PHOTOMETRIC BRIDGING USING FILTERED MONOCULAR OPTICAL FLOW

J. F. C. Silva* a, R. L. Barbosa b, M. Meneguette Junior a, R. B. A. Gallis a

a Faculty of Science and Technology, Unesp, Brazil - (jfcsilva, messias)@fct.unesp.br
b São Paulo Associate Faculty, São Paulo, Brazil - rluisbarbosa@uol.com.br

Commission ICWG V/1

KEY WORDS: Mobile Mapping, Optical Flow, Monocular Velocity, Image Sequence, Image Orientation

ABSTRACT:

Under certain conditions the positioning and orientation sensors such as INS and GPS of a land-based mobile mapping system may fail for a certain time interval. The consequence is that the images captured during this time interval may not be oriented properly or even may have no orientation. This article presents a solution to orient the images based only on image processing and a photogrammetric technique without any external sensors in order to overcome the lack of external orientation. A land-based mobile mapping system with a pair of video cameras and a GPS receiver was used to test the proposed methodology on an urban flat road. The video cameras were mounted on the roof of the vehicle with both optical axes parallel to the main road axis (Y). The methodology is based on the velocity estimation of the vehicle, which is done in two steps. Initially, the dense optical flow is computed then the velocity estimation is obtained through a filtering strategy that consists of using radial vectors in the low parts of the images. These radial vectors are detected by the Canny algorithm. The vehicle velocity is re-estimated after eliminating the optical flow outliers. With the re-estimated velocity and with the video sampling time the spatial displacement of each image (with respect to the previous one of the sequence) is determined. The displacement is used as an approximation to the image position. Homologous points are manually selected in the first stereo pair. Assuming an arbitrary orientation for the first pair, the corresponding object point coordinates are computed by photogrammetric intersection. The orientation is computed by phototriangulation with additional parameters. Following, new bases are included, one by one, and the phototriangulation is computed for every additional basis. The photocoordinates are determined by forward area correlation. This process has been called photogrammetric bridging or traverse. The results show that the estimated velocity is pretty close to the true one and the quality of the least square adjustment is quite acceptable, considering that no external sensors were used.

1. INTRODUCTION

The GPS (Global Positioning System) receivers and the INS (Inertial Navigation System) sensors are very important elements of a land-based mobile mapping systems (LBMMs), considering that the position and orientation of such system, which is built in a vehicle, are given by an integrated solution of GPS and INS data (Tao et al., 1999; Tao et al. 2001). In general, a LBMMs has also at least a pair of video cameras mounted on the roof of such vehicle. The camera optical axes are parallel to each other and they point forward so that it is possible to get a sequence of road image stereoscopic pairs as the vehicle moves along roads. The vehicle position and orientation are then extended to these image pairs so that it is possible to map interest objects such as poles, trees, garbage cans, curbs, road center lines, and many other features. Government and private companies have great interest on these data in order to get information for the purpose of planning, control, and maintenance.

However, under certain conditions such as in an urban environment with trees and tall buildings the GPS signal may be blocked for a certain time interval. Then INS data are used to complement the vehicle position and orientation. If the time interval lasts a few seconds the INS data may not be useful because they are degraded with the time due to the cumulative internal error process. Additionally, a low cost LBMMs may not have a complete system either without an INS mechanism or with only one GPS antenna.

This article presents a solution to orient the image pairs considering that the external sensors are not available. The optical flow is used to estimate the velocity of the vehicle and then the relative positioning among the image pairs considered that the image sequence time interval is known. The relative positioning and a set of homologous points manually extracted in the first stereo pair are used to compute the photogrammetric traverse which means to add and to orient the image pairs consecutively and automatically.

