LaBGen-P: A Pixel-Level Stationary Background Generation Method Based on LaBGen

Benjamin Laugraud, Sébastien Piérard, and Marc Van Droogenbroeck
INTELSIG Laboratory, University of Liège, Belgium
{BLaugraud, Sebastien.Pierard, M.VanDroogenbroeck}@ulg.ac.be

***Abstract*—Estimating the stationary background of a video sequence is useful in many applications like surveillance, segmentation, compression, inpainting, privacy protection, and computational photography. To perform this task, we introduce the LaBGen-P method based on the principles of LaBGen and the conclusions drawn in the corresponding paper. It combines a pixel-wise median filter and a pixel selection mechanism based on a motion detection performed by the frame difference algorithm. By working with pixels instead of patches, as originally done in LaBGen, it avoids some discontinuities between different spatial areas and generates better visual results. In this paper, we describe the LaBGen-P method, study its performance on the sequences of the SBMnet dataset, and compare it to that of LaBGen and other methods on the same dataset. Both algorithms emerged as the best ones during the IEEE Scene Background Modeling Contest (SBMC) organized in 2016. However, as there is not yet a good understanding of the recommended metrics, and due to the small amount of video sequences provided with the corresponding ground truth, we have performed a subjective evaluation. More precisely, 35 human experts were asked to compare background images estimated by LaBGen-P and LaBGen, and select the best one. From these experiments, it turns out that the results of LaBGen-P are preferred for about two thirds of the video sequences. Note that we provide an open-source C++ implementation at http://www.telecom.ulg.ac.be/labgen.**

I. Introduction

The *stationary background generation problem* (also known as *stationary background estimation*, *reconstruction*, or *initialization problem*) consists in generating a unique image estimating the stationary background of an input video sequence acquired from a fixed viewpoint. In this context, the *stationary background* is defined as the set of elements that are motionless for the duration of the whole sequence (note that this definition excludes the elements subject to periodic movements). Generating an estimation of the background is helpful for many applications including surveillance, segmentation, compression, inpainting, privacy protection, and computational photography [1]. As an example, *background subtraction algorithms* (*BGS algorithms*), able to classify any pixel of a sequence as belonging to the background or not, could benefit from such an estimation to initialize their model [2].

One of the simplest and most intuitive background generation method applies a pixel-wise temporal median filter on the frames composing the input sequence (this method is referred to as the *median method* hereafter). However, this method fails to generate a correct estimation when a sequence is highly cluttered or, in other words, when the background is observable for less than half of the time. To cope with the problems related to complex scenes (*i.e.* subject to illumination changes, camera jitter, intermittent motions, etc), more advanced methods have been proposed in the literature, as detailed in [1].

Fig. 1. Background estimated for the Board sequence of the SBMnet dataset by LaBGen and LaBGen-P with the parameters discussed in Section III-C. The discontinuities observable between different spatial areas make the estimation visually incoherent. They are avoided with the LaBGen-P method.

In order to create and evaluate new background generation methods, the Scene Background Modeling and Initialization (SBMI) workshop was organized by Maddalena and Bouwmans in 2015 [3]. Among its contributions, the first complete benchmarking framework [4] comprises the Scene Background Initialization (SBI) dataset[1], an evaluation methodology, and a set of relevant metrics. Moreover, our method named LaBGen was introduced during this workshop [5].

The LaBGen method combines a pixel-wise median filter and a patch selection mechanism based on a motion detection performed by a background subtraction algorithm. More precisely, the image plane is divided into several spatial areas. For each spatial area, a quantity of motion is computed for the corresponding patches seen over time. Then, the patches associated to the smallest quantities of motions are selected. At the end of the process, the background estimation for a given spatial area is generated by applying a pixel-wise median filter on the corresponding subset of selected patches.

Even though LaBGen generates almost perfect results on the sequences of the SBI dataset, it can still be improved. For instance, discontinuities might appear in the generated background estimation between different spatial areas as illustrated by Fig. 1. To solve this problem, we introduce the LaBGen-P method, which is a variant of LaBGen. Instead of computing a quantity of motion for a given patch, LaBGen-P computes quantities of motion per pixel by taking into account the motion in the spatial neighborhood of each considered pixel. Thus, by selecting pixels instead of patches, LaBGen-P avoids

[1] http://sbmi2015.na.icar.cnr.it/SBIdataset.html

B. Laugraud, S. Piérard, and M. Van Droogenbroeck. **LaBGen-P: A Pixel-Level Stationary Background Generation Method Based on LaBGen**. In *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, pages 107-113, Cancún, Mexico, December 2016. DOI: 10.1109/ICPR.2016.7899617.

discontinuities between spatial areas and generates background estimations that are visually more coherent.

Hereafter, we present, discuss, and evaluate the new LaBGen-P method. In Section II, we provide a complete description of the method, which appears to be simpler than LaBGen. In Section III, we describe our experimental setup and study the performance of the method on the new SceneBackgroundModeling.NET (SBMnet) dataset[2] proposed by Jodoin et al. We also compare the performance of LaBGen-P and LaBGen on this dataset, both objectively (with the recommended metrics) and subjectively (by asking 35 human experts to compare their results). Moreover, we present in Section IV the results of the IEEE Scene Background Modeling Contest (SBMC) 2016, where LaBGen-P and LaBGen emerged as the best presented methods. Finally, we draw some conclusions in Section V.

