Detecting motion by means of 2D and 3D information

Federico Tombari Stefano Mattoccia Luigi Di Stefano
Fabio Tonelli
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, Italy
{ftombari, smattoccia, ldistefano}@deis.unibo.it, fabio.tonelli@studio.unibo.it

Abstract

The use of depth information, such as that provided by a stereo system, can potentially enhance the robustness of motion detection algorithms with regards to typical factors such as sudden illumination variations, shadows and camouflage. In this paper we propose and compare two change detection strategies based on two views which exploits 2D and 3D information in order to deal with typical change detection issues. Preliminary experimental results provided using two different stereo matching algorithms on real stereo sequences show the usefulness of the proposed approach in the considered scenarios.

1 Introduction

Detecting motion in video sequences is a fundamental requirement for many higher-level vision tasks such as object classification, tracking, event detection (e.g. stolen or abandoned object). A common domain of application for such tasks is video-surveillance, e.g. with the aim of detecting intrusions. Major issues related to the motion detection process are as follows. It is difficult to correctly segment out moving objects when they look similar to the background of the scene (*camouflage*). The motion detection process is very sensitive to sudden illumination variations of the scene. It is difficult to filter out shadows from the detected foreground. By using more than one view, e.g. by means of a stereo camera, it is possible to obtain 3D information on the surveyed scene. This allows to exploit, in addition to scene radiance information, also depth information concerning scene 3D structure, so as to achieve higher robustness with regards to one or more of the above mentioned issues [4, 6–8]. In the next sections we propose and compare two change detection strategies based on two views which exploits 2D and 3D information in order to deal with typical change detection issues.

The paper is structured as follows. Section 2 describes in detail the processing stages of proposed change detection strategies based on the analysis of two synchronized views acquired by a stereo camera. Section 3 shows qualitative experimental results obtained by processing real stereo sequences. Finally, Section 4 summarizes the contributions of the paper and outlines future work.

2 Proposed approach

This section describes a novel change detection approach which jointly exploits depth information coming from a 3D device and 2D brightness information. Information on scene changes is recovered by means of two different strategies. The former, referred to as *3D Output*, mainly relies on depth information, and aims at being robust to camouflage, shadows and sudden illumination changes. The latter, referred to as *2D Output*, aims at obtaining robustness with regards to sudden illumination changes as well as accuracy in foreground segmentation. The final change masks determined by the two outputs will be referred to as, respectively, C_{3D} and C_{2D}.

As depicted in Fig. 1, the proposed approach can be outlined as a 4-stage algorithm. The imaging sensor used is a stereo camera. In particular, we assume that a calibrated stereo setup is available, so that the image pairs retrieved from the camera can be properly rectified.

Stereo Matching For each new rectified image pair coming from the stereo device, a stereo matching algorithm (see [10] for a survey on this topic) is used in order to retrieve a dense disparity map relative to the observed scene.

