ISSN 1060-992X, Optical Memory and Neural Networks (Information Optics), 2011, Vol. 20, No. 3 pp. 201–209. © Allerton Press, Inc., 2011.

Efficient Ghost Removal in Motion Detection with Patch-Corrected Background Differentiation

Yuri Boiko and Pierre Payeur

School of Electrical Engineering and Computer Science University of Ottawa, Ottawa, ON, Canada

e-mail: yboik074@uottawa.ca; ppayeur@site.uottawa.ca

Received May 10, 2011; in final form, July 10, 2011

Abstract—An efficient ghost removal technique is proposed as an extension to adaptive background differentiation for motion detection. The pixels of the first frame in the sequence representing moving objects are replaced with the values taken for the same pixels from the memory bank where those pixels are identified as non-moving ones. The memory bank is built of the frames immediately following or, alternatively preceding, the initial frame of the analyzed sequence. This allows creating the initial background model with no moving pixels. Parameters optimization is conducted for specific case of traffic control system application. Experiments demonstrate that threshold reduction is beneficial to achieve completeness of the ghost removal. Additionally, a second improvement is introduced to reduce for the noise by non-stationary cameras which are shown to be efficiently compensated by a second derivative in the temporal differentiation when working with videos at a sufficiently high frame rate.

Keywords: motion detection; ghosts reduction; background differentiation; traffic monitoring.

DOI: 10.3103/S1060992X11030088

INTRODUCTION

Traffic monitoring in urban areas becomes an increasingly important application where machine vision systems are found to be efficient in traffic management, control and its optimization [1–7]. As a by-product of the traffic optimization at the intersections comes reduction of environmental pollution and gas consumption, as well as ease of traffic congestions and reduction of commuting time losses. Automated visual traffic monitoring and control systems rely on a stream of images of selected parts of the road and/or of an intersection. The sequence of images is analyzed by the system to reach the control decision. An example can be detecting approaching cars at an intersection, for which the automated system may adjust the traffic light to minimize the waiting time at light intersection. In such circumstances a motion detection algorithm is used for identification of the moving vehicle.

STATE-OF-THE-ART REVIEW

The temporal differentiation and background differentiation algorithms are traditional image processing approaches suitable for the task of detecting a moving vehicle in the traffic flow [1–3]. There had been number of advances [4] aiming to improve the performance of the algorithms for specific tasks in traffic monitoring, in which adaptivity had been employed to extract and update the background image for reliable foreground detection, particularly aiming to separate moving objects from the background. Bi-models of background in HSV color were explored in [5] via the minimum, the maximum and the largest inter-frame absolute difference of per static pixel, which were adaptively updated by synthesizing pixel level, object level and frame level methods. In [6], the optical flow was computed and utilized as a feature in a higher dimensional space to subtract the background. Background registration technique allows adapting effectively to changing environments [7]. The proposed modification to background differentiation stems from earlier reports [2, 3], in which ghost removal approach combined temporal differentiation followed by weighted averaging of the background to achieve ghost elimination with robustness to illumination variations. The adaptivity here comes at a cost—in the limited time-frame the ghost elimination is not complete, as it is seen in [2] on Fig. 4b, where updated background after 6 s still contain residues of the moving vehicles seen in Fig. 4a of [2]. The problem of achieving a complete correction of the background, even at