2. OPTICAL FLOW

Optical flow is the 2D distribution of the apparent velocity of the intensity pattern movement in the image plane. In other words, the field of optical flow consists of a dense field of velocity where each pixel in the image plane is associated to only one velocity vector. If the time interval between two consecutive images is known the velocity vectors can be converted into displacement vectors and vice-versa (Shi; Sun, 2000). The optical flow – or the image velocity – computation is a fundamental problem in the image sequence processing and can be useful in many tasks such as scene interpretation, exploring navigation, object following, time-to-collision evaluation, object segmentation, video coding, robot vision, and many others (Tekalp, 1995). The methods to compute the optical flow can be classified in three main groups: differential,
correlation, and frequency-energy-based techniques (Beauchemin; Barron, 1995). In the differential techniques, the initial hypothesis for the optical flow computation is that the intensity between different frames of a sequence is approximately constant in a short time interval, which means that the displacement is minimal. The image velocity is computed from the space-time differentials of the image intensity. The image intensity (domonium or differentiable) in the space and time. This can be written according to the Horn and Schunck (1981) expression:

\[ I(\hat{x}; t) \approx I(\hat{x} + \vec{d} \hat{x}, t + \Delta t) \]  

(1)

Where \( I(\hat{x}, t) \) is the image intensity function, \( \hat{x} = (x, y) \) is the position vector in the image, and \( \vec{d} \hat{x} \) is the displacement of an image region \( (\hat{x}, t) \) after the time \( \Delta t \). The right side of (1) is expanded using Taylor series:

\[ I(\hat{x}; t) + \nabla I \cdot \vec{d} \hat{x} + \Delta t \nabla I + O^2 \]  

(2)

Where \( \nabla I = (I_x, I_y) \) is the intensity spatial gradient; \( I_i \) is the first order term related to the time of \( I(\hat{x}, t) \) and \( O^2 \) are the second order and higher terms which can be neglected. \( I(\hat{x}, t) \) is subtracted from both sides and the result is divided by \( \Delta t \), then:

\[ \nabla I \frac{\vec{d} \hat{x}}{\Delta t} + I_i = 0 \Rightarrow \nabla I \vec{v} + I_i = 0 \]  

(3)

\[ \vec{v} = (u, v) = \left( \frac{\partial x}{\partial t}, \frac{\partial y}{\partial t} \right) \]  

(4)

Where (4) is the velocity in the image or the optical flow and (3) is the optical flow constraint equation and it defines a unique local restriction on image displacement (Horn; Schunck, 1981; Barron et al., 1994). The accuracy of the differential techniques depends mainly on the estimation of the intensity function partial derivatives. Besides its simplicity, the finite difference method is not able to separate signal (actual data) from noise. In order to eliminate or reduce such drawback, in general the image is smoothed by a Gaussian filter for instance.

In the real world, the 2D displacement determination process is complex because both sensor and objects in the scene may be moving relating to each other. Camera vibration, due to irregular surface roads, is another important cause to make difficult the displacement determination based on optical flow. Clouds and shadows affect the scene illumination and this may change the image sequence intensities which bring more difficulties to this estimation process.

3. VELOCITY ESTIMATION

Considering that the vehicle translational velocity is parallel to the camera optical axis (Y) the optical flow is given by (Giachetti et al., 1998):

\[ u = \frac{\omega}{f} x^2 + \frac{V}{h f} xy + \omega f \]  

\[ v = \frac{\omega}{f} xy + \frac{V}{h f} y^2 \]  

(5)

where \( f \) is the camera focal length, \( h \) is the camera height taken from the floor, \( \omega \) is the angular velocity and \((x, y)\) are the image plane coordinates. The equations (5) describe the 2D movement when the vehicle travels on a flat road considering a static scene. In other words, they consider that the unique movement comes from the cameras fixed on the top of the vehicle. This kind of movement is called passive navigation (Giachetti et al., 1998) and the authors state that the estimated velocity can be very different from the true one.