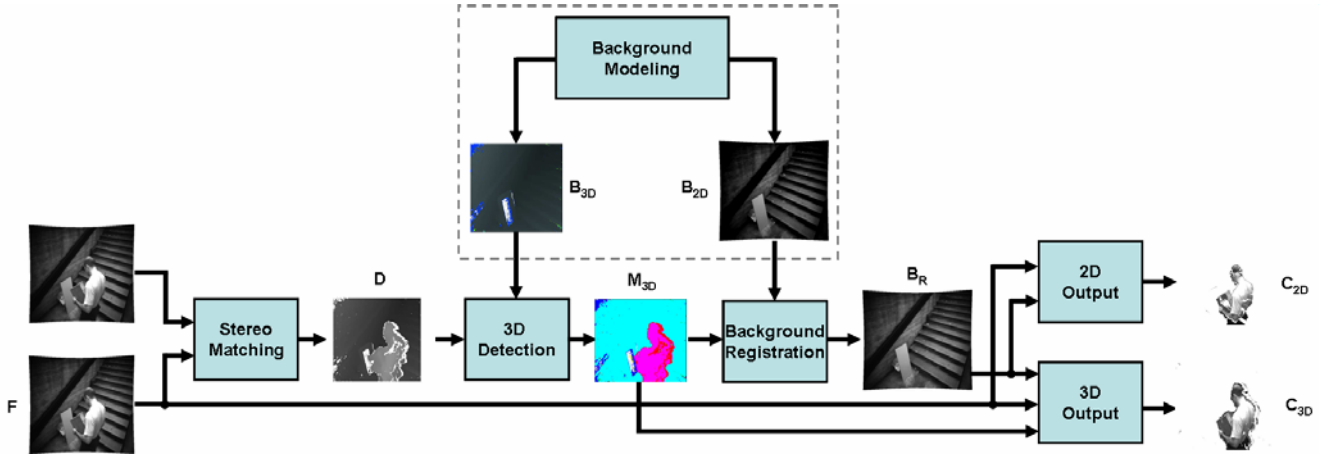

Figure 1. Flow diagram of the proposed approach

In particular, in our experiments we have used two different local stereo matching algorithms. The former, referred to as SMP (*Single Matching Phase*) [3], is an algorithm that allows to obtain dense disparity maps in real-time. The latter, referred to as *Variable Windows* [11], is an algorithm that holds the potential to retrieve more accurate depth borders compared to SMP thanks to the use of a variable aggregation stage, even if at a higher computational cost which renders it slower than SMP (near real-time).

We first describe briefly how SMP computes disparities for each point p_r of the reference image. Let I_r, I_t be respectively the reference image and the "other" image. Given a disparity range D, for each point $p_{t,d}$ on I_t belonging to the disparity range induced by p_r a similarity measure is applied between a squared window centered on p_r and all squared windows centered on $p_{t,d}$. The adopted similarity measure is the SAD (*Sum of Absolute Differences*). The selected disparity d_r for point p_r is that relative to the point $p_{t,d}$ that yielded the lowest SAD score on its window. Then *uniqueness*, *distinctiveness* and *sharpness* constraints (see [3] for details) are used to eliminate ambiguous disparity values. Hence, the pixels of the final disparity maps obtained by SMP are labeled either as valid disparity values, or as points violating the constraints (referred here to as non-matched points, NM). SMP relies on incremental calculations techniques [9] and delivers disparity maps in real-time.

Differently to SMP, Variable Windows uses the Birchfield-Tomasi [1] measure as a point wise matching cost and selects the more appropriate aggregation support evaluating a useful range of window sizes/shapes. Although slower than SMP, also this approach relies on incremental calculation techniques (i.e. integral images [2, 12]) for efficient disparity maps computation. The matching selection is based on a Winner Takes All (WTA) strategy and hence all points are labeled as valid.

Figure 2. An example of B_{2D} and B_{3D}

Background Modeling The proposed approach requires that, at initialization, two background models of the scene are built: the former, B_{2D}, is determined from the brightness values of the reference view, the latter, B_{3D}, is determined from the disparity values provided by the stereo matching stage. Both models are built by processing a short initialization sequence of frames. While B_{2D} captures the radiance information of the scene background, B_{3D} represents a model of the scene 3D structure. In order to obtain

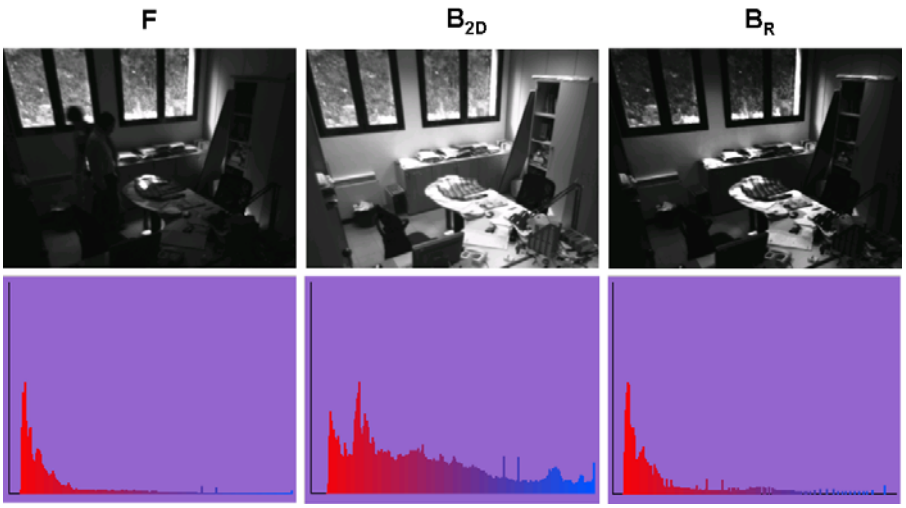

Figure 3. During the background registration stage, the histogram of the background model B_{2D} (center) is registered according to the specification given by the histogram of the frame F (left), yielding the new background model B_R (right)