II. Description of the method

The LaBGen-P method, whose open-source C++ implementation is available at http://www.telecom.ulg.ac.be/labgen, is a variant of the LaBGen method. To understand the former, we start with an overview of the latter. The LaBGen method was introduced in [5]. It combines a pixel-wise median filter and a patch selection mechanism based on the classifications performed by a background subtraction algorithm. LaBGen can be summarized in the five following steps:

1) An *augmentation step* increases the length of the input video sequence. It allows background subtraction algorithms to be better trained, and avoids artifacts like ghosts (*i.e.* false positive classifications due to bootstrapping problems) when short sequences are encountered. The parameter $\mathcal{P}$ controls the length of the augmented sequence.
2) *A motion detection step* determines, for each frame, which pixels belong to the background. This classification is performed by a background subtraction algorithm identified by the parameter $\mathcal{A}$, and stored into a binary *segmentation map*.
3) Based on the segmentation maps, *quantities of motion* are estimated locally by counting the number of pixels classified as foreground in each spatial area. The size of these areas depends on the parameter $\mathcal{N}$.
4) The resulting quantities of motion are used to select, for each spatial area, the *subset of patches* with the least motion. The size of the subsets is defined by the parameter $\mathcal{S}$.
5) The background image B is then estimated by applying a pixel-wise median filter on the subsets of selected patches.

Even though LaBGen and LaBGen-P share many principles, they also present significant differences. Unlike the former, LaBGen-P selects pixels instead of patches. The modifications induced by this major difference are discussed below.

[2]http://scenebackgroundmodeling.net

It has been shown in [5] that the frame difference is the background subtraction algorithm bringing the most valuable contribution to the performance of LaBGen in average. Thus, the parameter $\mathcal{A}$ is discarded and the frame difference is systematically used. As this algorithm does not require any training period, the augmentation step described above is removed and the parameter $\mathcal{P}$ is also discarded. This reduces the number of parameters by 2, and simplifies the method.

In LaBGen-P, the results of the motion detection step are not binary anymore. Thus, for a given frame at time t, *motion scores* are computed for each pixel $p^t_{x,y}$ (with (x, y) being the pixel coordinates starting from $(0, 0)$) and stored into the corresponding *motion map* m^t. This modification avoids the need to find a correct hard threshold and allows the method to capture some shades about observed motions. The motion score $m^t_{x,y}$ of the pixel $p^t_{x,y}$ is given by the frame difference:

$$m^t_{x,y} = \left| p^t_{x,y} - p^{t-1}_{x,y} \right| . \tag{1}$$

Furthermore, instead of estimating a quantity of motion per spatial area, a quantity of motion is estimated per pixel considering the spatial neighborhood. Thus, to compute the *quantity of motion* $q^t_{x,y}$, all the motion scores inside a window centered on the pixel $p^t_{x,y}$ are added. The size $W \times W$ of this window is defined by an odd natural number W such that:

$$W = 1 + 2 \left\lfloor \frac{\min(w, h)}{2\mathcal{N}} \right\rfloor , \tag{2}$$

with w and h being respectively the width and height of the input video sequence, and $\mathcal{N}$ a parameter. For simplicity, we choose to ignore pixels outside the limits of the image plane. To give a precise definition of the computed quantity of motion, let P_x (resp. P_y) be the predicate indicating whether a pixel coordinate x' (resp. y') is in a window centered on x (resp. y):

$$\begin{aligned} P_x(x') = \;& x' \geq \max\left(x - \lfloor W/2 \rfloor, 0\right) \wedge \\ & x' \leq \min\left(x + \lfloor W/2 \rfloor, w - 1\right) \\ P_y(y') = \;& y' \geq \max\left(y - \lfloor W/2 \rfloor, 0\right) \wedge \\ & y' \leq \min\left(y + \lfloor W/2 \rfloor, h - 1\right) \end{aligned} , \tag{3}$$

and let $\Psi_{x,y}$ be the set of pixel coordinates inside the window centered on pixel $p_{x,y}$:

$$\Psi_{x,y} = \left\{ (x', y') \,|\, P_x(x') \wedge P_y(y') \right\} . \tag{4}$$

The quantity of motion $q^t_{x,y}$ is then defined as follows:

$$q^t_{x,y} = \sum_{(x',y') \in \Psi_{x,y}} m^t_{x',y'} . \tag{5}$$

In our implementation, the computation of quantities of motion has been sped up by using summed area tables [6]. Note that after an initialization of a linear complexity $\mathcal{O}(wh)$, these tables allow to compute any quantity of motion in a constant time $\mathcal{O}(1)$, regardless the size $W \times W$ of the windows.

