Robust and accurate change detection under sudden illumination variations

Luigi Di Stefano Federico Tombari Stefano Mattoccia
Errico De Lisi
Department of Electronics Computer Science and Systems (DEIS)
Viale Risorgimento 2, 40136 - Bologna, Italy
Advanced Research Center on Electronic Systems (ARCES)
Via Toffano 2/2, 40135 - Bologna, Italy
University of Bologna
{ldistefano, ftombari, smattoccia}@deis.unibo.it, errico.delisi@studio.unibo.it

Abstract

A challenging task for most change detection applications is to accurately segment the foreground from the background under the presence of heavy illuminations changes. In fact, photometric changes can be easily misinterpreted as structural changes, thus leading to false positives in the change mask. In this paper we present a novel approach to deal with this problem, which deploys a recently proposed robust visual correspondence measure together with a tonal registration procedure. Furthermore, we provide an experimental comparison of our approach with other proposals specifically designed for the task of being robust to sudden illumination changes, demonstrating the effectiveness of our approach.

1 Introduction

Change detection aims at detecting structural changes occurring in time in a scene by analyzing a frame sequence. This is a key task in most advanced video-surveillance applications, for the mask highlighting changed pixels (*change mask*) typically represents the input data to higher level vision algorithms.This is the case of traditional single-view as well as more recent and advanced multiple-views systems [1, 6]. The most common change detection approach is referred to as *background subtraction*: given the current frame, F, and a model of the background of the scene, B, the change mask is obtained as the difference between F and B. This approach assumes that the background model is available or can be obtained by processing a short sequence of frames at initialization time (e.g. as shown in [5]). A wide variety of change detection algorithms has been proposed in literature, so as to address issues such as illumination changes, camouflage and vacillating background. A recent survey providing good coverage of this research area is given by [9].

A major issue for most practical change detection applications is represented by sudden illumination changes occurring in the scene. Properly dealing with such a problem is a challenging task for change detection algorithms since the resulting photometric variations can be easily misinterpreted as structural changes, leading to many false positives in the change mask. Algorithms [3, 8, 11] that specifically aim at detecting changes robustly with respect to sudden illumination variations typically take the decision of voting a pixel as changed or unchanged based on a given spatial support (e.g. a 3×3 or larger window centered at the pixel under evaluation). Typically such algorithms rely on a parametric (e.g. linear [3, 8]) or non parametric (e.g. order preserving [11]) model for the false image changes due to sudden illumination variations. However, it is well known that such algorithms suffer from an aperture problem, i.e. they cannot detect as changed the pixels belonging to untextured foreground regions. As a result, they typically enable to detect the borders of foreground objects but not accurately their interior parts. Moreover, the use of a spatial support rather than pointwise background subtraction implies inaccuracy as regards localization of the borders of the detected foreground objects. Coarse-to-fine approaches such as [2] can alleviate these problems.

In this paper we propose a novel approach aimed at obtaining robust and accurate foreground segmentation under sudden illumination variations. The paper is structured as follows. Section 2 describes the proposed algorithm along its various stages, then Section 3 discusses the computational requirements dealing with the implementation of the proposed approach. Section 4 shows some qualitative experimental results as well as a comparison with other approaches. Finally, Section 5 draws conclusions and future work.

2 Proposed approach

As depicted in Fig. 1, the proposed approach consists of three processing stages. In the first stage, a visual correspondence measure robust to photometric variations [10] is used to extract a subset of pixels in the current frame that can be marked as background with a high confidence level. Once such a subset, referred to as F_B, is obtained, it can be usefully employed to remove the photometric distortion between F and B. To this purpose, in the second stage the algorithm computes the transformation that aligns tonally the current frame, F, to the background image, B, using as support subset F_B. In the third stage, the final change mask is achieved by a pixelwise subtraction between F and the tonally registered background image, B_R. In the following we provide more details on these three processing stages.

Robust visual correspondence In order to get F_B we basically match the points in the background image to the current frame. To achieve robustness with respect to outliers and noise, a block-based approach is used: that is, for each pair of correspondent points in B and F, a $M \times M$ surrounding block is considered, and a matching score is computed between the two blocks. Points having a score higher than a given threshold are included into F_B.

