NIGHT–TIME OUTDOOR SURVEILLANCE WITH MOBILE CAMERAS

Ferran Diego[1,3], Georgios D. Evangelidis[2] and Joan Serrat[1]

[1]*Dept. Ciencies Computacio and Computer Vision Center, Universitat Autonoma de Barcelona, Barcelona, Spain*

[2]*Department of Computer Engineering & Informatics,University of Patras, Rio-Patras, Greece*

[3]*HCI, University of Heidelberg, Heidelberg, Germany*

Keywords: Video surveillance, Video synchronization, Video alignment, Change detection.

Abstract: This paper addresses the problem of video surveillance by mobile cameras. We present a method that allows online change detection in night–time outdoor surveillance. Because of the camera movement, *background* frames are not available and must be ”localized“ in former sequences and registered with the current frames. To this end, we propose a Frame Localization And Registration (FLAR) approach that solves the problem efficiently. Frames of former sequences define a database which is queried by current frames in turn. To quickly retrieve nearest neighbors, database is indexed through a visual dictionary method based on the SURF descriptor. Furthermore, the frame localization is benefited by a temporal filter that exploits the temporal coherence of videos. Next, the recently proposed ECC alignment scheme is used to spatially register the synchronized frames. Finally, change detection methods apply to aligned frames in order to mark suspicious areas. Experiments with real night sequences recorded by in-vehicle cameras demonstrate the performance of the proposed method and verify its efficiency and effectiveness against other methods.

1 INTRODUCTION

Lately, visual–surveillance systems have attracted increasing interest from urban and building security, military related field and security and video patrolling systems. Such systems aim at detecting potential suspicious items or signs of intrusion, and consequently generate a warning to a human operator. This detection mainly consists of identifying changes between images of the same scene but temporally separated. Most change detection methods proposed in the literature deal with stationary cameras as surveyed in (Radke et al., 2005). This amounts to detect differences using a stationary background thus reflecting minor applicability when multiple cameras are needed. This drawback can be overcome by mobile cameras, but the problem becomes more challenging due to the non-stationary background and varying ambient illumination.

The latter scenario is what we consider in this paper. Specifically, we present a method that helps the video analyst to robustly detect potential and suspicious signs of intrusion by vehicles that repeatedly patrol sensitive areas and private buildings at night–time. This detection cannot rely on specific classifiers mainly due to the following factors: (1) the video quality can be significantly degraded at night–time and (2) these signs may be random or stationary anomalies (*e.g.* intruder, suspicious suitcase), with arbitrary shapes, color or texture. To this end, we propose an efficient framework to detect potential anomalies exploiting similarities occurred by repeatedly patrolling the same ride. This consists of comparing a pair of video sequences recorded by a forward–facing camera attached to the windscreen of the vehicle whose view is what the driver sees. Hence, signs of intrusion or missing objects occurred in the interim of successive rounds can be detected by background subtraction methods. This obviously requires the spatio–temporal alignment of the current sequence with the one captured during the previous round, i.e. the video synchronization and the spatial registration of corresponding frames.

Video synchronization algorithms estimate the temporal relation between two sequences once they have been acquired. However, our goal is to online detect changes at a reasonable rate. Thus, instead of solving an offline global optimization problem, the proposed framework counts on a Frame Localization And Registration (FLAR) scheme. In short, given

Diego F., Evangelidis G. and Serrat J. (2012).
NIGHT–TIME OUTDOOR SURVEILLANCE WITH MOBILE CAMERAS.
In *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods*, pages 365-371
DOI: 10.5220/0003758103650371

each newly acquired frame, we *temporally localize* it against the *background sequence* of the previous ride. This aims in other words at assigning each current frame to a background frame so that their viewpoints are the closest ones. Since efficiency is of major importance in online solutions, the extraction of the corresponding frame relies on an image retrieval scheme based on the SURF descriptor (Bay et al., 2008). A temporal filter applies to the outcome of the retrieval task in order to handle false positives (outliers). Then, we have to spatially register the corresponding frames into the same coordinate system. As the video acquisition takes place at different times, the appearance of corresponding frames varies. To cope with such variations, we adopt the recently proposed ECC image alignment scheme (Evangelidis and Psarakis, 2008) that offers the desired robustness. As a final step, different metrics that count on image differences are applied to detect changes and mark areas of interest.