Once the quantities of motion have been computed, LaBGen-P iteratively builds, for each pixel $p_{x,y}$, subsets $\Omega_{x,y}$ of maximum $\mathcal{S}$ selected pixels. In order to initialize a subset,

B. Laugraud, S. Piérard, and M. Van Droogenbroeck. **LaBGen-P: A Pixel-Level Stationary Background Generation Method Based on LaBGen**. In *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, pages 107-113, Cancún, Mexico, December 2016. DOI: 10.1109/ICPR.2016.7899617.

the first encountered pixels are added into $\Omega_{x,y}$ while its cardinality is less than $\mathcal{S}$. After that, a pixel is added at time t only when a *selection criterion* is satisfied. This criterion checks whether the quantity of motion $q^t_{x,y}$ associated to a candidate pixel $p^t_{x,y}$ is less or equal to at least one quantity of motion associated to a pixel already in the subset. To keep the cardinality of a subset $\Omega_{x,y}$ equal to $\mathcal{S}$, we remove the pixel associated to the largest quantity of motion. If several pixels are associated to this quantity, we discard the oldest one.

Finally, the last step remains unchanged as a median filter is applied on each subset of selected pixels after the processing of the last frame. Note that a background estimation could be generated at any time t by applying a median filter on each subset $\Omega^t_{x,y}$. Considering this, our method could be used in an online mode enlarging the number of applications in which it could be useful.

III. Experiments and results

A. Experimental setup

Our experimental setup consists in 2 datasets (SBI and SBMnet) and 6 metrics used for evaluating our results. The SBI dataset is composed of 14 video sequences with ground truth from 6 to 740 frames and whose resolution varies from 144×144 to 800×600. The SBMnet dataset is composed of 79 video sequences scattered through 8 categories: Basic, Intermittent Motion, Clutter, Jitter, Illumination Changes, Background Motion, Very Long, and Very Short. They are composed of 6 to 9370 frames and their resolution varies from 240×240 to 800×600. To the contrary of what was done for the SBI dataset, ground truth is provided for only 13 sequences distributed among categories. Note that for this reason, we are unable to draw category specific conclusions in our experiments.

The six following metrics [4] are used to assess our experiments. The ones to minimize (resp. maximize) are followed by a ↓ (resp. ↑) symbol.

- Average Gray-level Error (AGE, ↓, from 0 to 255): average of the absolute difference between the gray-scale values of an input and a ground truth image.
- Percentage of Error Pixels (pEPs, ↓): a difference of gray-scale values larger than 20 is considered as an error.
- Percentage of Clustered Error Pixels (pCEPs, ↓): any error pixel whose 4-connected neighbors are also error pixels according to pEPs.
- Peak-Signal-to-Noise-Ratio (PSNR, ↑): defined by Eq. 6, with MSE being the Mean Squared Error:

$$\mathrm{PSNR} = 10 \log_{10} \frac{255^2}{\mathrm{MSE}}\, dB. \tag{6}$$

- Multi-Scale Structural Similarity Index (MS-SSIM, ↑, from 0 to 1): estimation of the perceived visual distortion.
- Color image Quality Measure (CQM, ↑, in dB): combination of per-channel PSNRs computed on an approximated reversible RGB to YUV transformation.

Although in our experiments the scores provided by all metrics are always reported, we use CQM as our reference metric when an optimization criterion is needed. Optimizing according to CQM offers the advantage to lead to results that are visually more satisfactory and coherent. Furthermore, CQM is the only metric to use the information provided by each RGB channel.

B. Parameters space

In order to evaluate the performance of the LaBGen-P method, an estimation of the background has been generated for each sequence provided with ground truth of the SBI and SBMnet datasets using each combination of $\mathcal{N} = 1, 2, 3, \ldots, 50$ and $\mathcal{S} = 1, 3, 5, \ldots, 201$. Even though LaBGen-P is embedded with a unique motion detection algorithm, we have tested two variants of the frame difference to compute motion scores. The first (named F. Diff. C1 hereafter) is used with gray-scale values and is defined by Eq. 1. The RGB to gray-scale conversion has been performed as follows, with $r^t_{x,y}$, $g^t_{x,y}$, and $b^t_{x,y}$ being respectively the red, green, and blue components of pixel $p^t_{x,y}$:

$$p^t_{x,y} = 0.299 \cdot r^t_{x,y} + 0.587 \cdot g^t_{x,y} + 0.114 \cdot b^t_{x,y}. \tag{7}$$

The second (named F. Diff. C3 hereafter) is used with RGB colors and is defined by Eq. 8:

$$m^t_{x,y} = \left|r^t_{x,y} - r^{t-1}_{x,y}\right| + \left|g^t_{x,y} - g^{t-1}_{x,y}\right| + \left|b^t_{x,y} - b^{t-1}_{x,y}\right|. \tag{8}$$

C. Best average performance

The best average performance achieved by LaBGen-P is provided by Table I. The best sets of parameters have been found by maximizing the CQM metric averaged over 27 video sequences (the 13 sequences of SBMnet with ground truth and 14 sequences of SBI). One can observe that the median method is always ranked below LaBGen-P, regardless of the used variant of the frame difference. Furthermore, as the scores reported for each variant are almost the same, any of them can be used without a significant performance loss. However, as we have to determine a *default set of parameters*, we define this set as $\mathcal{N} = 3$, and $\mathcal{S} = 19$ with F. Diff. C1.

