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

B. Laugraud, S. Piérard, M. Van Droogenbroeck

INTELSIG Laboratory, University of Liège, Belgium

December 4th 2016

IEEE Scene Background Modeling Contest (SBMC 2016)

Cancun, Mexico
LaBGen-P is a stationary background generation method.

It is a simpler pixel-based version of LaBGen.

LaBGen should be introduced to understand LaBGen-P.
- It combines a pixel-wise median filter and a patch selection mechanism.

- The selection mechanism is based on motion detection.

- This mechanism selects the patches with the smallest amounts of motion.

- The pipeline of the method comprises 5 steps.
LaBGen: Step 1 - Augmentation

- Increases the duration of the input video sequence.

- In fact, we process the sequence in $P$ passes.

- An odd pass is performed forwards while an even pass is performed backwards.
LaBGen: Step 2 - Motion detection

- We chose to work with background subtraction (bgs) algorithms.
- The training of the considered algorithm $\mathcal{A}$ is helped by the augmentation step.
- LaBGen does not use the model of $\mathcal{A}$, only segmentation maps.
- LaBGen can be used with any bgs algorithm “out-of-the-box”.

Background Subtraction
LaBGen: Step 3 - Local estimation of the quantity of motion

- The image plane is divided into $N \times N$ spatial areas.
- A quantity of motion $q$ is estimated for each patch.
- It represents the probability of observing pixels corresponding to moving objects.

$$q = \frac{\text{# pixels classified as foreground in the patch}}{\text{# pixels in the patch}}$$
In each spatial area, $S$ patches are selected.

The $S$ selected patches are associated to the smallest quantities of motion $q$. 
LaBGen: Step 5 - Background generation

- A pixel-wise median filter is applied on the sets of $S$ selected patches.

- The background is then generated.
LaBGen-P: Motivation

- Sometimes, with LaBGen, we have a "patch effect".

- We wanted to make a pixel-based method to avoid this effect.

- LaBGen-P(ixel).

Backgrounds estimated with the same parameters!
LaBGen-P: What is new?

LaBGen-P is now pixel-based!

LaBGen $\uparrow$ $\downarrow$ LaBGen-P

LaBGen $\uparrow$

LaBGen-P $\downarrow$

LaBGen-P is now pixel-based!
- The frame difference has the most valuable contribution in average for LaBGen.
- Only the frame difference is used in LaBGen-P (no $A$ and $P$ parameter).
LaBGen-P: Motion maps

- No threshold is applied on the resulting differences (*motion scores*) any more.
- The motions scores are put in a *motion map*.
- Such a map allows to capture some shades about motion.
- For instance: 200 > 20 → fg, 30 > 20 → fg, but $p(fg|200) > p(fg|30)$.
LaBGen-P: Local estimation of the quantity of motion

- Unlike in LaBGen, quantities of motion are estimated per pixel, but locally!
- The motion scores available in the local neighbourhood are aggregated (sum).
- The local neighbourhood is delimited by a window centered on the current pixel.
- The size of the window depends on the parameter $N$.

\[
\text{quantity of motion of } \blacksquare = \sum \blacksquare = 1120
\]
Drawbacks

<table>
<thead>
<tr>
<th>Default</th>
<th>AVSS2007</th>
<th>boulevardJam</th>
<th>CameraParameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per seq.</td>
<td><img src="default.png" alt="Image" /></td>
<td><img src="boulevardJam.png" alt="Image" /></td>
<td><img src="CameraParameter.png" alt="Image" /></td>
</tr>
<tr>
<td>Closest GT</td>
<td><img src="default.png" alt="Image" /></td>
<td><img src="boulevardJam.png" alt="Image" /></td>
<td><img src="CameraParameter.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Quantitative evaluation

- We have ground-truth (GT) for $\sim \frac{1}{6}$ of the sequences.
- Metrics consider LaBGen-P better for half of the sequences with GT.
- Is LaBGen-P better than LaBGen considering the overall dataset?
Subjective evaluation - Web platform

1. Video for which we would like to define a background image

[Video] video/Candela_m1.10.m4v
2. Question

Which background image do you prefer?

--- Please select the correct answer! ---

I don't know.
The one on the left hand side.
The one on the right hand side.

save the answers and display the next question.

Copyright Piérard Sébastien, 2012
Subjective evaluation

- **35** human experts participated.

- We collected **2210** answers (≈ **28** answers in average per video sequence).

- Unable to choose between LaBGen and LaBGen-P for **38** sequences.

- LaBGen-P was prefered for **26** sequences and LaBGen for **15** sequences.
### Results (September 12, 2016)

Results, all categories combined.

Click on method name for more details.