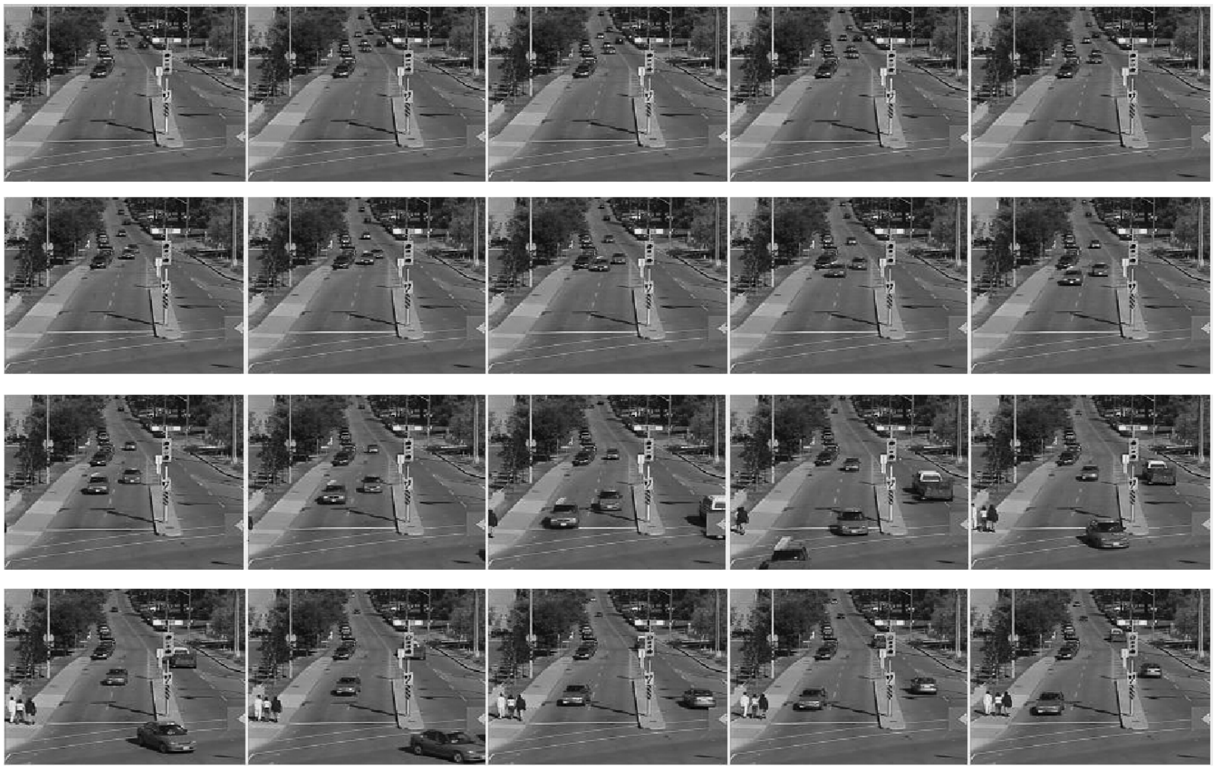


Fig. 1. A sequence of 20 consecutive frames from the video of the urban road intersection used in the experiments. (From top/left to the right/bottom).

the cost of sacrificing some adaptivity, requires its solution. The adaptivity may be redundant under some circumstances, particularly when there are no significant lighting changes at a scale of few seconds.

PROBLEM FORMULATION

The problem can be specified by its implementation example, where ghosts clearly occur and require algorithmic elimination. For that task the video sequence of 20 frames shown in Fig. 1 made with slightly non-stationary camera on urban road intersection at sunny weather condition is used. Non-stationary condition of the camera introduces additional difficulty in extra source of noise in motion detection, because non-stationarity causes all objects in the scene to experience slight synchronous movement. Gray scale of 256 levels is used in image representation $I(i, j, t)$ for the time stamp $t = 1, 2, 3, \ldots 20$, where (i, j) represent the coordinates of a given pixel in the frame. Adaptive background differentiation algorithm consists of the following steps.

Step 1. Assign the very first frame (time-stamped $t = 1$) of the sequence as an initial temporary background model $\mathrm{TB} = \mathrm{B}(i, j, t) = I(i, j, t)$.

Step 2. Find the absolute difference with the next frame time-stamped at $(t + 1)$ to obtain $\Delta I(i, j, t) = |I(i, j, t + 1) - \mathrm{TB}|$.

Step 3. Threshold the obtained differential map $\Delta I(i, j, t)$ against the selected threshold value Z to assign "1" to pixels (i, j) with intensity values $\Delta I(i, j, t) \geq Z^* \Delta I_{\max}$ (moving pixels) and "0" otherwise (static pixels), where $\Delta I_{\max}$ is the maximal value in the set $\Delta I(i, j, t)$, thus obtaining the motion map $M(i, j, t)$ associated with the last frame at $(t + 1)$. Note, that to distinguish motion maps from the correction masks the reverse gray scale is used in first case, i.e. in $M(i, j, t)$ the black represents moving pixels and white the static ones, as in Fig. 1, while in second case the opposite is true, as in Fig. 3 and Fig. 6. This choice is consistent with that in [2], which makes it easier to make comparisons.

Step 4. Update the background model TB with the following rule: if $M(i, j, t) = 1$ (i.e. moving pixel) then keep the pixel intact (i.e. $\mathrm{TB} = \mathrm{B}(i, j, t - 1)$), otherwise $\mathrm{TB} = I(i, j, t)$.