Table II provides the performance reached by LaBGen-P using the default set of parameters for each SBMnet sequence with ground truth (the generated background images are given Fig. 2), and Table III the one reached by LaBGen using a unique set of parameters for the same sequences. To compare both algorithms, the parameters used with LaBGen are the same than the ones used with LaBGen-P, with $\mathcal{A} =$ F. Diff. C1 (the motions scores produced by F. Diff. C1 are thresholded) and $\mathcal{P} = 1$ (no augmentation at all). According to the majority of metrics, one method is better than the other one for about half of the sequences with ground truth. This suggests that the use of patches or pixels is equivalent, on average, with respect to the recommended metrics.

D. Subjective evaluation

Because there is not yet a clear understanding of the various metrics, and the number of video sequences provided with the corresponding ground truth is too small, we have performed subjective tests. A total of 35 human experts saw pairs of

B. Laugraud, S. Piérard, and M. Van Droogenbroeck. **LaBGen-P: A Pixel-Level Stationary Background Generation Method Based on LaBGen**. In *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, pages 107-113, Cancún, Mexico, December 2016. DOI: 10.1109/ICPR.2016.7899617.

TABLE I
BEST AVERAGE PERFORMANCE OF LABGEN-P ON THE SEQUENCES OF THE SBI AND SBMNET DATASET.

		Best parameters		Averaged metrics					
F. Diff	**Rank**	$\mathcal{N}$	$\mathcal{S}$	**AGE** ↓	**pEPs** ↓	**pCEPs** ↓	**PSNR** ↑	**MS-SSIM** ↑	**CQM** ↑
C1	1	3	19	4.7150	3.5298%	1.6824%	31.6625	0.9530	32.3464
C3	2	5	25	5.2047	3.8983%	2.0591%	31.6332	0.9488	32.3119
Median method				9.7053	10.9965%	8.1457%	27.5757	0.8703	28.3883

TABLE II
PERFORMANCE OF LABGEN-P USING THE DEFAULT SET OF PARAMETERS FOR EACH SBMNET SEQUENCE WITH GROUND TRUTH.

		Metrics					
Category	**Sequence**	**AGE** ↓	**pEPs** ↓	**pCEPs** ↓	**PSNR** ↑	**MS-SSIM** ↑	**CQM** ↑
Intermittent Motion	AVSS2007	9.3098	7.4672%	5.4991%	21.7030	0.8793	22.5939
	busStation	2.8890	0.8218%	0.2905%	33.2031	0.9853	33.9428
Background Motion	advertisementBoard	1.6658	0.0083%	0.0000%	40.8086	0.9967	40.8971
Clutter	boulevardJam	17.0728	22.8177%	12.5742%	19.3285	0.5551	20.5361
	Board	6.5073	4.0244%	0.9573%	28.4209	0.9107	29.3391
Very Short	CUHK_Square	2.8796	0.4157%	0.0092%	34.8721	0.9885	35.0981
	DynamicBackground	7.3805	5.3680%	0.1056%	27.5241	0.9628	28.0665
Basic	Blurred	1.9237	0.0422%	0.0000%	38.6874	0.9961	38.9903
	511	4.7435	5.0996%	0.2923%	27.8479	0.9502	29.6688
Illumination Changes	CameraParameter	6.6406	3.0182%	2.6823%	18.4200	0.9387	20.2001
Very Long	BusStopMorning	6.2404	3.1602%	0.1602%	28.1427	0.9794	28.8607
Jitter	badminton	2.2521	0.9722%	0.1973%	35.0043	0.9801	35.6124
	boulevard	9.7225	12.9414%	2.0647%	21.7985	0.9000	23.2436

TABLE III
PERFORMANCE OF LABGEN USING A UNIQUE SET OF PARAMETERS FOR EACH SBMNET SEQUENCE WITH GROUND TRUTH.

		Metrics					
Category	**Sequence**	**AGE** ↓	**pEPs** ↓	**pCEPs** ↓	**PSNR** ↑	**MS-SSIM** ↑	**CQM** ↑
Intermittent Motion	AVSS2007	8.6178	7.3580%	5.8109%	21.0771	0.8952	21.9665
	busStation	7.0289	4.8796%	3.6817%	22.1009	0.8890	23.0759
Background Motion	advertisementBoard	1.7101	0.2185%	0.1122%	39.2685	0.9919	38.5256
Clutter	boulevardJam	8.2270	10.8958%	6.4232%	22.6539	0.6852	23.9999
	Board	8.0354	5.0030%	0.9878%	27.4030	0.8564	28.3531
Very Short	CUHK_Square	2.6470	0.3033%	0.0043%	35.5707	0.9908	35.7489
	DynamicBackground	6.7240	4.1970%	0.0374%	28.3835	0.9661	28.9499
Basic	Blurred	1.9293	0.0422%	0.0000%	38.6870	0.9961	38.9405
	511	4.8156	5.1956%	0.3457%	27.6669	0.9481	29.5005
Illumination Changes	CameraParameter	1.4271	0.0143%	0.0000%	42.5794	0.9970	43.2695
Very Long	BusStopMorning	6.2404	3.1602%	0.1602%	28.1427	0.9794	28.8607
Jitter	badminton	2.2731	1.0723%	0.3032%	34.6599	0.9807	35.2966
	boulevard	10.1775	13.8304%	2.1801%	21.4658	0.8949	22.9257

TABLE IV
BEST PERFORMANCE OF LABGEN-P FOR EACH SBMNET SEQUENCE WITH GROUND TRUTH ACCORDING TO CQM.