The angular and translational velocities are computed by the least square method and so the use of the dense optical flow is not recommended either for the computational effort (many points) or the presence of a lot of noises (low quality results). In order to reduce the quantity of vectors used in the translational velocity estimation, the following procedure is advisable: (i) to take the vectors in the region of interest (lower half of the image), (ii) to take only vectors with radial pattern (because of the forward vehicle displacement and the planar image displacement), (iii) and the radial vectors must have their origins in points (pixels) detected by the Canny algorithm (a borderline point or pixel) (Canny, 1986). This filtering technique is equivalent to a first stage which is complemented by a second one which eliminates those residuals that fall out of the interval:

\[ (\overline{u} - s_u, \overline{u} + s_u) \]  

(6)

\[ (\overline{v} - s_v, \overline{v} + s_v) \]

\( \overline{u} \) is the average (eq. 4) and \( s_u \) is the standard deviation of the residuals for the \( u \) component of the optical flow; \( \sigma \) and \( s_v \) are the same related to the \( v \) component. The second stage re-estimates the velocity based only on those radial vectors that are in the one standard deviation interval (eq. 6).

4. IMAGE ORIENTATION

With the velocity estimated from a stereo pair sequence the relative positioning is computed for each camera using the video sampling rate (30 frames per second). Homologous points are manually extracted in the first stereo pair. The related object point coordinates are estimated by photogrammetric intersection using the collinearity equations with additional parameters \((x_0, y_0, k_1)\):

\[ \begin{align*}
x &= x_0 - f \frac{N_z}{D} + (x - x_0) k_1 r^2 \\
y &= y_0 - f \frac{N_z}{D} + (y - y_0) k_1 r^2
\end{align*} \]  

(7)

These object coordinates are used for the photogrammetric resection of the second stereo pair together with the new values for the perspective center coordinates given by the forward displacement which is derived from the estimated velocity and
the video sampling rate. Rotation matrix is approximated by the identity matrix. A forward area correlation determines the
photocoordinates in the second pair which is pretty similar to the first one so that the optical flow technique is differential.
The correlation window is minimal so the accuracy is improved. This process is repeated for every stereo pair added to the
sequence and it is called photogrammetric bridging or traverse, which means a sequential procedure of photogrammetric
intersection and resection (Silva et al., 1999; Silva et al., 2000; Silva et al., 2003). Besides the functional model, two types of constraints are
considered to compute the spatial resection: stereo basis and velocity constraints. The stereo basis constraint uses the fixed
distance between the perspective centers (eq. 8) and the velocity constraint takes the stereo basis advances (given by the basis
displacement) divided by the time (eq. 9):

\[
B_i = \sqrt{(X_{C}^i - X_{C}^i)^2 + (Y_{C}^i - Y_{C}^i)^2 + (Z_{C}^i - Z_{C}^i)^2}
\]

(8)

\[
\begin{align*}
V_x &= \frac{X_{C(i+1)} - X_{C(i)}}{t} \\
V_y &= \frac{Y_{C(i+1)} - Y_{C(i)}}{t} \\
V_z &= \frac{Z_{C(i+1)} - Z_{C(i)}}{t}
\end{align*}
\]

(9)

and i+1 stand for two consecutive bases.

When the velocity is considered as a parameter, the collinearity equations can be rewritten in order to estimate the velocity of the
perspective centers (Gosh, 1985):

\[
x = -f_r \left( X - X_{C(i)} - tV_x \right) + r_v \left( Z - Z_{C(i)} - tV_y \right) + r_v \left( Y - Y_{C(i)} - tV_z \right) \\
y = -f_r \left( X - X_{C(i)} - tV_x \right) + r_v \left( Z - Z_{C(i)} - tV_y \right) + r_v \left( Y - Y_{C(i)} - tV_z \right)
\]

(9)

\[tV_x, tV_y, tV_z\] are the displacements between two

5. RESULTS FROM THE PROPOSED METHODOLOGY

Firstly, the results for the velocity estimation are presented. The images were taken on a flat urban road with low illumination at
30 fps. The optical flow was computed and the Canny algorithm
was applied to each image to detect the edges. Then each flow
was classified according to the procedure advised in section 3.
Figure 1a shows the original image and figure 1b shows an example of dense optical flow (5 x 5 needle diagram).
Considering the first sequence that corresponds to the first
second for both images (left and right) the reference velocity
was 19.83 km/h, as computed from the Ashtech Reliance GPS
}

Figure 1a. Original image

Figure 1b. Dense optical flow

Tables 1 until 4 show the average velocity and the standard deviation (SD) of the two sequences (T: 1 means the 1st second
and 2 means the 2nd second) considering the camera at left (L) independently from the camera at the right side (R). Each
sequence has 30 images. Table 1 shows the results computed using all the vectors in the region of interest.