Since the task of the first stage is to find a set of points in F that can be reliably classified as background even in presence of strong illumination changes, we need to determine a suitable matching measure for comparing blocks. To this purpose, let's discuss Fig. 2 where, for the sake of simplicity, we consider only two kind of regions, i.e. uniform and highly-textured. When dealing with a uniform region in both F and B (case *a* in Fig. 2), photometric differences between F and B can occur due to either variations of the illumination conditions of the background scene as well as to structural changes induced by a uniform foreground object. Thus, in this case the required matching measure should yield a low score, for nothing can be said reliably on whether the point belongs to the background or not. As for cases *b,c,d*, it is easy to observe that the matching score should be low too, since there's evidence of the presence of a foreground object. Finally, when the background is highly textured and the texture pattern does not change in spite of possible photometric changes (case *e*), it is reasonable to flag the point as background with a high confidence level. Hence, in case *e* we should get a high score from the required matching measure.

Based on the above considerations, we adopt the visual correspondence measure recently proposed in [10], referred to here as MF (*matching function*). This measure matches

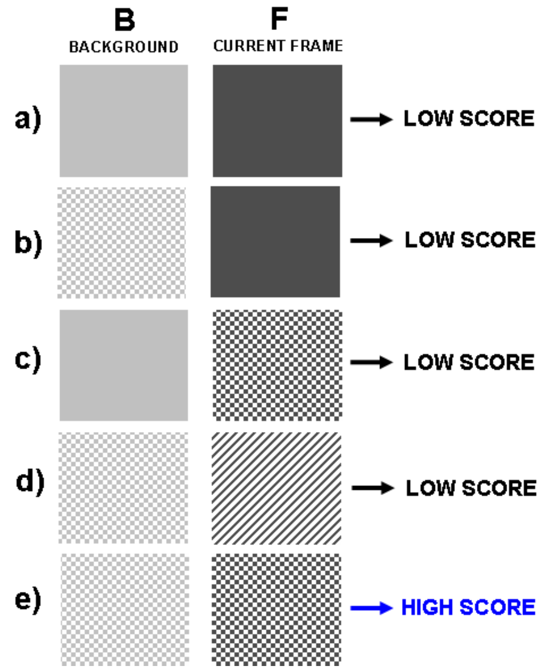

Figure 2. Reasoning concerning the robust visual correspondence stage

corresponding blocks of two images by implicitly checking an ordering constraint. Since photometric variations tend not to violate the ordering of intensities in a neighborhood of pixels, MF allows to handle sudden and strong illumination variations between the background scene and the current frame. In particular, let F and B be functions of two coordinates, (i, j). Then, once defined a vector of pixel differences computed at a point (i, j) on F:

$$\delta_F(i,j) = \begin{bmatrix} F(i-1,j) - F(i+1,j) \\ F(i,j-1) - F(i,j+1) \end{bmatrix} \tag{1}$$

and, similarly, at a point (i, j) on B:

$$\delta_B(i,j) = \begin{bmatrix} B(i-1,j) - B(i+1,j) \\ B(i,j-1) - B(i,j+1) \end{bmatrix} \tag{2}$$

the MF function computed at block centered at coordinates (x, y) is defined as:

$$MF(x,y) = \frac{N(x,y)}{D_F(x,y) \cdot D_B(x,y)} \tag{3}$$

where $\circ$ represents the dot product between two vectors, and

$$N(x,y) = \sum_{i=1}^{M} \sum_{j=1}^{M} \delta_F(x+i, y+j) \circ \delta_B(x+i, y+j) \tag{4}$$