		Best parameters			Metrics					
Category	**Sequence**	$\mathcal{N}$	$\mathcal{S}$	**F. Diff.**	**AGE** ↓	**pEPs** ↓	**pCEPs** ↓	**PSNR** ↑	**MS-SSIM** ↑	**CQM** ↑
Intermittent Motion	AVSS2007	5	1	C1	9.0387	5.8232%	4.0456%	22.1251	0.8778	23.0481
	busStation	8	23	C3	2.6209	0.5347%	0.1574%	35.2979	0.9897	35.8762
Background Motion	advertisementBoard	2	125	C3	1.4821	0.0053%	0.0000%	41.6413	0.9971	41.6582
Clutter	boulevardJam	41	81	C1	3.4112	2.8177%	0.8307%	28.8930	0.8979	30.1544
	Board	2	69	C3	5.4661	1.3140%	0.1067%	30.8233	0.9318	31.6931
Very Short	CUHK_Square	1	7	C1	2.8796	0.4157%	0.0092%	34.8721	0.9885	35.0981
	DynamicBackground	1	5	C1	7.3805	5.3680%	0.1056%	27.5241	0.9628	28.0665
Basic	Blurred	2	67	C3	1.7839	0.0325%	0.0000%	39.3885	0.9968	39.6530
	511	1	199	C1	3.4379	2.2096%	0.0368%	31.4552	0.9828	32.9898
Illumination Changes	CameraParameter	2	113	C1	1.3925	0.0130%	0.0000%	42.9167	0.9975	43.6507
Very Long	BusStopMorning	16	201	C3	5.1679	1.9036%	0.0547%	29.8094	0.9872	30.5270
Jitter	badminton	1	121	C3	1.7968	0.3895%	0.0480%	37.6193	0.9884	38.1299
	boulevard	1	5	C1	8.9716	11.5691%	1.7020%	22.4568	0.9120	23.8302

B. LAUGRAUD, S. PIÉRARD, AND M. VAN DROOGENBROECK. **LaBGen-P: A Pixel-Level Stationary Background Generation Method Based on LaBGen**. In *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, pages 107-113, Cancún, Mexico, December 2016. DOI: 10.1109/ICPR.2016.7899617.