The contribution of this paper is summarized as follows: 1) A challenging case of night–time outdoor surveillance by mobile cameras is investigated. 2) The proposed FLAR scheme reflects a solution for online surveillance instead of postprocessing. 3) It incorporates efficient tasks that allows us to envision a real-time execution in GPU-based environment. 4) The desired invariance to the motion style of surveillance vehicle (speed, backward motion) is fulfilled.

1.1 Related Work

The challenging problem of detecting changes between videos acquired by mobile cameras at different times is considerably less tackled than the case of stationary cameras (Radke et al., 2005). Marcenaro *et al.* (Marcenaro et al., 2002) proposed an outdoor–surveillance based on fixed and pan/tilt mobile cameras that exceeds the limitations of the fixed camera which monitors the entire scene, but the position of the mobile camera must be known anytime. Primdahl *et al.* (Primdahl et al., 2005) presented a method for automatic navigation of cameras in a specific, well–defined corridor. Sand and Teller (Sand and Teller, 2004) proposed a video matching scheme for two sequences recorded by moving cameras following nearly identical trajectories. Although it allows pixel–wise comparisons to detect differences, its key limitation is the computational time of computing a robust image–alignment for several possible pairs of corresponding frames. To make it efficient, Kong *et al.* (Kong et al., 2010) temporally aligned sequences using GPS data only and detect abandoned suspicious objects via inter–sequence homographies. In contrast, Soiban *et al.* (Soibam et al., 2009) and Haberdar and Shah (Haberdar, 2010) found manually the corresponding frame in the first video for each observed frame of the second one. Finally, Diego *et al.* (Diego et al., 2011) proposed a video alignment framework based on fusing image–based and GPS observations to spot differences between sequences taken at different times and by independently moving cameras, while Chakravarty *et al.* (Chakravarty et al., 2007) presented a mobile robot capable of repeating a manually trained route that detect any visual anomalies using stereo–based algorithm; these anomalies are subsequently tracked using a particle filter.

The rest of this paper is organized as follows: Section 2 describes the whole framework and specifically, subsection 2.1 presents the frame localization approach, while the spatial registration and the change detection tasks are discussed in subsections 2.2 and 2.3 respectively. Experiments to validate the proposed algorithm are presented in Section 3 and results are discussed. Finally, in Section 4, the main conclusions are drawn.

2 FRAME LOCALIZATION AND REGISTRATION

Suppose we are given two video sequences represented as $\mathcal{I}^r = \{I^r_m(\hat{\mathbf{x}})\}_{m=1}^M$ and $\mathcal{I}^c = \{I^c_n(\mathbf{x})\}_{n=1}^N$, being M,N their number of frames and $\hat{\mathbf{x}} = [\hat{x},\hat{y}]^t$, $\mathbf{x} = [x,y]^t$ their spatial coordinates respectively. The former denotes the *reference* or *background*, taken in a previous ride, whereas the latter is the *current* sequence being recorded in the current ride following a similar trajectory. Then, the anomalies occurred in the meanwhile between successive rounds can be detected by matching and comparing the two sequences. That is, the proper thresholding of image differences between spatio-temporally aligned sequences allows the detection of changes.

To solve the above defined problem we propose a Frame Localization And Registration (FLAR) framework that is shown in Figure 1. The only assumption we make is that the vehicles follow a similar, approximately coincident, route. The most likely frame of a previous ride is extracted for each newly acquired frame in the current ride (localization step). This implies a challenging task because of the independently moving cameras and the non-coincident trajectories. As a result, the speed and the position of the cameras vary, while the ambient illumination can be different. A few video alignment approaches (Sand and Teller, 2004; Liu et al., 2008; Diego et al., 2011) could be adjusted to our problem. However, none of them is able to estimate the frame correspondence during