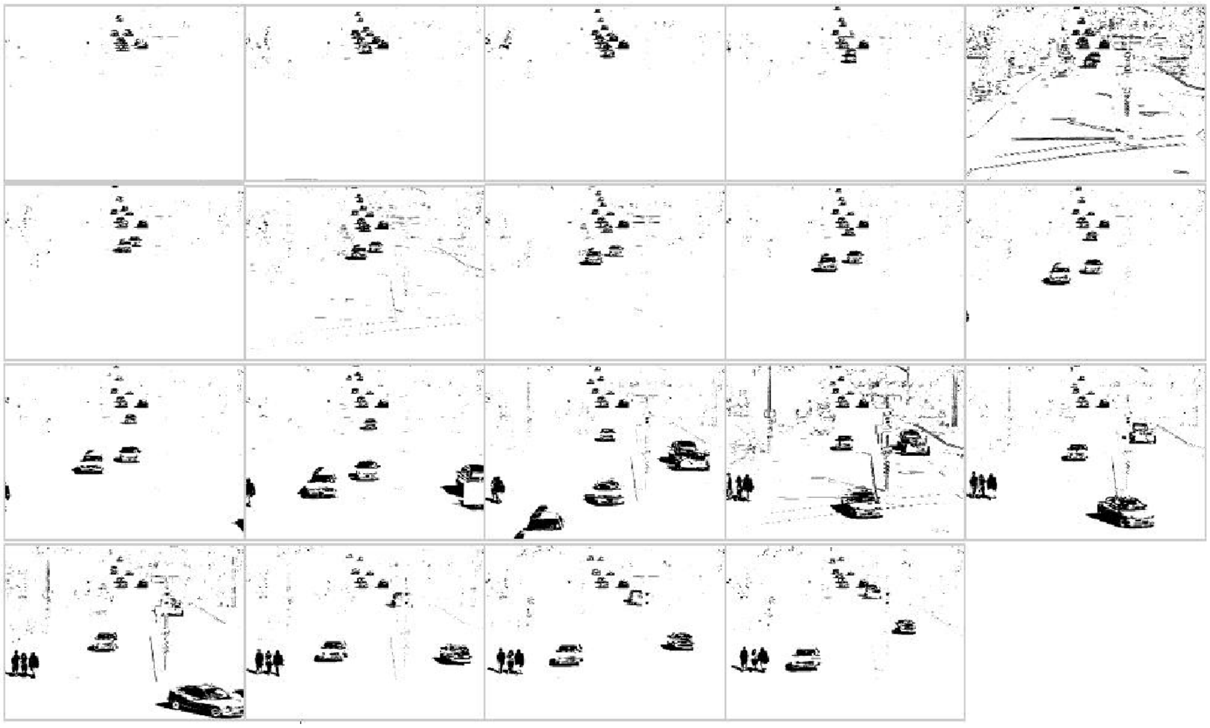
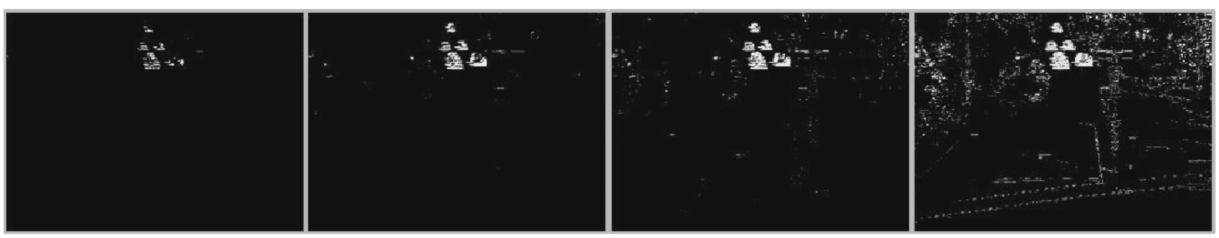


Fig. 2. Motion maps for the video sequence of Fig. 1, obtained by traditional adaptive background approach.



(a) $Z^\circ = 0.33$ (b) $Z^\circ = 0.16$ (c) $Z^\circ = 0.10$ (c) $Z^\circ = 0.05$

Fig. 3. Generated correction masks for various threshold levels Z° in completion of the steps 1° through 3°: (a) 0.33; (b) 0.16; (c) 0.10 and (d) 0.05

(a) $Z^\circ = 0.33$ (b) $Z^\circ = 0.16$ (c) $Z^\circ = 0.10$ (c) $Z^\circ = 0.05$

Fig. 4. Effect of threshold value Z° on the moving pixels elimination from background model - lower Z° benefits in completeness of the operation: (a) 0.33; (b) 0.16; (c) 0.10 and (d) 0.05.

Step 5. Upgrade the time-stamp by 1, i.e. set $t = t + 1$.

Step 6. If new time-stamp leads to a valid frame in the sequence, then go to step 2 for next iteration. Otherwise, end.

This algorithm offers adaptivity of the background due to its updating step 4. If there are no moving pixels in the first frame, the described adaptive background differentiation would be a viable option for motion detection of urban traffic monitoring systems. However, this algorithm is vulnerable to the presence of moving objects in the very first frame of the series, because the pixels associated with it are not discriminated against in step 4 and therefore these pixels are excluded from the updating process. This results in persisting ghosts, which are present in all motion maps as if the same moving objects are present on all the frames and do not abandon the place they are supposed to move from. This is illustrated in Fig. 2. The objects from $M(t = 1)$ continue to be present on all other motion maps even though no objects at that location are seen in the respective frames in Fig. 1.

Therefore the task here can be formulated as the ghost removal from the motion maps, where the ghosts are originated in the adaptive background differentiation from the moving objects of the very first frame in the video sequence.