B_{2D}, a classical method is used, that is each value in B_{2D} represents the mean brightness of a pixels over the initialization sequence. Conversely, by means of three different thresholds T_1,T_2,V, each pixel of B_{3D} is associated to 4 different classes:

1. Valid disparity: if a valid disparity is retrieved by the matching algorithm for more than T_1 frames during the initialization sequence, and the variance of disparities is less than V. The final disparity assigned to the pixel in B_{3D} is the mean disparity over the initialization sequence.
2. High-variance disparity: if a valid disparity is retrieved by the matching algorithm for more than T_1 frames during the initialization sequence, but the variance of disparities is equal or higher than V. The pixel is depicted in green in B_{3D}.
3. Non-match: if a NM is retrieved by the matching algorithm for more than T_2 frames during the initialization sequence. This pixel is depicted in white in B_{3D}.
4. Unreliable point: if a valid disparity is retrieved by the matching algorithm for equal or less than T_1 frames during the initialization sequence, and a NM for equal or less than T_2 frames. The pixel is depicted in blue in B_{3D}.

An example of B_{2D} and B_{3D} is shown in Fig. 2.

It is worth observing that when a WTA strategy is adopted (as in [11]) the last two conditions are not meaningful.

3D detection Once the background models are built, at each time instant a new frame from the reference view, F, together with its disparity map, D, is obtained. At this stage B_{3D} and D are deployed to compute a mask, M_{3D}, which encodes with different colors the various correspondences between D and B_{3D}. In particular, as shown in Figure 1:

1. $\mathbf{b_1}$ (*light blue*): a valid disparity point in B_{3D} corresponding to a valid disparity point in D, the difference between the two disparity values being less than a certain threshold.
2. $\mathbf{b_2}$ (*pink*): a valid disparity point in B_{3D} corresponding to a valid disparity point in D, the difference between the two disparity values being equal or higher than a certain threshold.
3. $\mathbf{b_3}$ (*blue*): an unreliable point in B_{3D}.
4. $\mathbf{b_4}$ (*green*): a high-variance disparity in B_{3D}.
5. $\mathbf{b_5}$ (*white*): a non-match point in B_{3D} corresponding to a NM point in D.
6. $\mathbf{b_6}$ (*red*): a non-match point in B_{3D} corresponding to a valid disparity point in D.
7. $\mathbf{b_7}$ (*yellow*): a valid disparity point in B_{3D} corresponding to a NM point in D.

The information encoded in mask M_{3D} is useful to perform the tonal alignment procedure performed in the background registration stage, as well as in the generation of the final change mask C_{3D}.