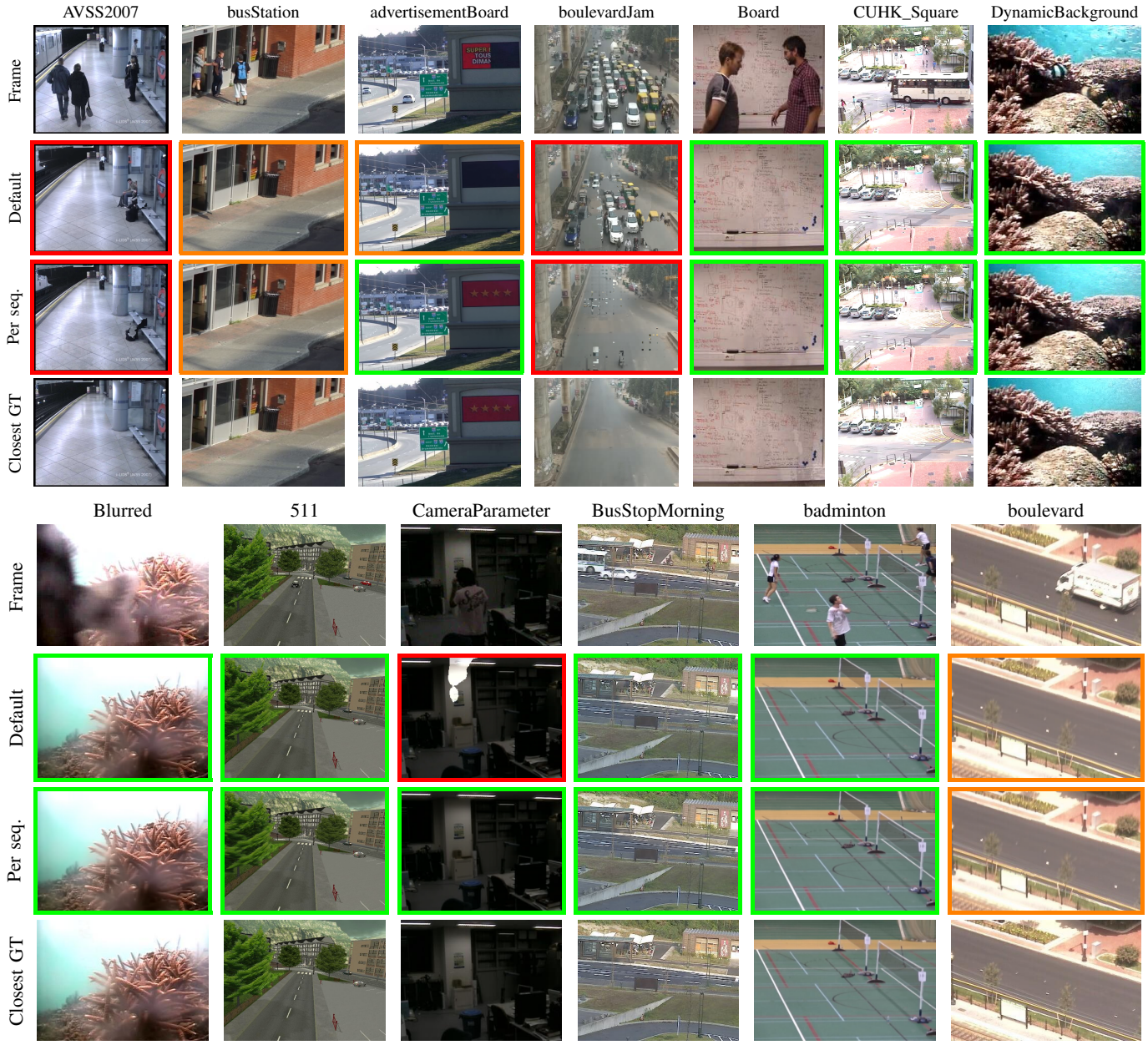

Fig. 2. Visual results obtained for the sequences of the SBMnet dataset with ground truth (a frame randomly selected is provided on the 1st row) by LaBGen-P with the default set of parameters (2nd row, see Section III-C), LaBGen-P with sets of per sequence optimized parameters (3rd row, see Table IV). The closest ground truths to our per sequence optimized results are provided in the 4th row. The estimations with major, minor, and no visual errors are respectively surrounded by a red □, orange □, and green □ frame.

Fig. 3. Examples of SBMnet sequences without ground truth for which the backgrounds generated by LaBGen-P are preferred by the human experts (see Section III-C).

background images generated by LaBGen-P and LaBGen displayed on a screen, side by side, with the associated video sequence. For each of the 79 SBMnet video sequences, we asked which background image is the best, and the experts had to choose among three answers: "the left one", "the right one", and "I don't know". In order to avoid any bias, the two images were shown in random order, and the order of the video sequences was also randomized. Moreover, there was no time limit to answer. It was also possible to stop before the end of the questions, in order to avoid any bias due to the fatigue of human experts. We collected a total of 2210

B. Laugraud, S. Piérard, and M. Van Droogenbroeck. **LaBGen-P: A Pixel-Level Stationary Background Generation Method Based on LaBGen**. In *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, pages 107-113, Cancún, Mexico, December 2016. DOI: 10.1109/ICPR.2016.7899617.

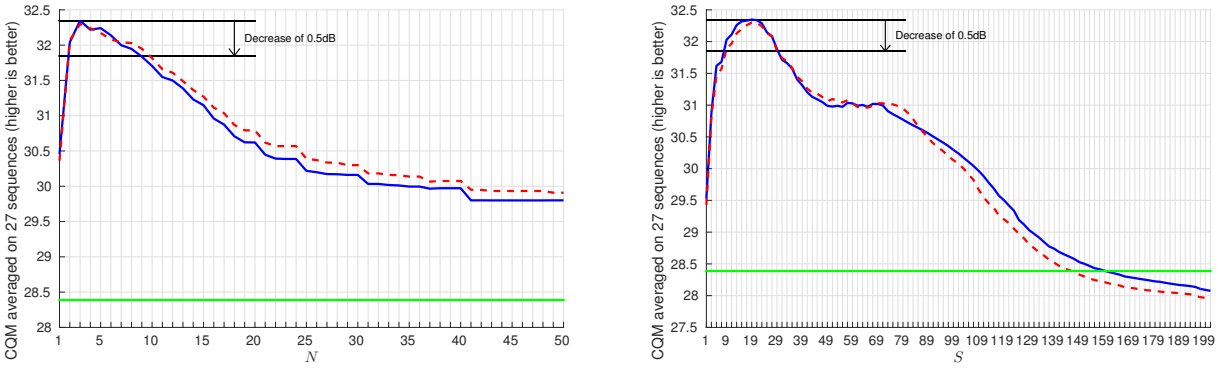

Fig. 4. Performance stability of LaBGen-P with respect to the $\mathcal{N}$ and $\mathcal{S}$ parameters, when they vary around the default set of parameters ($\mathcal{N} = 3$, $\mathcal{S} = 19$) with F. Diff. C1 —, and F. Diff. C3 ---. The median method is our baseline —.

answers (between 26 and 30 answers per video sequences with 27.975 in average). Most people were unable to choose between the two methods for 38 sequences. These sequences (for which LaBGen-P and LaBGen seems to perform equally well) put aside, the human raters preferred LaBGen-P as there are, according to the human experts, more video sequences for which LaBGen-P performs better than LaBGen (26 vs 15). Some of these sequences are shown in Fig. 3.