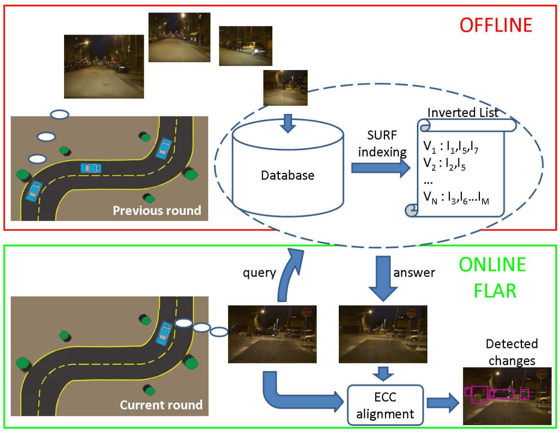

Figure 1: FLAR system for video surveillance with mobile cameras.

the acquisition of the current sequence due to their complexity. Therefore, we propose an efficient on–line video synchronization algorithm that relies on an image retrieval scheme based on the SURF descriptor (Bay et al., 2008) and a temporal filter. This essentially assigns the n^{th} current frame to the reference index t_n with $t_n \in [1, M]$, thereby providing the desired invariance to the motion style of cameras. Once the correspondence pair (n, t_n) has been found, the dense alignment of the assigned frames is required (registration step) in order to compare them pixel–wise. This kind of comparison is necessary for our endmost goal, that is, the identification of regions that changed in the interim between the records.

2.1 Image–retrieval Scheme

The image–retrieval scheme based on SURF descriptors aims to efficiently evaluate a confidence matrix that measures the similarity of all possible pairs, thus allowing the association of the current frame to the most similar reference frame. Our implementation resembles (Sivic and Zisserman, 2009) but we disable the vector quantization step and use only the inverted file. In short, we run the SURF algorithm to localize keypoints and describe their neighborhood in all background frames. A visual dictionary is then learned and an inverted index list is built as shown in Fig. 1. Note that all this can be done off-line without having the current sequence at our disposal. Given now a current frame of the new ride, we extract its SURF descriptors and look for their closest visual words, thus voting for the assigned reference frames. To ignore very frequent visual words, we enable the inverse–document–frequency (IDF) weighting scheme (Sivic and Zisserman, 2009).

2.1.1 Temporal Filtering

To extract the time mapping result, for each observed frame one could simply choose the reference index with the maximum confidence value. However, it might return an erroneous synchronization signal with sharp changes due to isolated points. To avoid sharp transitions, we choose for any query frame the maximum reference index subject to the constraint that lies in a tolerance interval. The latter is defined around the reference index assigned to the previous query frame. By recalling that t_n is the correspondence of the n^{th}, $[t_n - 10, t_n + 10]$ is the tolerance interval of the $(n+1)^{th}$ current frame for our experiments (the value of 10 can vary with the application). In order to obtain a far smoother signal, we propose the use of a filter that applies to the signal t_n. Specifically, such a filter can be described by the standard *difference equation* (Lathi, 1998)

$$T_n = \sum_{i=0}^{K} b_i t_n - \sum_{j=1}^{L} a_j T_{n-j} \tag{1}$$

where T_n defines its output. In general, it constitutes an Infinite Impulse Response (IIR) filter, but when $a_j = 0$, it turns into a Finite Impulse Response (FIR) filter of order K (Lathi, 1998). It is important to note that this filter is a causal system and the current output depends only on previous input and output values, being capable for online and real-time solutions. Both type of filters were tested using $K = L = 3$, $b_0 = 0.4$, $b_1 = 0.3$, $b_2 = 0.2$, $b_3 = 0.1$ and $a_1 = 4$, $a_2 = -2$, $a_3 = -1$. In either case, these values establish a low-pass filter. IIR provides smoother results because of its higher, theoretically infinite, order. On the other hand, FIR deals better with peaks (outliers) due to its finite order. The frequency response of the filters and the Discrete Fourier Transform (DFT) (Lathi, 1998) magnitude of the output signals are shown in Fig. 2. The input is the time mapping sequence resulted when the proposed method applies to a real sequence. The ground smooth signal obtained by postprocessing (curve fitting) is given for comparison. Although both filters behave similarly in low frequencies, IIR output is more close to the ground signal in high frequencies.