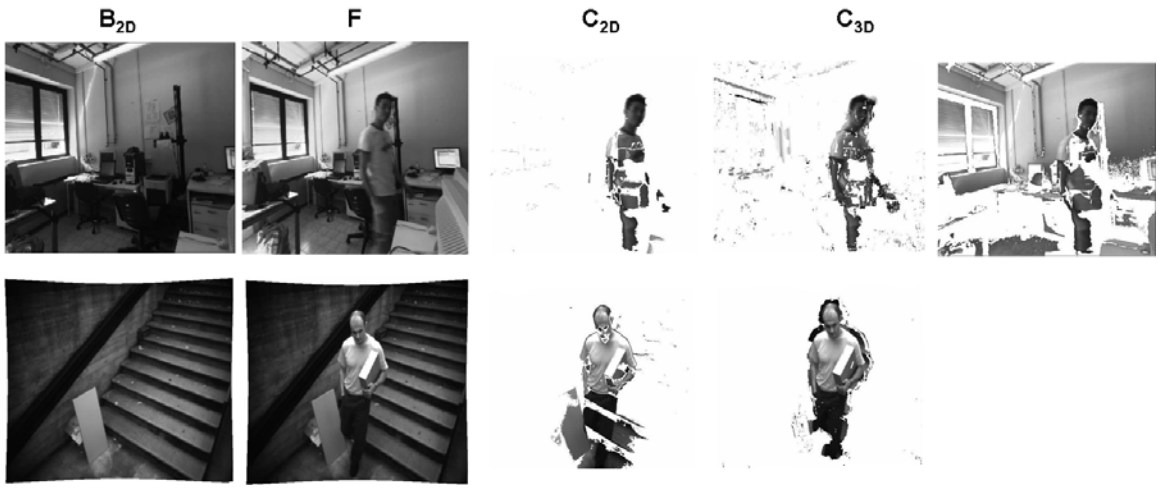

Figure 4. Experimental results on an indoor (above) and outdoor stereo sequence. Top right picture: output of a classical background subtraction algorithm

Background registration A further stage of the algorithm, which will be particularly useful for the generation of C_{2D}, deals with the elimination of photometric distortions between F and B_{2D} by tonally registering B_{2D} with respect to F. In particular, the evaluation of the Intensity Mapping Function that tonally aligns B_{2D} to F is done by applying the *histogram specification* method [5]. For this aim a set of pixels belonging to F which reliably belongs to the scene background has to be extracted. This can be easily done by exploiting the information included in M_{3D}: in particular, the set of pixels chosen as representative of the background of the scene are selected as those tagged as $\mathbf{b_1}$ (i.e. in *light blue* color) on M_{3D}, as they denote unchanged valid disparities between D and B_{3D}. The output of this stage is a novel background model, B_R, where photometric distortions with respect to the current frame F have been removed. In Fig. 3 an example is presented, which shows the application of the background registration to a frame. In particular, the histogram of the current frame F (left) is used as a model to tonally register the histogram of the 2D background model, B_{2d} (center). The histogram and the image of the registered background B_R, obtained as output of this stage, are shown on the right side of the Figure.

2D Output Once the background is tonally aligned to the current frame, a simple pixel wise frame difference can highlight structural changes robustly with respect to possible brightness distortions. Hence, C_{2D} is generated by subtracting B_R from F. The main strength of this approach is robustness against sudden illumination changes as well its accuracy in the foreground segmentation stage. Nevertheless, the shadow and camouflage issues are not properly dealt with.

3D Output Conversely to 2D Output, this approach relies more on 3D information. In particular, all pixels whose disparity value was reliably determined on M_{3D} as unchanged ($\mathbf{b_1}$, light blue color) are set as unchanged on C_{3D}. Similarly, all pixels whose disparity value was reliably determined on M_{3D} as changed ($\mathbf{b_2}$, pink color) are set as changed on C_{3D}. For what means pixels whose disparity can not be determined reliably, all pixels that *might* denote a structural change (i.e. from NM to a valid disparity or vice versa, that is $\mathbf{b_6}$ and $\mathbf{b_7}$ on M_{3D}) are set on C_{3D} as the correspondent value on C_{2D}. Finally, all remaining pixels ($\mathbf{b_3}$, $\mathbf{b_4}$, $\mathbf{b_5}$ on M_{3D}) for whom nothing can be said are set as unchanged on M_{3D} [1]. This solution represents a robust approach toward shadows, camouflage and sudden illumination changes, but foreground segmentation is less accurate compared to the other approach due to the depth borders inaccuracy brought in by the stereo matching process.

3 Experimental results

In this section we show some qualitative experimental results obtained on three different sequences [2], acquired with a Videre Design stereo camera, referred to as Indoor, Outdoor and Office. In the first sequence, which is indoor, photometric distortions are induced by real illumination changes. The second sequence, which is outdoor, is

[1] Another solution is to have these pixels represent regions in the final change mask where detection can not be performed.