<table>
<thead>
<tr>
<th>Method</th>
<th>Average ranking</th>
<th>Average ranking across categories</th>
<th>Average AGE</th>
<th>Average pEPs</th>
<th>Average pCEPs</th>
<th>Average MS-SSIM</th>
<th>Average PSNR</th>
<th>Average CQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>LaBGen [6]</td>
<td>2.00</td>
<td>4.75</td>
<td>6.7090</td>
<td>0.0631</td>
<td>0.0265</td>
<td>0.9266</td>
<td>28.6396</td>
<td>29.4668</td>
</tr>
<tr>
<td>LaBGen-P [7]</td>
<td>2.63</td>
<td>5.38</td>
<td>7.0738</td>
<td>0.0706</td>
<td>0.0319</td>
<td>0.9278</td>
<td>28.4860</td>
<td>29.3106</td>
</tr>
<tr>
<td>Photomontage [3]</td>
<td>3.33</td>
<td>6.50</td>
<td>7.1950</td>
<td>0.0606</td>
<td>0.0257</td>
<td>0.9169</td>
<td>28.0113</td>
<td>28.6719</td>
</tr>
<tr>
<td>SC SOBS-C4 [9]</td>
<td>4.33</td>
<td>6.00</td>
<td>7.5183</td>
<td>0.0711</td>
<td>0.0242</td>
<td>0.9160</td>
<td>27.6533</td>
<td>28.5601</td>
</tr>
<tr>
<td>MAGICPA [10]</td>
<td>5.63</td>
<td>6.83</td>
<td>8.3132</td>
<td>0.0994</td>
<td>0.0567</td>
<td>0.9401</td>
<td>28.4556</td>
<td>29.3152</td>
</tr>
<tr>
<td>Temporal median filter [2]</td>
<td>7.17</td>
<td>5.50</td>
<td>8.2761</td>
<td>0.0984</td>
<td>0.0546</td>
<td>0.9130</td>
<td>27.5364</td>
<td>28.4434</td>
</tr>
<tr>
<td>BE-AAPSA [14]</td>
<td>7.17</td>
<td>7.33</td>
<td>7.9065</td>
<td>0.0873</td>
<td>0.0447</td>
<td>0.9127</td>
<td>27.0714</td>
<td>27.9611</td>
</tr>
<tr>
<td>Bidirectional Analysis [13]</td>
<td>7.67</td>
<td>6.83</td>
<td>8.3419</td>
<td>0.0756</td>
<td>0.0131</td>
<td>0.9086</td>
<td>26.1722</td>
<td>27.1637</td>
</tr>
<tr>
<td>Bidirectional Analysis and Consensus Voting [12]</td>
<td>8.67</td>
<td>7.75</td>
<td>8.5318</td>
<td>0.0724</td>
<td>0.0257</td>
<td>0.9070</td>
<td>26.1018</td>
<td>27.1000</td>
</tr>
<tr>
<td>TMFG [11]</td>
<td>10.00</td>
<td>6.25</td>
<td>7.4020</td>
<td>0.1061</td>
<td>0.0588</td>
<td>0.9043</td>
<td>27.1347</td>
<td>28.0690</td>
</tr>
<tr>
<td>FC-FlowNet [5]</td>
<td>10.17</td>
<td>9.00</td>
<td>9.1131</td>
<td>0.1128</td>
<td>0.0599</td>
<td>0.9102</td>
<td>26.9559</td>
<td>27.8767</td>
</tr>
<tr>
<td>RSL 2011 [4]</td>
<td>11.17</td>
<td>10.25</td>
<td>9.0443</td>
<td>0.1008</td>
<td>0.0497</td>
<td>0.8891</td>
<td>25.8951</td>
<td>26.7986</td>
</tr>
<tr>
<td>AAPSA [1]</td>
<td>12.17</td>
<td>10.88</td>
<td>9.2044</td>
<td>0.1057</td>
<td>0.0523</td>
<td>0.9000</td>
<td>25.3947</td>
<td>26.3021</td>
</tr>
<tr>
<td>RMR [8]</td>
<td>12.50</td>
<td>10.00</td>
<td>9.5383</td>
<td>0.1176</td>
<td>0.0582</td>
<td>0.8790</td>
<td>26.5217</td>
<td>27.4549</td>
</tr>
</tbody>
</table>
## Results for SBMnet 2016