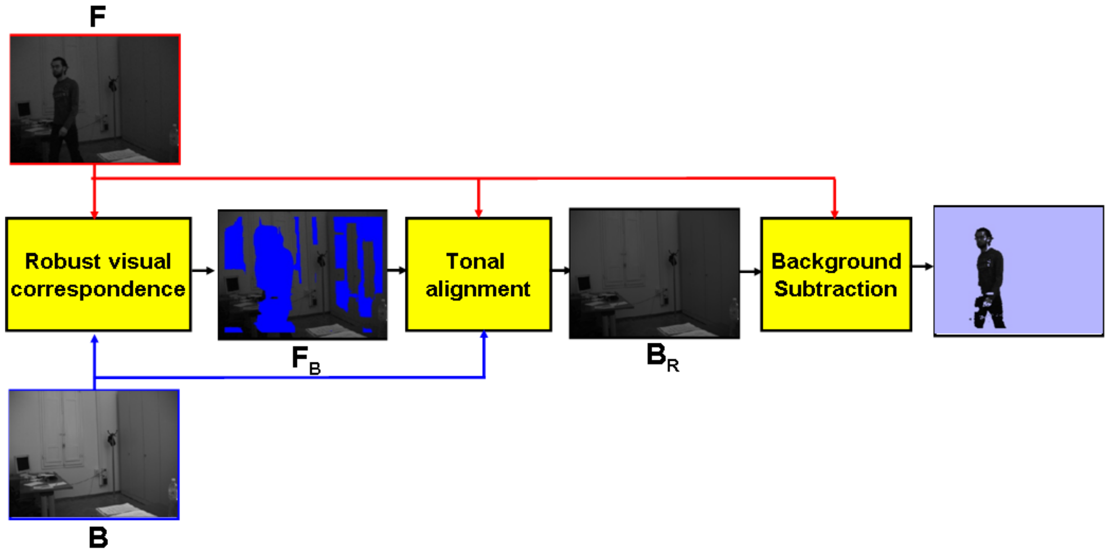

Figure 1. Flow diagram of the proposed change detection algorithm

$$D_F(x,y) = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} \delta_F(x+i,y+j) \circ \delta_F(x+i,y+j)} \quad (5)$$

$$D_B(x,y) = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} \delta_B(x+i,y+j) \circ \delta_B(x+i,y+j)} \quad (6)$$

Hence, this function computes the normalized correlation between differences of intensities of neighboring pixels, the normalization at denominator of (3) allowing the measure to range between $[-1, 1]$. As previously introduced, once $MF(x, y)$ is computed for all point coordinates (x, y) of F and B, a threshold is adopted so that only points having a score higher than a certain value are included into F_B.

It can be argued easily that MF tries to match the high contrast regions (i.e. the intensity edges) of the two blocks under comparison, since only high intensity differences can provide high contributions to the correlation score. Hence, MF behaves exactly as pointed out in Fig. 2. In fact, only two highly textured and highly correlated patterns can provide a high matching score (case *e*), while the presence of at least one untextured region (cases *a,b,c*) or of two textured but uncorrelated patterns (case *d*) yields a low score.

Tonal alignment At this point of the algorithm, F_B represents a subset of F denoting pixels that reliably belong to the current background. Hence, B is tonally aligned to F by applying the *histogram specification* method [4]. In the evaluation of the IMF (*Intensity Mapping Function*) that aligns B to F only the set of corresponding points that belong to the mask F_B is taken into account. By applying the IMF obtained from the histogram specification method to B we get a novel background, B_R, where the photometric distortions have been removed.

Background subtraction Finally, a simple pixelwise difference between B_R and F highlights structural changes by correctly extracting foreground regions. It is worth pointing out that since background subtraction is carried out pixelwise, it is not affected by the aperture problem mentioned in the introduction and allows for accurately detecting the borders as well as interior parts of foreground objects. Obviously, false negatives can still be found due to the possible camouflage between the tonally registered background and the foreground objects.

3 Computational requirements

The bottleneck of the proposed algorithm can be identified in the first stage. In fact, denoting as W and H respectively the width and height of images, the computation of N (4) requires theoretically $2M^2WH$ differences, multiplications and summations, and similarly D_F (5) and D_B (6) require each $2M^2WH$ differences, multiplications and summations plus WH square roots.

Nevertheless, all differences concerning the background model B can be computed once for all at initialization, while all differences concerning the current frame can be computed once for all at each new frame, accounting for a total of $2WH$ differences. Furthermore, since the matching

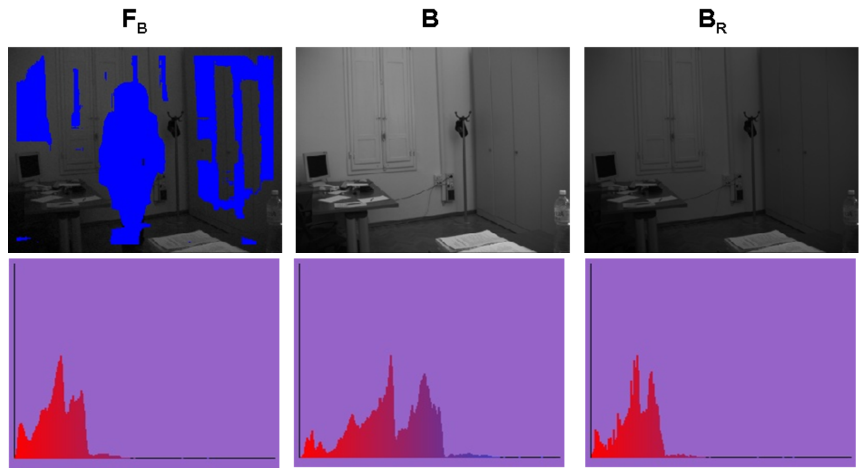

Figure 3. The tonal alignment stage registers, on the basis of the histogram specified by F_B (left), the histogram of the background B (center), obtaining the tonally registered background B_R (right)

