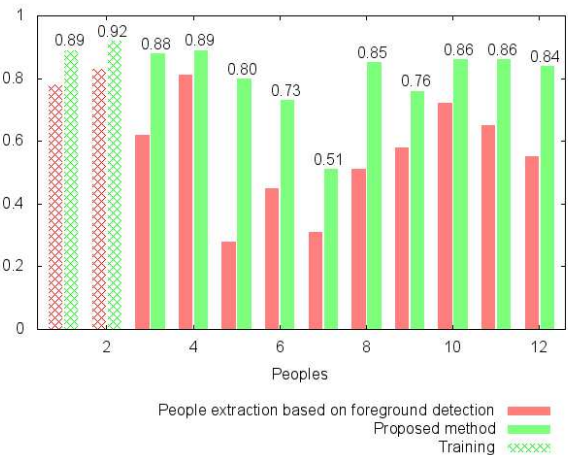


Fig. 5. People extraction results: the proposed method versus state-of-the-art foreground detection.

Figure 5 shows, on one sequence of the BOSS project, the people detection results obtained by the proposed method as





Fig. 6. Displacement of one people in the train.

compared to those obtained by the best foreground detection algorithm of the literature (found by the genetic algorithm). It is important to notice that F-Measures provided in Figures 5 and 7 correspond to an average one calculated on all the displacements of the different persons (around 120 frames per person). Globally, the proposed method (in green) gives better results than the best foreground detection method of the literature (in red). Indeed, we obtain a F-Measure of 0.82 as compared to 0.59. For the proposed method, the F-Measure is always upper than 0.73, excepted for the 7<sup>th</sup> person (a difficult case illustrated in the first image of second row of Figure 4). For a better visualization, a representative image of each people (during his displacement in the train) is given in Figure 4. As explain previously, the proposed method permits to obtain very good people extraction results. Nevertheless, we can explain the poor value of F-Measure (0.51) of the 7<sup>th</sup> person by the detection of a seat classified as people (due to high similar texture). This problem can however be easily corrected with a tracking process. Indeed, for the moment, we only consider independent frames and no tracking of the detected persons is considered within adjacent video frames. Figure 7 illustrates more precisely the people extraction results obtained by the best foreground detection method and the proposed method. To that aim, we give the average precision and recall scores for each people (around 120 frames per people). For our application and the considered database, the best foreground detection method of the literature (found by the genetic algorithm) gives satisfactory results in terms of precision (few background pixels are classified as foreground) but it gives a very poor recall score (foreground pixel are not detected at all). In contrast, our proposed method gives both very good precision and recall scores and the gap between these values is small. Figure 6 shows few representative images of the displacement of one person moving in the train. One can notice, that person is well extracted in each frame whether near or far from the camera. On the one hand, there are few background pixels classified as foreground (pixels located between arms or legs). On the other hand, the contour of the extracted person is not always accurate. This problem is due to the superpixel segmentation that has some difficulties to segment blurred

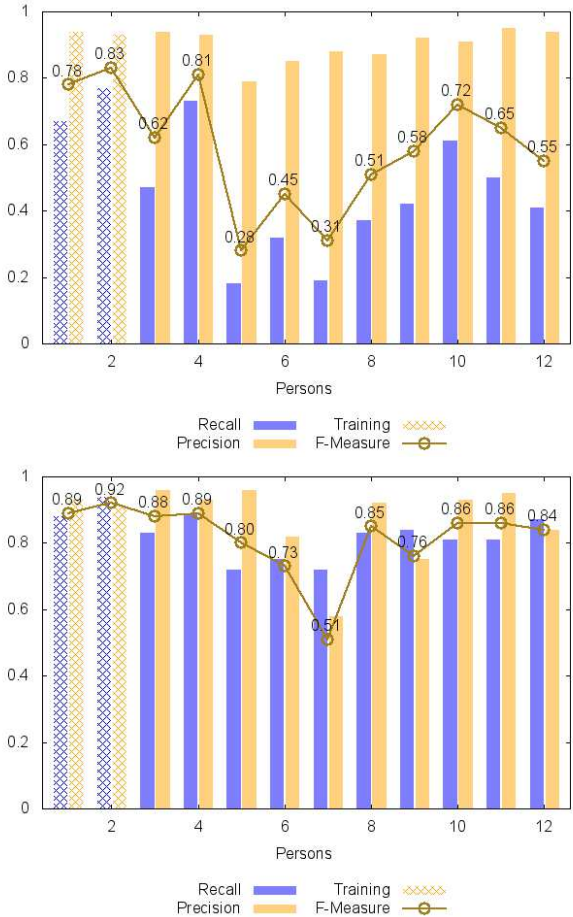


Fig. 7. Precision/Recall results of people extraction obtained by the best foreground detection method (top) and the proposed method (bottom).

regions when the displacement is fast. Figure 8 shows the F-measure for the displacement of this person in front of a camera installed in the train (100 frames). For all images of the sequence (except for one frame), one can notice that the proposed method gives better results than those obtained by the best foreground detection method of the literature. However, we have actually one image (43<sup>th</sup> frame) that gives a bad people extraction, but this problem will be corrected in future works by an integrated tracking method.