2.2 Spatial Alignment

In order to obtain accurate alignment between a reference frame $I^r(\hat{\mathbf{x}})$ and the current observed frame $I^c(\mathbf{x})$, we propose the use of a recently introduced algorithm that is called ECC algorithm (Evangelidis and Psarakis, 2008). This scheme seems to be robust in noise while at the same time is insensitive to global illumination changes. The algorithm uses

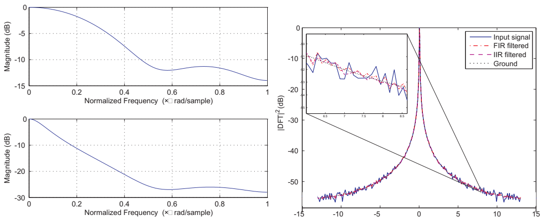

Figure 2: *Left:* Frequency response of the (top) FIR and (bottom) IIR filter. *Right:* The DFT of the input, the outputs and the ground signal (video rate: 25fps).

an enhanced version of the correlation coefficient as an objective function and the goal is its maximization through an iterative scheme.

Let us suppose that the warp $W(\mathbf{x};\mathbf{p})$ is a 2D mapping based on the standard homography model with eight parameters (Szeliski, 2010), i.e. $\hat{\mathbf{x}} = W(\mathbf{x};\mathbf{p})$, that provides dense correspondences. Then, ECC algorithm tries to estimate the warp so that the observed and the warped reference images are similar. In other words, it solves the following maximization problem

$$\max_{\mathbf{p}} \frac{\mathbf{i}_c{}^t \mathbf{i}_r(\mathbf{p})}{\|\mathbf{i}_c\| \|\mathbf{i}_r(\mathbf{p})\|} . \tag{2}$$

where $\mathbf{i}_c$ and $\mathbf{i}_r(p)$ are the zero-mean vectorized forms of images $I^c(x)$ and $I^r(W(\mathbf{x};\mathbf{p}))$ respectively. Since the above maximization problem is highly non-linear, the solution of a sequence of secondary problems that follow a closed form solution is proposed in (Evangelidis and Psarakis, 2008). By considering the update rule $\mathbf{p} = \mathbf{p}_0 + \Delta\mathbf{p}$, the vector $\mathbf{i}_r(\mathbf{p})$ can be approximated by $\mathbf{i}_r(\mathbf{p}) \simeq \mathbf{i}_r(\mathbf{p}_0) + \mathrm{J}\Delta\mathbf{p}$ using the first-order Taylor expansion formula, where J is the Jacobian of $\mathbf{i}_r(\mathbf{p})$ with respect to $\mathbf{p}$ evaluated at $\mathbf{p}_0$ (see (Evangelidis and Psarakis, 2008) for details). Although after linearization the objective function remains nonlinear in $\Delta\mathbf{p}$, it has been proved that the optimum correction vector obeys the following closed form solution

$$\Delta\mathbf{p} = (\mathrm{J}^t\mathrm{J})^{-1}\mathrm{J}^t \left\{\lambda\bar{\mathbf{i}}_o - \bar{\mathbf{i}}_r(\tilde{\mathbf{p}})\right\} , \tag{3}$$

with λ being

$$\lambda = \frac{\bar{\mathbf{i}}^t_{\tilde{\mathbf{p}}} \mathrm{P}_\mathrm{J} \bar{\mathbf{i}}_{\tilde{\mathbf{p}}}}{\mathbf{i}^t_q \mathrm{P}_\mathrm{J} \mathbf{i}_{\tilde{\mathbf{p}}}} , \tag{4}$$

where $\mathrm{P}_\mathrm{J} = \mathrm{I} - \mathrm{J}(\mathrm{J}^t\mathrm{J})^{-1}\mathrm{J}^t$ is an orthogonal projection operator and I the identity matrix. Finally, by iteratively following the above parameter update rule, we can obtain an acceptable solution by setting a stopping criterion or fixing the number of iterations. Note that the complexity of this scheme is $O(N_i N_p^2)$ per iteration, where N_p is the number of parameters and N_i is the number of pixels.

2.3 Change Detection

Although we present a video surveillance algorithm, we do not focus on change detection since this subject has been extensively studied. A nice survey for image detection algorithms can be found in (Radke et al., 2005). Hence, since FLAR provides the corresponding background frame appropriately warped, we use known methods to detect changes between registered images. Specifically, we use the Simple Differencing (SD) method by thresholding the image differences, a Mimimum Description Length (MDL) model to classify changed and unchanged regions and a statistical method that considers a Gaussian model for the noise (GN) (for details see (Radke et al., 2005)). All these methods return a binary image (mask) at which we apply trivial morphological operations in order to locate bounding boxes in the image of interest.