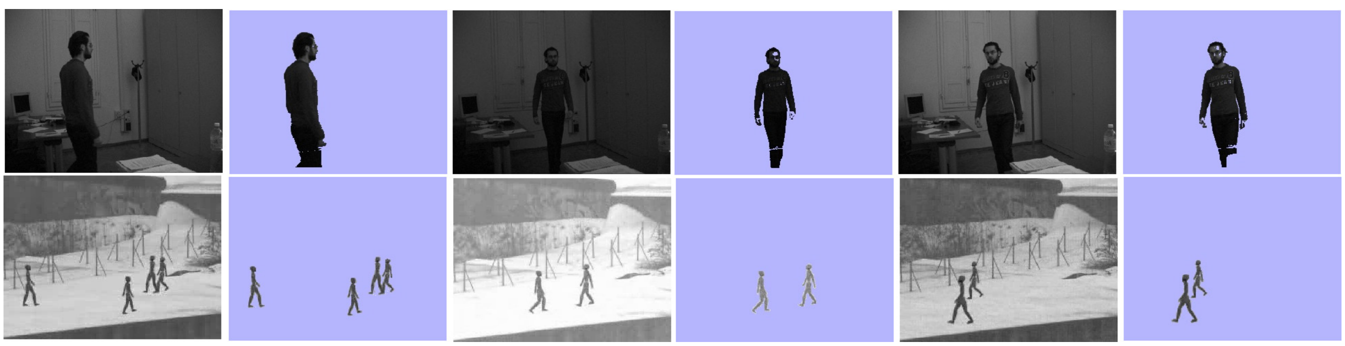

Figure 4. Change masks yielded by the proposed algorithm in two sequences affected by sudden brightness variations

metric has to be applied on neighboring blocks in a ”sliding window” fashion, well-known incremental approaches such as [7] allow for further shrinking the total number operations required for the computation of N. In particular, this can be done by computing the product of corresponding pixel differences once for all at each new frame (accounting for $2WH$ multiplications), then running two Box-Filter instances to compute the final accumulation term, which accounts for $8WH$ summations overall. Similar deductions apply to the computation of the two denominator terms, D_F and D_B. The latter can be computed once for all at initialization, while, by means of a strategy analogous to that used to compute N, the former requires only $2WH$ additional multiplications, $8WH$ summations and WH square roots overall at each new frame.

Thanks to these optimizations, our implementation of the proposed algorithm easily deals with the real-time requirements of many change-detection applications (e.g. video-surveillance), with an average frame rate of 15 fps on a 320×240 frame size.

4 Experimental results

In this section we provide some experimental results dealing with the proposed approach. In particular, our algorithm has been tested with real as well as with a synthetic benchmark sequence: real sequences are affected by sudden and strong brightness variations due to illumination changes, while the synthetic one[1] by artificial brightness changes.

First of all, we show some qualitative results. In Fig. 4 some screenshots of the change masks obtained by the proposed algorithm on a real sequence (above) and a synthetic one (below) are presented, which clearly prove that

[1] available at: *http://muscle.prip.tuwien.ac.at/data_here.php*

the proposed approach is able to accurately segment foreground objects in presence of heavy photometric changes.It is worth pointing out that no morphology operator was used at any stage of the algorithm in order to obtain these results: nevertheless, uniform regions of the foreground are correctly segmented and no false positives arise on low textured regions of the background.

Moreover, we also propose a quantitative comparison between our approach and other proposals. The testing sequence was the syntethic sequence used for obtaining the previous results, since it comes together with the groundtruth. For what regards the comparison, as representative of change detection algorithms that model false image changes according to a linear relation we consider the Normalised Cross Correlation (NCC) between pixel intensities. As for algorithms relying on checking the order preservation of intensities we consider the Rank transform [12]. We also consider as baseline for comparison the basic pixelwise background subtraction approach (BBS).