[2] Sequences available at: www.vision.deis.unibo.it/smatt

affected by illumination changes as well as by the strong presence of shadows and camouflage problems. The Office sequence, which is indoor, shows the strong photometric distortions induced by switching lights on and off. Fig. 4 shows the two outputs C_{3D} and C_{2D} on a frame of the Indoor and Outdoor sequences using the disparity maps computed by the SMP algorithm. No morphology operator was used at any stage of the algorithm in order to obtain these results, which demonstrate that our approach is in general robust to photometric distortions. Moreover, it can be noted that 2D Output generally retrieves more accurately foreground borders (both sequences), and that 3D Output suffers much less of shadows (outdoor sequence) and camouflage (both sequences). Finally, top right frame in Fig. 4, which shows the output of a classical background difference algorithm, demonstrates the strong entity of the photometric distortions, and how less accurate is the segmentation compared to the proposed approaches. Figures 5 and 6 show the results provided by the change detection strategies proposed in this paper on the more challenging Office sequence using, respectively, the SMP and Variable Windows algorithms. In both figures we show 9 out of 195 frames. Similarly to the Indoor and Outdoor sequences no morphology operator was used at any stage. The Office sequence presents dramatic artificially induced illumination changes clearly observable comparing frames $\sharp 15$, $\sharp 85$ and $\sharp 185$ in Figures 5 and 6. It is worth observing that under these difficult conditions the disparity maps generated by the two stereo matching algorithms are significantly noisy. Nevertheless, as shown in the two rightmost columns of Figure 5, although the shapes of the objects are not accurately retrieved, the proposed strategies provide a robust detection. Moreover, similarly to the previous sequences, C_{3D} output provides less accurate detection of borders compared to C_{2D} output but it the seems less affected by shadows (on the wall and on the table). Similar considerations apply by observing the results shown in Figure 6. However, in this case, the WTA strategy adopted by the Variable Windows algorithm results in even more noisy disparity maps and consequently in more noisy results provided by C_{2D} and C_{3D} outputs. As for Variable Windows, these results do not highlight a better accuracy in recovering the object borders in C_{3D} compared to SMP. We think that this is due to the WTA strategy adopted in our current implementation of the algorithm and that a higher accuracy may be obtained by enforcing into the algorithm constraints such as uniqueness, distinctiveness and sharpness or left-right consistency.

4 Conclusions and future work

In this papers we propose and compare two change detection strategies which exploit 2D and 3D information in order to deal with typical change detection issues. Preliminary experimental results on three real stereo sequences show that both approaches can deal with strong natural and artificial disturbance factors. Experimental results also show that the 2D Output generally retrieves more accurately foreground borders while the 3D Output is more robust to shadows and camouflage. Future work is aimed at evaluating more accurate and more robust stereo matching algorithms as well as at embodying within the proposed framework robust multi-view change detectors such as [7, 8].

5 Acknowledgments

We acknowledge with thanks Elisa Addimanda for the implementation of the Variable Windows algorithm.

References

[1] S. Birchfield and C. Tomasi. A pixel dissimlarity measure that is insensitive to image sampling. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(4):401–406, 1998.

[2] F. Crow. Summed-area tables for texture mapping. *Computer Graphics*, 18(3):207–212, 1984.

[3] L. Di Stefano, M. Marchionni, and S. Mattoccia. A fast area-based stereo matching algorithm. *Image and Vision Computing*, 22(12):983–1005, 2004.

[4] C. Eveland, K. Konolige, and R. Bolles. Background modeling for segmentation of video-rate stereo sequences. In *1998 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'98)*, pages 266–271, 1998.

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

[6] G. Gordon, M. Darrel, M. Harville, and J. Woodfill. Background estimation and removal based on range and color. In *1999 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'99)*, 1999.

[7] Y. Ivanov, A. Bobick, and J. Liu. Fast lighting independent background subtraction. *International Journal of Computer Vision*, 37(2):199–207, 2000.

[8] S. Lim, A. Mittal, L. Davis, and N. Paragios. Fast illumination-invariant background subtraction using two views: Error analysis, sensor placement and applications. In *2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 1071–1078, 2005.

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

[10] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3):7–42, 2002.

[11] O. Veksler. Fast variable window for stereo correspondence using integral images. In *2003 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, 2003.

[12] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

Figure 5. Experimental results on 9 frames of the Office stereo sequence using the disparity map provided by the SMP algorithm [3]. (First column) - Reference image F of the stereo pair. (Second column) Background model B_{2D} registered according to the specification given by the histogram of the frame F. (Third column) - Disparity map D computed by the SMP algorithm. (Fourth column) - Change mask C_{2D} provided by the proposed *2D Output* approach. (Fifth column) - Change mask C_{3D} provided by the proposed *3D Output* approach.