### Overall

<table>
<thead>
<tr>
<th>Method</th>
<th>Average ranking</th>
<th>Average ranking across categories</th>
<th>Average AGE</th>
<th>Average pEPs</th>
<th>Average pCEPS</th>
<th>Average MSSSIM</th>
<th>Average PSNR</th>
<th>CQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSCL [15]</td>
<td>1.00</td>
<td>4.13</td>
<td>5.9547</td>
<td>0.0524</td>
<td>0.0171</td>
<td>0.9410</td>
<td>30.8952</td>
<td>31.7049</td>
</tr>
<tr>
<td>BEWIS [24]</td>
<td>3.17</td>
<td>4.88</td>
<td>6.7004</td>
<td>0.0822</td>
<td>0.0256</td>
<td>0.9282</td>
<td>28.7728</td>
<td>20.6342</td>
</tr>
<tr>
<td>LabGen [6]</td>
<td>3.67</td>
<td>6.33</td>
<td>6.7090</td>
<td>0.0631</td>
<td>0.0295</td>
<td>0.9266</td>
<td>28.5395</td>
<td>29.4668</td>
</tr>
<tr>
<td>LabGen-P [7]</td>
<td>4.83</td>
<td>7.13</td>
<td>7.0738</td>
<td>0.0706</td>
<td>0.0319</td>
<td>0.9278</td>
<td>28.4660</td>
<td>29.3196</td>
</tr>
<tr>
<td>Photomontage [3]</td>
<td>5.33</td>
<td>9.53</td>
<td>7.1950</td>
<td>0.0666</td>
<td>0.0257</td>
<td>0.9189</td>
<td>28.0113</td>
<td>26.6719</td>
</tr>
<tr>
<td>SC-SOBS-C4 [9]</td>
<td>6.33</td>
<td>8.00</td>
<td>7.5183</td>
<td>0.0711</td>
<td>0.0242</td>
<td>0.9160</td>
<td>27.6533</td>
<td>28.5601</td>
</tr>
<tr>
<td>MAGRPCA [10]</td>
<td>7.67</td>
<td>8.25</td>
<td>8.3132</td>
<td>0.0994</td>
<td>0.0507</td>
<td>0.9401</td>
<td>26.4556</td>
<td>29.3152</td>
</tr>
<tr>
<td>Temporal median filter [2]</td>
<td>9.33</td>
<td>7.50</td>
<td>8.2761</td>
<td>0.0984</td>
<td>0.0548</td>
<td>0.9130</td>
<td>27.5384</td>
<td>28.4434</td>
</tr>
<tr>
<td>BE-AAPSA [14]</td>
<td>9.33</td>
<td>10.38</td>
<td>7.9006</td>
<td>0.0673</td>
<td>0.0447</td>
<td>0.9127</td>
<td>27.0714</td>
<td>27.9811</td>
</tr>
<tr>
<td>Bidirectional Analysis [13]</td>
<td>10.00</td>
<td>9.38</td>
<td>8.3449</td>
<td>0.0756</td>
<td>0.0181</td>
<td>0.9085</td>
<td>26.1722</td>
<td>27.1637</td>
</tr>
<tr>
<td>Bidirectional Analysis and Consensus Voting [12]</td>
<td>11.00</td>
<td>11.38</td>
<td>8.5010</td>
<td>0.0724</td>
<td>0.0257</td>
<td>0.9070</td>
<td>26.1018</td>
<td>27.1000</td>
</tr>
<tr>
<td>TMEC [11]</td>
<td>12.17</td>
<td>9.00</td>
<td>9.4020</td>
<td>0.1061</td>
<td>0.0586</td>
<td>0.9043</td>
<td>27.1347</td>
<td>28.0530</td>
</tr>
<tr>
<td>FC-FlowNet [5]</td>
<td>12.33</td>
<td>11.25</td>
<td>9.1131</td>
<td>0.1126</td>
<td>0.0599</td>
<td>0.9162</td>
<td>26.9559</td>
<td>27.8767</td>
</tr>
<tr>
<td>RMAMR [20]</td>
<td>13.50</td>
<td>13.63</td>
<td>9.6995</td>
<td>0.1243</td>
<td>0.0770</td>
<td>0.9258</td>
<td>26.5380</td>
<td>27.4880</td>
</tr>
<tr>
<td>RSL2011 [4]</td>
<td>13.83</td>
<td>15.38</td>
<td>9.0443</td>
<td>0.1006</td>
<td>0.0497</td>
<td>0.8851</td>
<td>25.8051</td>
<td>26.7966</td>
</tr>
</tbody>
</table>

Click on method name for more details.
LaBGen-P is a variant of the LaBGen method.

It combines a pixel-wise median filter and a pixel selection mechanism.

It uses the frame difference as a motion detection algorithm.

Quantities of motion are computed spatially by aggregating motion scores.

It performs well on the SBMnet dataset.

The metrics consider LaBGen-P less effective than LaBGen.

A subjective evaluation has shown the contrary.

Shall we find a metric even more correlated with the human eye?
Thank you for your attention!

Do you have questions?

LaBGen website