3 EXPERIMENTAL RESULTS

In this section, we present qualitative and quantitative results to validate the proposed approach. Specifically, we compare the performance of different counterparts of the proposed algorithm with the most related works (Diego et al., 2011; Liu et al., 2008; Yang et al., 2007). The evaluation counts on experimenting with six real video sequence pairs recorded by in-vehicle cameras, whose trajectories are approximately coincident. Although we aim at registering nighttime sequences, we consider essential to also test the algorithms with daylight sequences. To this end, we used three sequences of each class denoted as *Night1*, *Night2* and *Night3* (Serrat et al., 2007), and as *Day1*, *Day2* and *Day3* (Kong et al., 2010) respectively. Their alignment implies a quite challenging task, since the speed of vehicles varies. The average length of night sequences is 2500 frames and the spatial resolution is 720 × 540 pixels, whereas daylight sequences are shorter in both space and time (200 frames of size 512 × 384 pixels).

3.1 Synchronization Evaluation

In this section, we evaluate the performance of temporally localize each newly acquired frame during the current ride against the background sequence of the previous ride. To properly assess the quality of the results, we have manually annotated the ground–truth for these datasets, *i.e.* a narrow reference interval $[l_n, u_n]$ that each current frame must correspond to; the length of these intervals is 3 frames on average. Similar to (Diego et al., 2011), the synchronization error

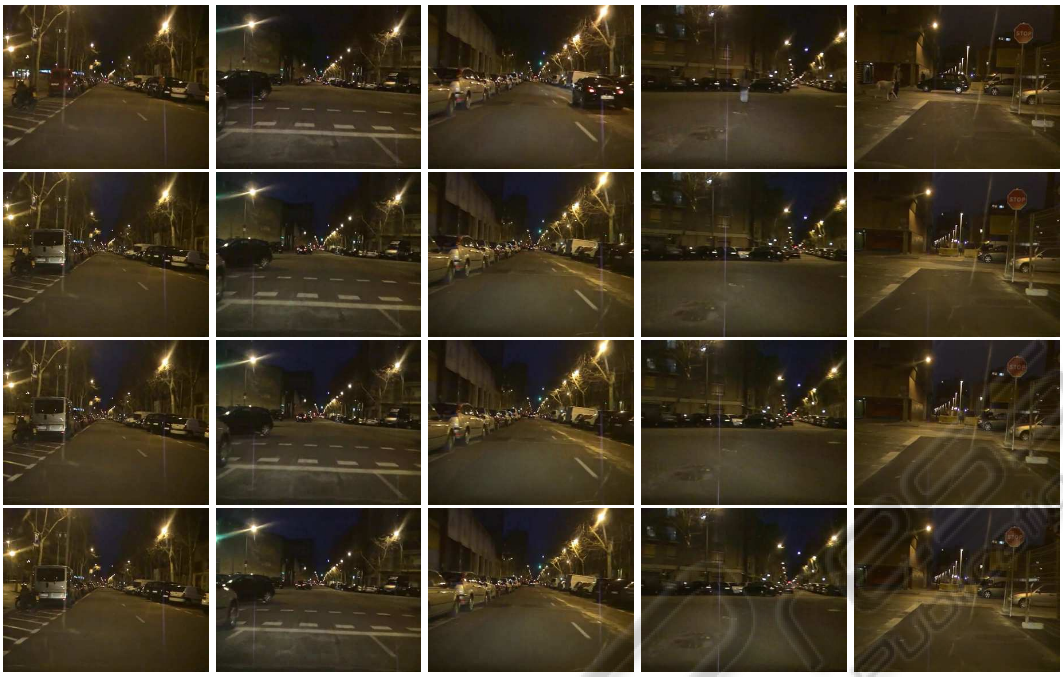

Figure 3: (First row) Query (current) frames of *Night1* sequence and synchronization results obtained by (second row) exhaustive search, (third row) SIFT-based retrieval and (fourth row) SURF-based retrieval.

for a candidate pair (n,t_n) is defined as

$$\mathrm{err}(t_n) = \begin{cases} 0 & \text{if } l_n \le t_n \le u_n \\ \min\left(|l_n - t_n|, |u_n - t_n|\right) & \text{otherwise} \end{cases} \quad (5)$$

The performance of synchronization is quantified through the percentage $1 - \sum_i^N (err(t_n) > \varepsilon)/N$ for $\varepsilon = 0, 1$.