PATCHED BACKGROUND DIFFERENTIATION ALGORITHM

In order to achieve the goal of persistent ghost removal, it is proposed to eliminate the very source of the ghosts' appearances, i.e. the moving objects from the initial background model in step 1 of the above algorithm. To accomplish that, first the correction mask needs to be obtained by the same temporal differentiation between the initial two frames, namely the 1st and the 2nd one, of the video sequence as proposed in [2]. The suggested modification to the algorithm introduced in [2] is in correcting not the mask, but rather the background model as well as in substituting weighted averaging of the masked background by algorithmic recollection from the recorded memory of the values of the pixels identified by the initially obtained mask as moving ones and therefore requiring replacement. This is to be achieved by forming a bank of memory frames, which is assigned to serve a specific set of frames in the sequence of the recorded video observation as a source of additional information about the background model. Depending on particular application tasks, such bank of memory frames may either coincide (totally or in part) with the set of frames it is assigned to serve, or precede it. Understandably, the mode of selecting the memory frames has to be predefined in the algorithm. The proposed version of the patched background differentiation algorithm therefore includes the following steps (marked with "°" to distinguish it from the above step 1 to step 6 of the adaptive background differentiation algorithm).

<u>Step 1°.</u> Assign the first frame to be a temporary background: $\mathrm{TB} = \mathrm{B}(i, j, t = 1) = I(i, j, t = 1)$.

<u>Step 2°.</u> Apply temporal differentiation between the initial two frames in the sequence to identify the moving pixels, which require correction for the background, to obtain $\Delta I(i, j, t = 1) = |I(i, j, t = 2) - \mathrm{TB}\,|$.

<u>Step 3°.</u> Based on the result of step 2°, generate the correction binary mask $C(i, j, t = 1)$, in which according to the chosen threshold value Z° the moving pixels are distinguished from the non-moving ones (attributing 1's to moving pixels and 0's to the rest of the field) , i.e.

$$C(i, j, t = 1) = \begin{cases} 1 \text{ if } \Delta I(i, j, t = 1) \geq Z^\circ * \Delta I_{\max} (\text{moving pixels}); \\ 0 \quad \text{esle} \quad (\text{static pixels}); \end{cases}$$

<u>Step 4°.</u> Assign the memory bank frames $P(i, j, n)$, $n = 1, 2, 3, \ldots, n_{\max}$, where $n_{\max}$, is the maximal number of the frames in the memory bank. In general, the memory bank can be built from either the frames immediately preceding or following the initial frame of the sequence, at the condition that in either case it must be based on the same background model as that of the sequence under consideration.

<u>Step 5°.</u> Generate replacement values $R(i,j)$ for the TB pixels marked on the correction mask $\mathrm{C}(i, j, t = 1)$ as moving ones, taking those values from the memory bank. Here prior to taking the replacement value, the two sequential frames from the memory bank are to be subjected to temporal differentiation to confirm that for this particular frame-pair the replacement pixel is a non-moving one (while it was a moving one for step 2°). A double condition, connected with a logical AND has to be satisfied here: "$\mathrm{C}(i, j, t = 1) = 1$ AND $\Delta P(i, j, n) \geq Z^* \Delta P_{\max}$", where $\Delta P(i, j, n) = |P(I, j, n + 1) - P(i, j, n)|$, $n = 1, 2, 3, \ldots, n_{\max} - 1$; $\Delta P_{\max}$ is the maximal value of $\Delta P(i, j, n)$ in the set, and Z is threshold for this operation, selected in the same way as the one in step 3 of the traditional adaptive background differentiation algorithm and which therefore differs from that of $Z°$ in step 3° above. Frames in the bank are sequentially scanned for each relevant pixel either until the replacement is found or until the end of the bank is reached. In the last case, averaging of the replacement value via the stored bank frames offers potential to compensate for the illumination changes in the scene and thus to minimize its impact on background model formulation.