For a fair comparison, we used the same block side for each algorithm (i.e. equal to 7). Then, for what regards the other parameters of each algorithm (in particular, the threshold for the final change mask), in order to determine the best parameter set of each algorithm we selected as a measure of comparison the *Precision*, i.e. the ratio between the true positives (TP) and the sum between true positives and false positives (FP):

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

and the *Recall*, i.e. the ratio between the true positives and the sum between true positives and false negatives (FN):

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

In order to obtain experimental results, we started from the observation that most change detection algorithms, especially for video-surveillance applications, require to have a minimum guaranteed value of Recall. Hence, for different thresholds of minimum Recall (i.e. 70%, 80%, 90%, we selected for each algorithm the optimal parameter set maximizing the Precision value. Such results are shown in Tab. 1. It is worth pointing out that we fixed the maximum constraint value of Recall to 90%, since with higher values all algorithms would provide Precision values lower than 50%, which would result in very poor change masks (the number of false positives being higher than the number of true positives). Moreover, it is worth noting that also for these results no post processing was added to the output of the evaluated algorithms, similarly no morphology operator was used at any stage of the evaluated algorithms.

From the Table it is easy to infer that the proposed algorithm is the most robust and accurate between the evaluated ones, since it always outperforms the other approaches in terms of Precision for all different constraint values of Recall. In addition, Fig. 5 shows, for a single frame of the evaluated testing sequence, the outputs of the various algorithms at the different constraint values of Recall. In addition, in the first row of the Figure the background model as well as the current frame together with the correspondent ground truth frame are shown. These results qualitatively confirm the trend shown in Tab. 1, proving that our approach provides overall the most accurate results.

Table 1. Best values of Precision yielded by the evaluated algorithms with different constraint values on Recall.

	$> 70\%$	$> 80\%$	$> 90\%$
Proposed	87.3	81.7	52.2
NCC	59.6	57.2	43.0
RANK	24.5	18.8	13.1
BBS	2.2	1.9	1.7

5 Conclusions and future work

This paper proposed a novel approach to perform change detection robust to sudden illumination variations. The proposed strategy is based on the use of a recently proposed visual correspondence measure robust to these kind of distortions, which allows to reliably extract parts of the background in the current frame. This allows to succesively apply a tonal alignment procedure in order to remove the photometric distortions, which allows to highlight structural changes keeping in the same time a high grade of accuracy in the foreground segmentation. Experimental results demonstrate our approach is able to accurately detect motion along video sequences highly subject to sudden brightness distortions, outperforming previous approaches.

6 Acknowledgements

We thank Advanced Computer Vision GmbH - ACV for providing the benchmark motion detection video sequence used in the experimental section.

References

[1] J. Berclaz, F. Fleuret, and P. Fua. Robust people tracking with global trajectory optimization. In *Proc. CVPR*, volume 1, pages 744–750, 2006.