An an exhaustive search scheme given a query frame, we can succeed in temporal matching by obtaining a short-list (i.e. top-10) of background frames using the image–appearance model proposed in (Diego et al., 2011). Then, a spatial coherence step using ECC algorithm re-ranks the list w.r.t. the correlation coefficient, thus emerging the closest frame. In the context of retrieval, we also try our scheme by only changing the SURF descriptor with the SIFT one (Lowe, 2004).

Table. 1 shows the synchronization performance achieved by these three methods. We provide results for $\varepsilon = 0$ and $\varepsilon = 1$ to show the error variance. We observe that the SURF–based method achieves higher synchronization scores than the two other methods across all sequences. It is important to note that the Frame Localization (FL) based on SURF or SIFT descriptors accurately discriminates the background frame by just retrieving the best neighbor. IIR filter provides slightly better scores with both descriptors. However, the contribution of SURF descriptor instead of SIFT is clearly evident especially for night–time sequences. Specifically, SURF–FL (SURF–based FL) outperforms SIFT–FL (SIFT–based FL) by 6% on average while the proposed scheme achieves a 8% better score than that of the exhaustive method. Note that we do not count on geometric constraints, since we aim at investigating the performance of the net algorithm. However, it is obvious that SURF–FL scheme would be benefitted by such constraints. Note also that SURF-FL and SIFT-FL need 0.88 and 2.8 secs respectively to synchronize a night frame.

3.2 Alignment and Detection Assessment

To assess the alignment, we use a color *RGB* representation, where the *G* channel of the current frame has been replaced by the warped *G* channel of the background corresponding frame. This way, changes are marked by green and pink colors. In Fig. 3 the corresponding frames obtained by the synchronization methods are shown for various night frames including challenging cases. Given the results of the proposed method (Fig. 3 (bottom)), Fig. 4 presents alignment instances obtained by the SIFT-flow algorithm, the Generalized Dual-Bootstrap version of the ICP algorithm (Yang et al., 2007) (GDB-ICP) and the ECC scheme. Note that the goal of SIFT-flow is a pixel–wise alignment instead of estimating a global geometric transformation as ECC and GDB-ICP do. All

Table 1: Synchronization scores (%) obtained by the proposed methods and the competitors for two values of error tolerance ε. Symbol ”–“ means that the exhaustive method totally fails for *Day1* due to repeated patterns in frames.

		Synchronization scores (ε = 0\ε = 1)						
		Night1	*Night2*	*Night3*	*Day1*	*Day2*	*Day3*	Average
Exhaustive search		71.5\84.5	**61.4\78.8**	68.9\83.8	–	93.2\98.6	85.0\99.3	76.0\89.0
SIFT–FL	FIR	67.5\82.9	48.7\68.6	66.9\83.1	70.0\85.0	99.3\100	92.5\96.6	74.2\86.0
	IIR	71.8\86.7	52.2\70.7	77.1\88.8	74.0\93.5	**100\100**	95.2\100	78.4\90.0
SURF–FL	FIR	72.6\86.3	53.4\71	73.6\87.4	74.0\90.5	99.3\100	88.4\95.2	76.9\88.4
	IIR	**78.8\90.6**	60.6\76.6	**82.6\92.8**	**96.5\99.5**	**100\100**	**100\100**	**86.4\93.3**