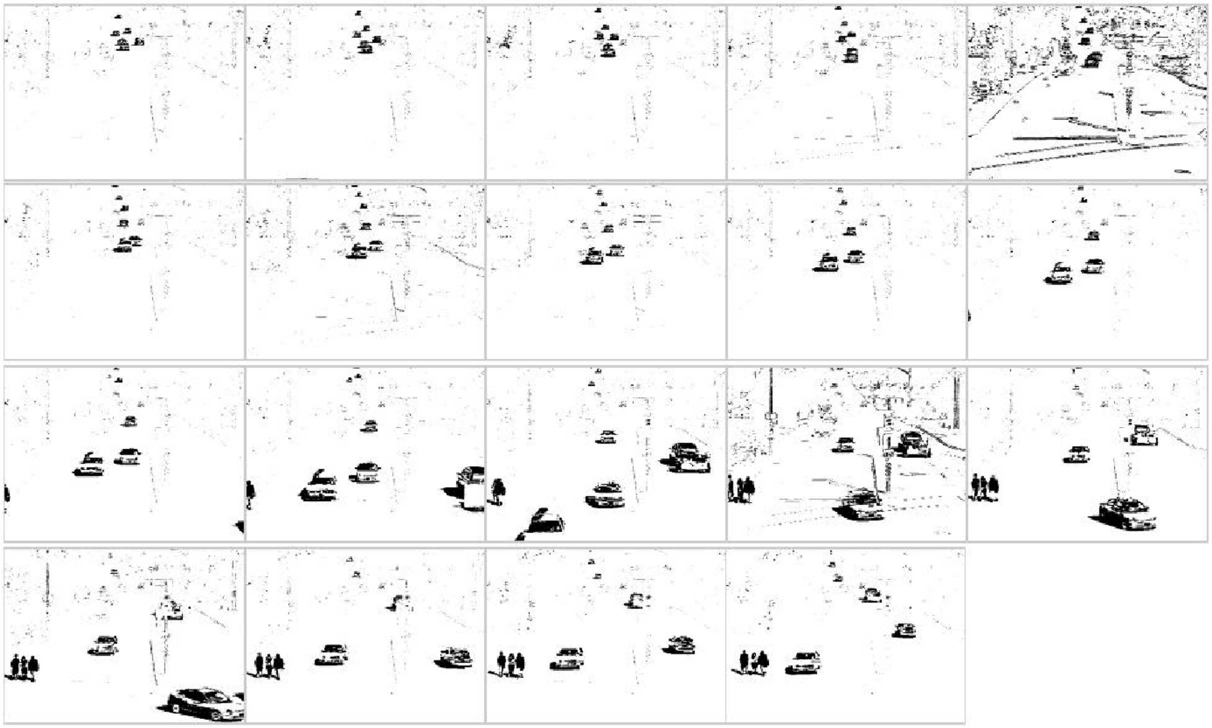


Fig. 5. Motion maps for the video sequence of Fig. 1, obtained by patched adaptive background approach at $Z° = 0.05$.

Step 6°. Replace in the temporary background the pixels identified by the correction mask as the moving ones with values generated at step 5°. Assign the resulting frame to be actual background for the selected video sequence:

$$\mathrm{TB}° = \begin{cases} R(i,j) & \text{if } \; C(i,j,t=1) = 1 \;\; \text{(moving pixels);} \\ B(i,j,t=1) & \text{if } \; C(i,j,t=1) = 0 \;\; \text{(static pixels);} \end{cases}$$

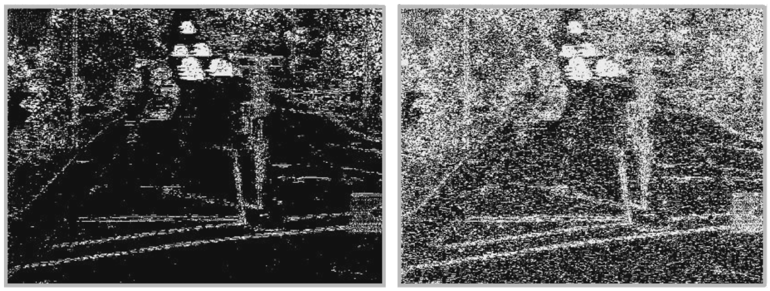
Step 7°. Complete motion detection following the steps 2 through 6 of the traditional adaptive background differentiation while employing as a background model, TB, the one resulting from step 6°: TB = TB°. Here it is seen that the modified algorithm replaces step 1 of the initial algorithm with steps 1° through 7° of the above, therefore providing patch-correction of TB model and upgrading it to corrected version TB°, which does not contain moving pixels and therefore eliminates the source of the persisting ghosts.

The memory bank here has been built by taking in the frames immediately following the initial frame of the sequence. For simplicity, the whole sequence of 19 frames starting from the second one was included into the memory bank of the experiment described in this article. This is considered to be appropriate for the mode of operation in which the pre-recorded frame sequences undergo processing, i.e. when collection of the data and its processing are separated in time. However in a real-time operation mode only the frames preceding in time are available to serve as memory bank content. In this case the restriction is to be applied as to the scene representation, namely, the frames stored in the memory bank need to represent the same scene and be connected to the same background model as the analyzed sequence.

EXPERIMENTAL VALIDATION

Experimental implementation of the suggested approach starts from generating the correction mask according to steps 1° through 3°. Correction masks for various threshold levels $Z°$ are presented in Fig. 3a–3d.