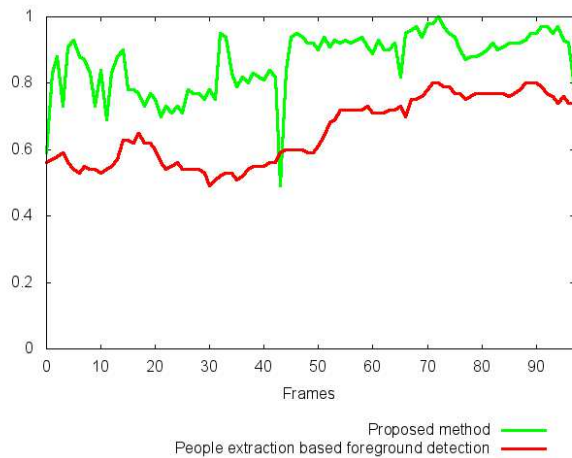


Fig. 8. F-Measure of one people for his displacement in front of the camera.

#### IV. ACKNOWLEDGMENTS

The authors would like to thank the BOSS project for the use of the database on which the proposed approach has been tested.

#### V. CONCLUSION

In this paper, we have shown that the best foreground detection method of the literature are not enough robust and accurate to obtain good people extraction in transport environment considering some scientific locks as fast brightness changes, noise, shadow, scrolling background, etc. Nevertheless, using foreground detection seems to provide a good basis for determining the position of people. On this observation, we have proposed a new method of people extraction based on superpixel segmentation coupled with graph cut clustering initialized by the information of foreground detection. To strengthen this foreground detection, we have proposed to use robust pre-processing (based on filters and colorimetric invariants) and post-treatment steps (shadow removal and morphological filtering). The best settings of the proposed method are determined by a genetic algorithm and evaluated with the F-Measure criteria. One can conclude that the people detection results obtained with the proposed method are better than those obtained with the best foreground detection method of the literature (F-Measure of 0.82 versus 0.59). Moreover, our method allows to decrease the gap between precision and recall scores, thereby further robustifying our method. Even if the proposed method widely improves people extraction results, the contours of the extracted persons are not perfect and some people areas may be not detected in several frames. To resolve these problems, our future works will consists in introduce a temporal information in the graph cut clustering along with tracking across frames.

#### REFERENCES

- [1] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4 – 21, 2014.
- [2] S. D. Hordley, G. D. Finlayson, G. Schaefer, and G. Y. Tian, "Illuminant and device invariant colour using histogram equalisation," *Pattern Recognition*, vol. 38, p. 2005, 2005.
- [3] T. Bouwmans, F. E. Baf, and B. Vachon, "Background modeling using mixture of gaussians for foreground detection - a survey," in *Recent Patents on Computer Science*, vol. 1, no. 3, 2008, pp. 219–237.
- [4] F. E. Baf, T. Bouwmans, and B. Vachon, "Fuzzy integral for moving object detection," in *FUZZ-IEEE*, 2008, pp. 1729–1736.
- [5] H. Zhang and D. Xu, "Fusing color and texture features for background model," in *Fuzzy Systems and Knowledge Discovery*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2006, vol. 4223, pp. 887–893.
- [6] M. Sigari, N. Mozayani, and H. Pourreza, "Fuzzy running average and fuzzy background subtraction: Concepts and application," in *Int. J. Comput. Sci. Network Security*, vol. 8, no. 2, 2008, pp. 138–143.
- [7] J. Yao and J. M. Odobez, "Multi-layer background subtraction based on color and texture," in *CVPR, Workshop on Visual Surveillance (VS2007)*, 2007, pp. 1–8.
- [8] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1168–1177, 2008.
- [9] Y. Goyat, T. Chateau, L. Malaterre, and L. Trassoudaine, "Vehicle trajectories evaluation by static video sensors," *IEEE International Conference on Intelligent Transportation Systems*, pp. 864 – 869, 2006.
- [10] A. Sobral, "BGSLibrary: An opencv c++ background subtraction library," in *IX Workshop de Visao Computacional (WVC'2013)*, Rio de Janeiro, Brazil, Jun 2013. [Online]. Available: <https://github.com/andrewsobral/bgslibrary>
- [11] A. Sanin, C. Sanderson, and B. C. Lovell, "Shadow detection: A survey and comparative evaluation of recent methods," *Pattern Recognition*, vol. 45, no. 4, pp. 1684 – 1695, 2012.
- [12] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 918–923, 2003.
- [13] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1337–1342, 2003.
- [14] J. Huang and C. Chen, "Moving cast shadow detection using physics-based features," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2310–2317, 2009.
- [15] J. Hsieh, W. Hu, C. Chang, and Y. Chen, "Shadow elimination for effective moving object detection by gaussian shadow modeling," *Image and Vision Computing*, vol. 21, no. 6, pp. 505–516, 2003.
- [16] A. Leone and C. Distanto, "Shadow detection for moving objects based on texture analysis," *Pattern Recognition*, vol. 40, no. 4, pp. 1222–1233, 2007.
- [17] A. Sanin, C. Sanderson, and B. C. Lovell, "Improved shadow removal for robust person tracking in surveillance scenarios," in *Proceedings of the 20th International Conference on Pattern Recognition*, 2010, pp. 141–144.
- [18] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [19] Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary region segmentation of objects in n-d images," in *IEEE International Conference on Computer Vision*, vol. 1, 2001, pp. 105–112.
- [20] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.
- [21] Boss european project (on bord wireless secured video surveillance). [Online]. Available: <http://www.celtic-boss.org>