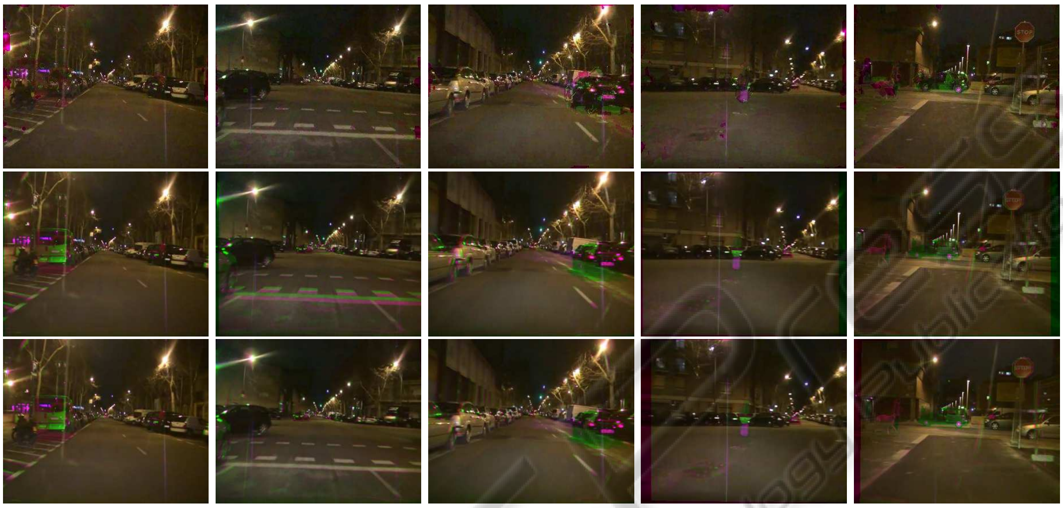

Figure 4: Alignment instances in negative color for (first row) SIFT-flow, (second row) GDB-ICP and (third row) ECC algorithm based on the frame pairs between the top and bottom row in Fig. 3.

algorithms behave quite well in the absence of occlusions. As we can see, however, when the scene contains objects visible only in the one sequence, SIFT-flow fails as it creates artifacts or disappears objects. This is probably because it works in a flow (local) basis. On the other hand, ECC and GDB-ICP achieve remarkable results despite the noise and the low information content, with GDB-ICP providing local misalignments in some of two of the depicted frames. The average registration time of half-size images is 29.2, 42.2 and 0.48 sec/frame for SIFT-flow, GDB-ICP and ECC algorithms respectively.

Detection results for SD, MDL and GN models are shown in Fig. 5. Instead of presenting binary masks, we use bounding boxes superimposed in query frames to annotate detected changes. An ”empty” bounding box means that something is missing compared to the background frame (see also the bottom row of Fig. 3). Otherwise, it may be due to local misalignment, different illumination and reflectance, shading etc. We observed that GN method provides slightly better result than MDL and SD methods. We must point out here that, normally, errors in alignment and detection do not happen in successive frames but randomly (see supplemental material). This is helpful for the video analyst who can ignore instant changes. The time required by SD method is meaningless. The complexity of the MDL and GN method is slightly higher, but not prohibitive for real–time applications.

Please refer to http://www.cvc.uab.es/∼fdiego/Surveillance/ for video results of the proposed method.

4 CONCLUSIONS

We presented a novel framework for helping a video analyst to robustly detect changes in night-time outdoor surveillance by mobile cameras. In order to avoid exhaustive cross-frame search of finding background frames, a Frame Localization And Registration (FLAR) is proposed to solve the problem efficiently. The frame localization builds upon retrieving the most similar background frame based on the SURF descriptor together with a temporal filtering applied to the retrieval results to handle outliers. Then, a recently proposed alignment scheme that overcomes appearance variations between frames acquired at different times is used to register the corresponding frames in space; thus applying a simple change