The result of respective patching of the background for the same $Z°$ values is shown in Fig. 4. The final result of the application of the patch-corrected background differentiation is presented in Fig. 5 for the case of $Z° = 0.05$, which represents the lowest possible value for the given level of camera stability. Com-



(a) $Z^{\circ} = 0.025$ (b) $Z^{\circ} = 0.010$

Fig. 6. Noise increase in the correction masks at $Z^{\circ} < 0.05$: (a) $Z^{\circ} = 0.025$; (b) $Z^{\circ} = 0.010$.

paring Fig. 5 and Fig. 2, it is seen that the main goal is achieved, namely, the efficient elimination of the persisting ghosts from the moving maps when patch-corrected background differentiation is applied. Persisting ghosts, present in Fig. 2 motion maps have been successfully eliminated, as shown in Fig. 5. This demonstrates that the correction of the initial mask based on extrinsic information retrieved from the memory bank does allow restoring the background model to remove the moving pixels that it contains, which gave rise to persisting ghosts. The patch-corrected background differentiation therefore becomes ghosts free in a sense that the persisting ghosts (such as those in Fig. 2) do not occur due to elimination of its very source—the moving pixels from the background model. Some residues of the ghosts still can be detected in Fig. 5 and these are due to the threshold limitation stemming from the non-stationarity of the camera in this sequence. Logically, the improvement here can be achieved via stable camera positioning, which will be explored in future work. However, in realistic outdoor environments, such as in the case of traffic control considered here, such instability of the camera is very likely to occur as it can be associated with the wind induced movements of the supporting post as well as structural vibrations. From Fig. 3 it is seen that lowering the threshold has twofold effects: from one hand, the areas requiring correction in background model (white pixels in Fig. 3) are expanded. On the other hand, the mask also becomes noisier, thus increasing the risk to introduce noise in the motion maps.

Comparing the masks and respective corrected background models between Fig. 3 and Fig. 4, it is seen that a reduction of Z° allows more complete elimination of the moving pixels from the background model, which are the source of the ghosts in the final motion maps. In terms of completeness of moving pixels elimination, it is seen from Fig. 4d that even lower thresholds are preferable.

However (as seen in Fig. 6) the masks obtained for $Z^{\circ} < 0.05$ exhibit much stronger noise level (as compared to the masks in Fig. 3). Among the two masks shown in Fig. 6, the mask (b) for $Z^{\circ} = 0.010$, appears to be extremely noisy, while the mask (a) for $Z^{\circ} = 0.025$, may still be acceptable despite the increased noise level. Taking into account the fact that apparent noise is generated from objects being detected as moving, while several of them are steady by nature, such as posts and lines on the roads, it can be inferred that a large part of the noise originates from the slightly non-stationary configuration of the camera, as was mentioned above (specifically, here the camera was held by the hands of the operator during recording). This allows to conclude that stationarity of the camera may permit further reduction of Z°, when no non-stationary noise is present and therefore further improvement of moving pixels elimination can be achieved via a reduction of Z° thus achieving better discrimination of the moving pixels against stationary ones with the proposed background model of the patched background differentiation algorithm. In Fig. 7 the best possible (under given stationarity of the camera conditions used in this experiment) background model is shown when derived with $Z^{\circ} = 0.025$. The resulting motion maps for) $Z^{\circ} = 0.025$ are shown in Fig. 8.

Fig. 7. Deep elimination of the moving pixels at $Z^{\circ} = 0.025$. Secondary ghost pixels start to appear as background noise.