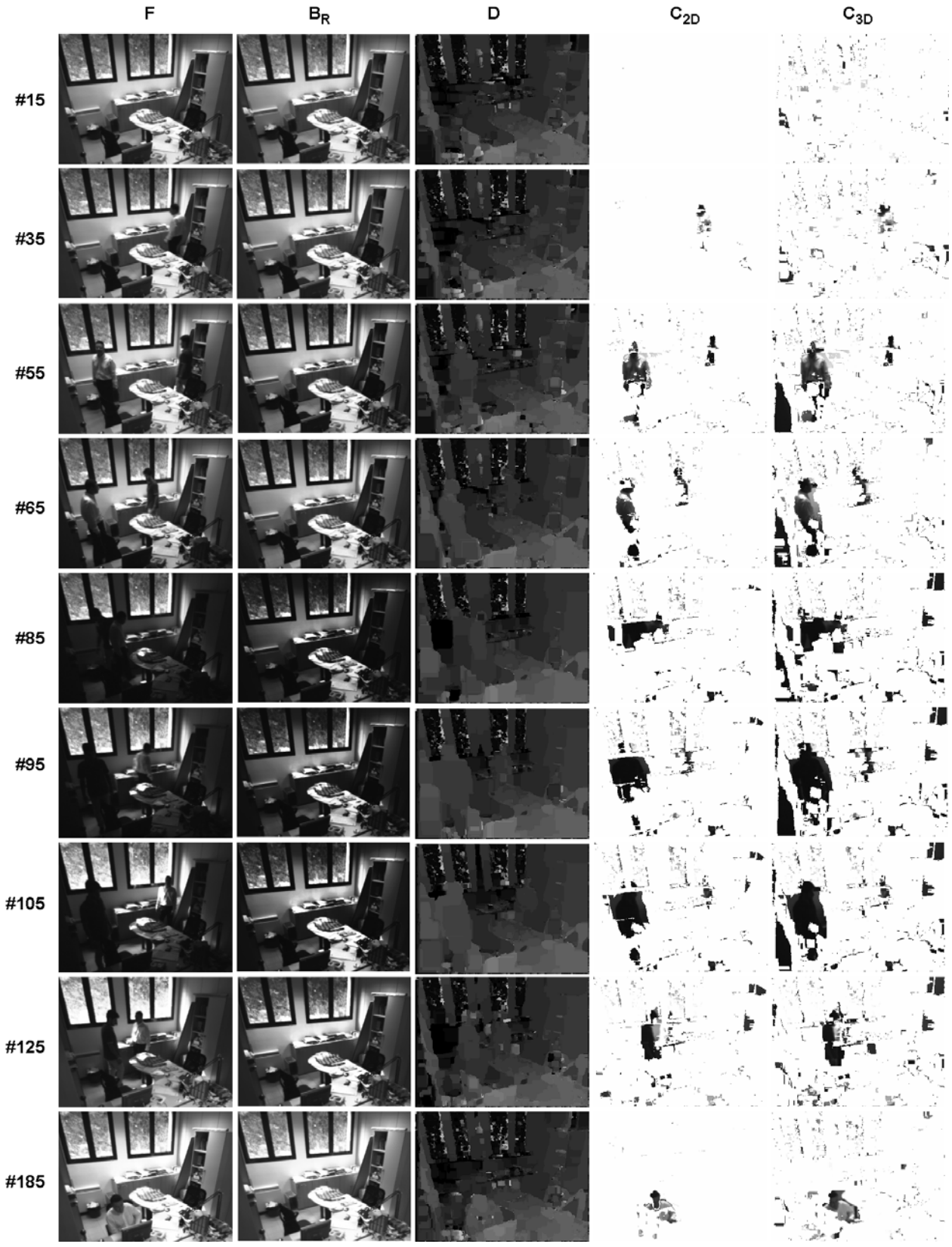

Figure 6. Experimental results on 9 frames of the Office stereo sequence using the disparity map provided by the Variable Windows [11] algorithm. (First column) - Reference image F of the stereo pair. (Second column) Background model B_{2D} registered according to the specification given by the histogram of the frame F. (Third column) - Disparity map D computed by the Variable Windows algorithm. (Fourth column) - Change mask C_{2D} provided by the proposed *2D Output* approach. (Fifth column) - Change mask C_{3D} provided by the proposed *3D Output* approach.