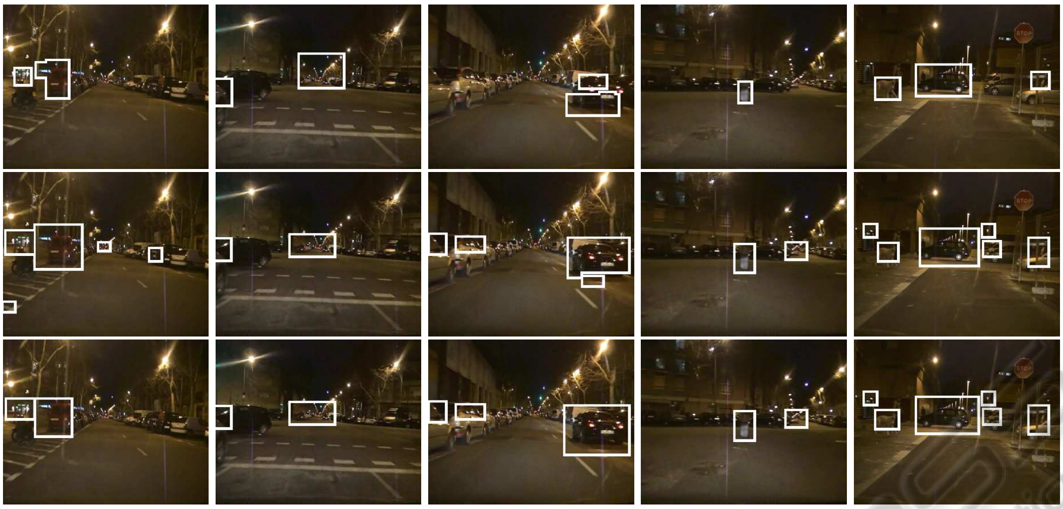

Figure 5: Change detection results using (top) SD, (middle) MDL and (bottom) GN method.

detection to aligned frames allows the detection of suspicious areas. Experiments with real night sequences recorded by in-vehicle cameras demonstrate the performance of the proposed method and verify its efficiency and effectiveness against other methods. Moreover, the ability of the proposed scheme to deal with daylight sequences was experimentally verified.

ACKNOWLEDGEMENTS

This work is supported by Spanish MICINN project TRA2011-29454-C03-01, and Consolider Ingenio 2010: MIPRCV (CSD200700018).

REFERENCES

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *CVIU*, 110(3):346–359.

Chakravarty, P., Zhang, A. M., Jarvis, R., and Kleeman, L. (2007). Anomaly detection and tracking for a patrolling robot. In *Australasian Conf. on Robotics and Automation*.

Diego, F., Ponsa, D., Serrat, J., and Lopez, A. (2011). Video alignment for change detection. *IEEE Trans. on Image Processing*, 20(7):1858 –1869.

Evangelidis, G. D. and Psarakis, E. Z. (2008). Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Trans. on PAMI*, 30(10):1858–1865.

Haberdar, H. (2010). Disparity map refinement for video based scene change detection using a mobile stereo camera platform. In *Proc. of ICPR*.

Kong, H., Audibert, J.-Y., and Ponce, J. (2010). Detecting abandoned objects with a moving camera. *IEEE Trans. on Image Processing*, 19(8):2201 –2210.

Lathi, P. (1998). *Signal Processing and Linear Systems*. Berkeley Cambridge Press.

Liu, C., Yuen, J., Torralba, A., and Freeman, W. T. (2008). Sift flow: dense correspondence across different scenes. In *Proc. of ECCV*.

Lowe, D. (2004). Distinctive image features from scale invariant keypoints. *IJCV*, 60(2):91–110.

Marcenaro, L., Marchesotti, L., and Regazzoni, C. (2002). A multi-resolution outdoor dual camera system for robust video-event metadata extraction. In *Proc. of the 5th Int. Conf. on Information Fusion*, volume 2, pages 1184 – 1189.

Primdahl, K., Katz, I., Feinstein, O., Mok, Y. L., Dahlkamp, H., Stavens, D., Montemerlo, M., and Thrun, S. (2005). Change detection from multiple camera images extended to non-stationary cameras. In *Proc. of Field and Service Robotics*.

Radke, R. J., Andra, S., Al-Kofahi, O., and Roysam, B. (2005). Image change detection algorithms: A systematic survey. *IEEE Trans. on Image Processing*, 14:294–307.

Sand, P. and Teller, S. (2004). Video matching. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 22(3):592–599.

Serrat, J., Diego, F., Lumbreras, F., and Àlvarez, J. (2007). Alignment of videos recorded from moving vehicles. In *Proc. of 14th Int. Conf. on Image Analysis and Processing*.

Sivic, J. and Zisserman, A. (2009). Efficient visual search of videos cast as text retrieval. *IEEE Trans. on PAMI*, 31(4):591–606.

Soibam, B., Shah, S. K., Chaudhry, A., and Eledath, J. (2009). Quantitative comparison of metrics for change detection for video patrolling. In *ICCV Workshop on Video-Oriented Object and Event Classification*.

Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Springer.

Yang, G., Stewart, C., Sofka, M., and Tsai, C.-L. (2007). Registration of challenging image pairs: Initialization, estimation, and decision. *IEEE Trans. on PAMI*, 29(11):1973–1989.