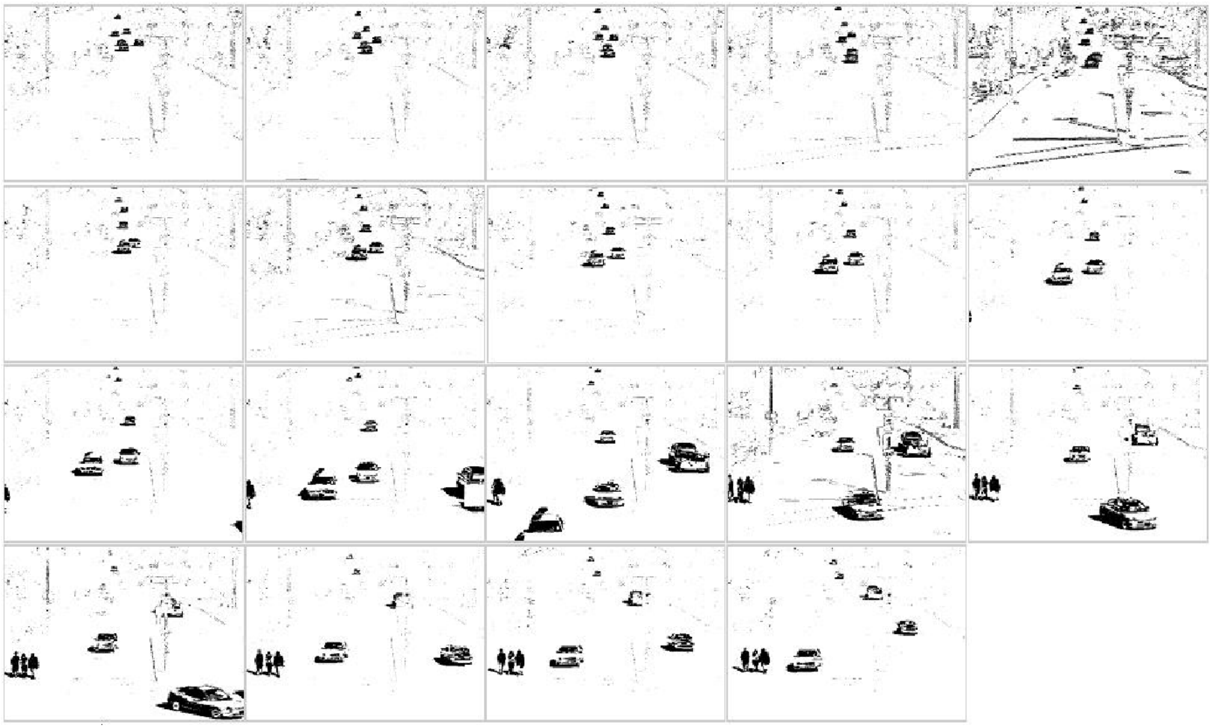


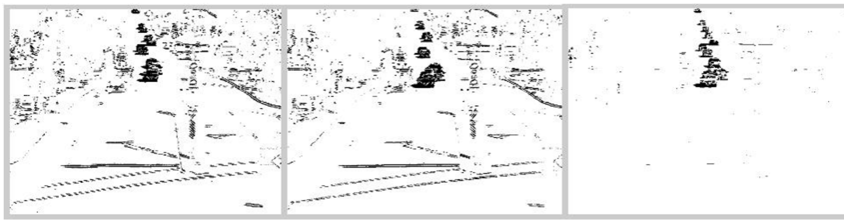
Fig. 8. Motion maps for the video sequence of Fig. 1, obtained by patched adaptive background approach at $Z^{\circ} = 0.025$.

Further reduction of Z°, i.e. below 0.025, leads to occurrence of secondary ghost-pixels from the memory frames. Some of those secondary ghost-pixels already are detectable in Fig. 7, however they are still perceived as a background noise. This effect becomes stronger when dropping Z° down to 0.01 and 0.001 levels, which then results in unacceptably noisy motion maps. The Z° level of 0.025 for a given video sequence appears to provide the best trade off between the elimination of the moving pixels from the background and secondary ghost prevention. The trade off in performance between ghosts elimination and non-stationary noise level for two masks,—namely, (a) the one shown in Fig. 3.d for $Z^{\circ} = 0.05$ and (b) another one from Fig. 6.a for $Z^{\circ} = 0.025$—is seen when comparing the respective motion maps of Fig. 5 and that of Fig. 8. The stronger ghosts in Fig. 5 lead to lower noise level as compared to that in Fig. 8. It is also worth noting that strongest difference in noise level between maps of Fig. 5 and that of Fig. 8 takes place for the initial maps, i.e. from $M(t = 1)$ to about $M(t = 7)$.

In this regard, later maps in the series, such as those from $M(t = 15)$ through $M(t = 19)$ exhibit almost the same noise level while keeping the same difference in the strength of the ghosts, which contributes in the favor of using lower Z° values even for non-stationary camera conditions.

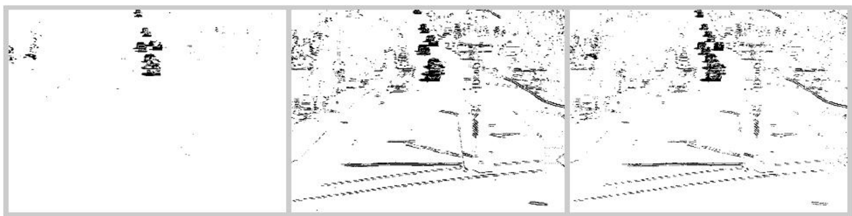
NOISY GHOSTS ELIMINATION

Apart from the persisting ghosts described above, there are more widely spread noisy ghosts caused by camera instabilities. The origin of such instabilities for traffic control applications can be the wind induced vibrations of the post or structure supporting the camera. In surveillance applications such camera's movements can be due to oscillations occurring on the pan-tilt device. In these cases the slight variations in camera's position during video acquisition result in all objects to appear as moving, as it is shown in Fig. 9a and Fig. 9b: Figure 9a presents the temporal differentiation motion map between frames #5 and #6, while in Fig. 9b—the motion map is computed between frames #6 and #7. These are two consecutive temporal differentiation motion maps. Significant noisy ghosts in these motion maps are due to motion of the camera rather than of the objects. The proposed approach to eliminate these ghosts assumes that the camera motion does not vary significantly between neighboring motion maps. Therefore such noisy ghost patterns should be sufficiently similar to one another. In this case, the derivative of the motion maps, that is a second derivative applied over the temporal differentiation, should efficiently eliminate the ghosts created by cam-



(a) (b) (c)

Fig. 9. Demonstration of the proposed technique of 2-nd derivative in temporal differentiation: motion map (c) is obtained by differentiating between motion maps (a) and (b). Significant reduction of noisy ghosts is observed.