E. Best per sequence performance

Although it is common to provide a unique set of parameters with a method as done in Section III-C, LaBGen-P generates better estimations when the parameters are tuned per sequence. Table IV provides the best performance reached by LaBGen-P by maximizing the CQM metric for each sequence of SBMnet provided with ground truth independently (the generated background images are given Fig. 2). It can be easily observed that the majority of metrics agree on an improvement. Note that the best performances for the Very Short sequences are already achieved with the default set of parameters.

F. Performance stability

Fig. 4 shows the stability of the performance achieved by LaBGen-P, measured by CQM, when the parameters $\mathcal{N}$ and $\mathcal{S}$ vary around their local optimum. It turns out that avoiding very low values for these two parameters is critical, but that selecting high values is not much penalized as long as $\mathcal{S}$ remains less than 145. In this case, LaBGen-P performs better than the median method, with an improvement up to about $4\,dB$. Given an arbitrary tolerance of $0.5\,dB$, the average performance is not harmed and remains about $3.5\,dB$ above the median method when the parameter $\mathcal{N}$ ranges from 2 to 9, or when $\mathcal{S}$ ranges from 9 to 29. Moreover, it can also be noted on Fig. 4 that the performance is almost similar when the variants C1 and C3 are considered. This suggests that LaBGen-P is mostly insensitive to the input color space. All these observations make us believe that the performance of LaBGen-P is stable, and that a default set of parameters is suitable for most video sequences.

IV. The IEEE SBMC 2016 contest

The LaBGen-P and LaBGen methods were submitted to the IEEE Scene Background Modeling Contest (SBMC) 2016, both using the default parameters introduced in Section III-C. This contest aims at ranking background generation methods according to their results on the 79 sequences of the SBMnet dataset. Whereas most ground truths are not publicly available to avoid overfitting, an online platform performs an evaluation on the overall dataset following the ranking strategies first introduced for ChangeDetection.NET (CDnet) [17].

Table V presents the results reported on this platform. According to the first ranking strategy, LaBGen and LaBGen-P are ranked respectively first and second, just before the temporal median filter. According to the second strategy, they are ranked respectively first and third, with Photomontage becoming the second best method. As the SBMC results are derived from all sequences, observations can be made per category. Thus, it should be noted that LaBGen-P is ranked first in the Intermittent Motion category, and LaBGen second in the Illumination Changes and Very Long categories.

Even though the parameters were optimized in Section III-C based on a small number of video sequences, our top ranks reveal that our methods generalize well to most video sequences. This should not be a surprise as the study presented in Section III-F pointed out that the performance is stable with respect to the chosen parameters. Therefore fine tuning them for the complete dataset was not necessary.

Last but not least, according to Table V, both ranking strategies agree to rank LaBGen above LaBGen-P, in contrast to the opinion of human experts. Indeed, it was shown in Section III-D that both methods perform equally well on many video sequences, and that LaBGen-P is preferable to LaBGen in most of the other ones (remember that, as for SBMC, our experiments involving human experts consider the whole dataset). This contradiction tends to prove that the recommended metrics fail to capture some aspects relevant to humans for the problem of background generation. Clearly, there is a need to acquire a better understanding of these metrics and their relationships with the visual perception.

B. Laugraud, S. Piérard, and M. Van Droogenbroeck. **LaBGen-P: A Pixel-Level Stationary Background Generation Method Based on LaBGen**. In *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, pages 107-113, Cancún, Mexico, December 2016. DOI: 10.1109/ICPR.2016.7899617.

TABLE V
RESULTS OF THE IEEE SBMC 2016 CONTEST, TAKEN FROM HTTP://WWW.SCENEBACKGROUNDMODELING.NET.

Method	Average ranking across categories ↓	Average ranking ↓	Average AGE ↓	Average pEPs ↓	Average pCEPs ↓	Average PSNR ↑	Average MS-SSIM ↑	Average CQM ↑
LaBGen [5]	**4.25**	**3.33**	**6.7090**	**6.31%**	2.65%	**28.6396**	0.9266	**29.4668**
LaBGen-P (this paper)	4.88	4.50	7.0738	7.06%	3.19%	28.4660	0.9278	29.3196
Temporal median filter [7]	5.13	6.67	8.2761	9.84%	5.46%	27.5364	0.9130	28.4434
SC-SOBS-C4 [8]	5.63	4.67	7.5183	7.11%	2.42%	27.6533	0.9160	28.5601
Bidirectional Analysis and Consensus Voting [9]	5.75	7.33	8.5816	7.24%	2.57%	26.1018	0.9078	27.1000
Bidirectional Analysis [9]	5.75	6.67	8.3449	7.56%	**1.81%**	26.1722	0.9085	27.1637
Wei-Liu-Aug-16-2 [10]	5.88	8.33	9.4020	10.51%	5.66%	27.1347	0.9043	28.0530
Photomontage [11]	6.13	4.33	7.1950	6.86%	2.57%	28.0113	0.9189	28.8719
MAGRPCA [12]	8.13	7.33	8.3132	9.94%	5.67%	28.4556	**0.9401**	29.3152
FC-FlowNet [13]	9.00	10.00	9.1131	11.28%	5.99%	26.9559	0.9162	27.8767
RMR [14]	9.13	9.50	9.5363	11.76%	5.82%	26.5217	0.8790	27.4549
RSL2011 [15]	9.38	8.50	9.0443	10.08%	4.97%	25.8051	0.8891	26.7986
AAPSA [16]	10.25	9.83	9.2044	10.57%	5.23%	25.3947	0.9000	26.3021