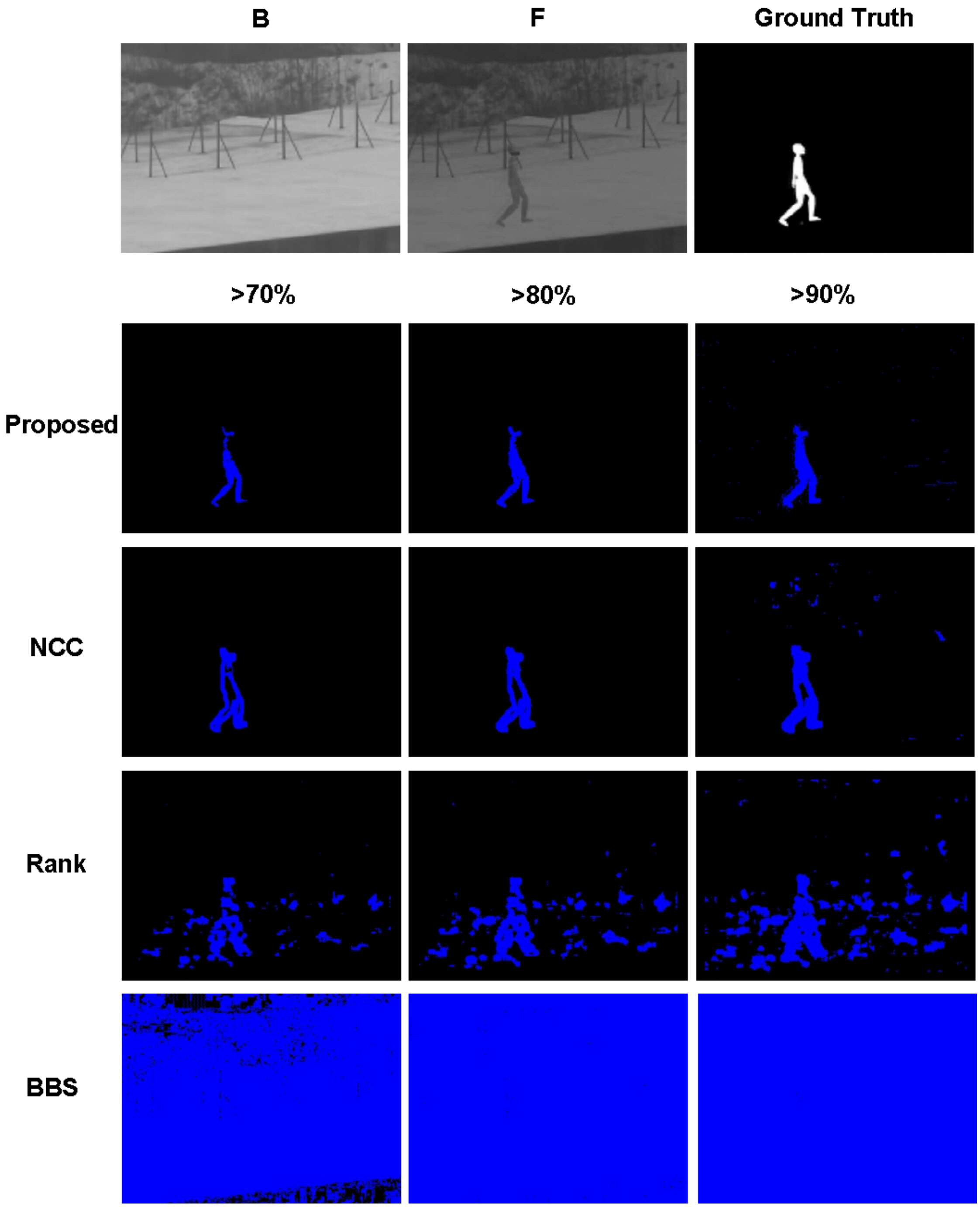

Figure 5. Comparison of outputs yielded by the evaluated algorithms on the same sequence and with the same constraint values on Recall used for results in Tab. 1. First row, from left to right: background model B, current frame F, Ground Truth

[2] A. Bevilacqua, L. Di Stefano, and A. Lanza. Coarse-to-fine strategy for robust and effcient change detectors. In *Proc. AVSS*, 2005.

[3] E. Durucan and T. Ebrahimi. Change detection and background extraction by linear algebra. *Proc. IEEE*, 89(10):1368–1381, 2001.

[4] R. Gonzalez and R. Woods. *Digital Image Processing*. Prentice Hall, 2nd edition, 2002.

[5] D. Gutchess and al. A background model initialization algorithm for video surveillance. In *Proc. ICCV*, volume 1, pages 733–740, 2001.

[6] S. Khan and M. Shah. A multiview approach to tracking people in crowded scenes using a planar homography constraint. In *Proc. ECCV*, volume 4, pages 133–146, 2006.

[7] M. Mc Donnel. Box-filtering techniques. *Computer Graphics and Image Processing*, 17:65–70, 1981.

[8] N. Ohta. A statistical approach to background subtraction for surveillance systems. In *Proc. ICCV*, volume 2, pages 481–486, 2001.

[9] R. Radke, S. Andra, O. Al-kofahi, and B. Roysam. Image change detection algorithms: a systematic survey. *IEEE Trans. Image Processing*, 14(3):294–307, 2005.

[10] F. Tombari, L. Di Stefano, and S. Mattoccia. A robust measure for visual correspondence. In *Proc. ICIAP*, 2007.

[11] B. Xie, V. Ramesh, and T. Boult. Sudden illumination change detection using order consistency. *Image and Vision Computing*, 2(2):117–125, 2004.

[12] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proc. European Conf. Computer Vision*, pages 151–158, 1994.