(a) (b) (c)

Fig. 10. Trade off in the suggested technique of 2-nd derivative in temporal differentiation: motion map (c) is obtained by differentiating between motion maps (a) and (b). Noise transfer is observed when the original motion maps differ strongly enough in terms of noise patterns.

era motion. The resulting motion map after ghost patterns are removed from the initial motion map is presented in Fig. 9c for the two consecutive motion maps. The latter represents the difference between the maps in Fig. 9a and Fig. 9b. However, the significant improvement observed in Fig. 9c comes at a cost. Given that most of the actually moving objects exhibit stronger positional shift than that produced by camera instability, a split of the moving objects is observed on some of them. The change in position for these objects between the initial and second motion maps are not zeroed out by the second derivative.

Another trade off comes at the border of the map sequence, where the camera stability is actually passing from being stationary to becoming non-stationary or vice versa. This situation is illustrated in Fig. 10, where (a) and (b) represent the motion maps for the frame pairs #4,#5 and #5,#6 respectively. It is seen that the pattern of the ghosts in these two consecutive motion maps differ significantly, reflecting the fact of strong change in the camera stability. Differentiation between these two motion maps result in the map presented in Fig. 10c in which compensation of ghosts is very weak. Therefore at the transition points where the camera stationarity varies, the use of motion maps differentiation looses efficiency for ghost's elimination. Therefore this approach is suitable for video acquired with a sufficiently high frame rate. At a high frame rate, the transitions between various levels of vibration of the camera are smoother as they get distributed over a larger number of frames. The proposed method then performs better as two successive motion maps exhibit ghosts that are related to camera motion which are of similar magnitude. In this case the operational range of the proposed ghost elimination algorithm is, thus preserving its positive impact on the video surveillance system performance in a wider set of conditions.

CONCLUSIONS

A patched correction technique for the background differentiation algorithm involved in motion detection is proposed which makes use of frames memory bank to eliminate moving pixels from the background model. Slightly non-stationary camera conditions are considered. An experimental validation of the effectiveness of the proposed approach suggests an optimum threshold value for the correction mask computation. This optimum value defines the trade off between ensuring elimination of the moving pixels from the background, while preventing the appearance of secondary ghost pixels. When considering a non-stationary camera, as in these experiments, the optimal threshold value Z° for the correction mask is found to be in the range $0.025 \le Z^{\circ} \le 0.050$. Improving stability of the camera allows pursuing deeper Z° values for more complete elimination of the moving pixels from background model without risking secondary ghosts' noise, which results in cleaner background models. Additionally, an approach is proposed for the reduction of ghosts related to the camera pose variations. It builds on a second differentiation between neighboring motion maps. The trade off associated with such operation, which is equivalent to a second temporal derivative between frames, is splitting couples of the actually moving objects in the resulting maps and a lower performance on video recorded a low frame rate.

REFERENCES

1. Karmann, K. and von Brandt, A., Moving Object Recognition Using an Adaptive Background Memory, in *Time-Varying Image Processing and Moving Object Recognition*, 2, Cappellini, V., Ed., Amsterdam, the Netherlands: Elsevier, 1990.
2. Liu,Y. and Payeur, P., Robust Image-Based Detection of Activity for Traffic Control, *Can. J. Elect. & Comp. Eng.*, 2003, vol. 28, no. 2, pp. 63–67.
3. Payeur, P. and Liu,Y., Video Traffic Monitoring for Flow Optimization and Pollution Reduction, *Proc. 2nd IEEE/AEI Int. Workshop Adv. Environm. Monitoring Techn.*, Como, Italy, 24–25 July 2003, pp. 53–58.
4. Elhabian, S.Y., El-Sayed, K.M., and Ahmed, S.H., Moving Object Detection in Spatial Domain Using Background Removal Techniques—State-of-Art, *Recent Patents on Comp. Science*, 2008, vol. 1, pp.32–54.
5. Hu, F.Y., Zhang, Y.N., Wong, H.S., and Pan, Q., A Robust Moving Object Detection Approach, *Image and Signal Processing, 2008. CISP '08. Congress on*, 2008, pp. 145–149.
6. Mittal, A. and Paragios, N., Motion-Based Background Subtraction Using Adaptive Kernel Density Estimation, *Proc. IEEE Comp. Soc. Conf. on Comp. Vis. & Pattern Recognition (CVPR'04)*, 2004, vol. 2, pp. 302–309.
7. Vibha, L., Venkatesha, M., Prasanth, G., Rao, N., Suhas, N., Deepa Shenoy, P., Venugopal, K.R., and Patnaik, L.M., Moving Vehicle Identification using Background Registration Technique for Traffic Surveillance, *Proc. Int. Multi Conf. Eng. & Computer Sci. IMECS 2008*, 19–21 March, 2008, Hong Kong, Vol. 1.