V. CONCLUSION

In this paper, we presented a new method for stationary background generation called LaBGen-P (see http://www.telecom.ulg.ac.be/labgen for the C++ source code). It is a variant of LaBGen that mainly avoids the discontinuities between different spatial areas and generates better visual results. Moreover, it has fewer parameters and it is simpler. We optimized its parameters using the SBI dataset and a subset of the SBMnet one, and studied its performance on this subset. Even though the performance achieved by LaBGen-P and LaBGen are close, we show that LaBGen-P generates better background images considering the overall SBMnet dataset. To reach this conclusion, we have proceeded to a thorough subjective evaluation. More precisely, we have asked 35 human experts to perform 2210 comparisons in order to select which one is best. The results of LaBGen-P are preferred for 26 sequences and the ones of LaBGen for 15 sequences; no choice was made for 38 sequences. Moreover, the results published online for the IEEE SBMC 2016 contest show that LaBGen-P and LaBGen are, to date, the two best known methods to generate a stationary background image given a video sequence.

REFERENCES

[1] L. Maddalena and A. Petrosino, "Background model initialization for static cameras," in *Background Modeling and Foreground Detection for Video Surveillance*. Chapman and Hall/CRC, 2014, ch. 3, pp. 3.1–3.16.

[2] M. Cristani, M. Farenzena, D. Bloisi, and V. Murino, "Background subtraction for automated multisensor surveillance: A comprehensive review," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, p. 24 pages, 2010.

[3] L. Maddalena and T. Bouwmans, "Scene background modeling and initialization (SBMI) workshop," http://sbmi2015.na.icar.cnr.it, Genova, Italy, September 2015.

[4] L. Maddalena and A. Petrosino, "Towards benchmarking scene background initialization," in *International Conference on Image Analysis and Processing Workshops (ICIAP Workshops)*, ser. Lecture Notes in Computer Science, vol. 9281, September 2015, pp. 469–476.

[5] B. Laugraud, S. Piérard, M. Braham, and M. Van Droogenbroeck, "Simple median-based method for stationary background generation using background subtraction algorithms," in *International Conference on Image Analysis and Processing (ICIAP), Workshop on Scene Background Modeling and Initialization (SBMI)*, ser. Lecture Notes in Computer Science, vol. 9281. Springer, September 2015, pp. 477–484.

[6] F. Crow, "Summed-area tables for texture mapping," in *Proceedings of SIGGRAPH 1984*, ser. Computer Graphics, vol. 18. ACM, July 1984, pp. 207–212.

[7] M. Piccardi, "Background subtraction techniques: a review," in *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, vol. 4, The Hague, The Netherlands, October 2004, pp. 3099–3104.

[8] L. Maddalena and A. Petrosino, "Extracting a background image by a multi-modal scene background model," in *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, Cancun, Mexico, 2016.

[9] T. Minematsu, A. Shimada, and R.-I. Taniguchi, "Background initialization based on bidirectional analysis and consensus voting," in *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, Cancun, Mexico, 2016.

[10] W. Liu, Y. Cai, M. Zhang, H. Li, and H. Gu, "Scene background estimation based on temporal median filter with gaussian filtering," in *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, Cancun, Mexico, 2016.

[11] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 294–302, August 2004.

[12] S. Javed, A. Mahmmod, T. Bouwmans, and S. K. Jung, "Motion-aware graph regularized rpca for background modeling of complex scene," in *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, Cancun, Mexico, 2016.

[13] I. Halfaoui, F. Bouzaraa, and O. Urfalioglu, "CNN-based initial background estimation," in *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, Cancun, Mexico, 2016.

[14] D. Ortego, J. M. SanMiguel, and J. M. Martínez, "Rejection based multipath reconstruction for background estimation in video sequences with stationary objects," *Computer Vision and Image Understanding*, vol. 147, pp. 23–37, 2016.

[15] V. Reddy, C. Sanderson, and B. Lovell, "A low-complexity algorithm for static background estimation from cluttered image sequences in surveillance contexts," *EURASIP Journal on Image and Video Processing*, vol. 164956, p. 13 pages, 2011.

[16] M. Chacon-Murguia, G. Ramirez-Alonso, and J. Ramirez-Quintana, "Evaluation of the background modeling method auto-adaptive parallel som architecture in the sbmnet dataset," in *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, Cancun, Mexico, 2016.

[17] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "changedetection.net: A new change detection benchmark dataset," in *IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Providence, Rhode Island, USA, June 2012.

B. LAUGRAUD, S. PIÉRARD, AND M. VAN DROOGENBROECK. **LaBGen-P: A Pixel-Level Stationary Background Generation Method Based on LaBGen**. In *IEEE International Conference on Pattern Recognition (ICPR), IEEE Scene Background Modeling Contest (SBMC)*, pages 107-113, Cancún, Mexico, December 2016. DOI: 10.1109/ICPR.2016.7899617.