Table 1. Estimated velocity from all vectors in the regions of interest.

<table>
<thead>
<tr>
<th>AV (km/h)</th>
<th>SD (km/h)</th>
<th>Nr. of vectors</th>
<th>SD (vectors)</th>
<th>Cam</th>
<th>T</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.57</td>
<td>0.54</td>
<td>230,186</td>
<td>0.00</td>
<td>L</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.15</td>
<td>0.37</td>
<td>230,186</td>
<td>0.00</td>
<td>R</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.83</td>
<td>0.76</td>
<td>230,186</td>
<td>0.00</td>
<td>L</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2.07</td>
<td>0.82</td>
<td>230,186</td>
<td>0.00</td>
<td>R</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12.48</td>
<td>7.83</td>
<td>489.59</td>
<td>145.15</td>
<td>L</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9.13</td>
<td>5.19</td>
<td>411.17</td>
<td>129.68</td>
<td>R</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8.02</td>
<td>2.35</td>
<td>697.03</td>
<td>202.56</td>
<td>L</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9.65</td>
<td>3.84</td>
<td>443.72</td>
<td>130.18</td>
<td>R</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

As the number of vectors are high (230,186), the velocity estimation has no good quality when compared to the reference
values (stage 1). The standard deviation is zero because all vectors of all 30 images of each side were used. In the second
stage, the number of vectors was reduced so another velocity value was computed although too low still far from the
reference values.
Table 2. Estimated velocity from the regions of interest and with only radial pattern vectors.

<table>
<thead>
<tr>
<th>AV (km/h)</th>
<th>SD (km/h)</th>
<th>Nr. of Vectors</th>
<th>SD (Vectors)</th>
<th>Cam</th>
<th>T</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.34</td>
<td>0.59</td>
<td>110,367</td>
<td>25972.80</td>
<td>L</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4.07</td>
<td>0.56</td>
<td>107,169</td>
<td>22780.10</td>
<td>R</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3.54</td>
<td>0.57</td>
<td>113,889</td>
<td>27590.10</td>
<td>L</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3.81</td>
<td>0.82</td>
<td>107,726</td>
<td>24595.40</td>
<td>R</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12.39</td>
<td>7.46</td>
<td>485.66</td>
<td>142.82</td>
<td>L</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9.29</td>
<td>5.04</td>
<td>412.90</td>
<td>129.78</td>
<td>L</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>8.13</td>
<td>2.18</td>
<td>705.38</td>
<td>210.07</td>
<td>R</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9.91</td>
<td>3.85</td>
<td>447.17</td>
<td>128.70</td>
<td>R</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

In the first stage of the second experiment, the velocity was computed with the radial vectors of the regions of interest so a little better value was estimated as compared to the first test. The velocity estimated in the second stage was close to the previous experiment. A possible partial conclusion may be that the second stage acts as filter that separates the radial pattern vectors.

The third experiment was done with the vectors which origins lay on points (pixels) detected by the Canny algorithm in the regions of interest no matter the pattern, radial or transversal. Table 3 shows the contribution of those vectors referred to the third experiment. It is evident that the estimated velocities are closer to the reference ones for both stages as compared to the previous experiments.

Table 3. Estimated velocity from the regions of interest and Canny.

<table>
<thead>
<tr>
<th>AV (km/h)</th>
<th>SD (km/h)</th>
<th>Nr. of vectors</th>
<th>SD (Vectors)</th>
<th>Cam</th>
<th>T</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.72</td>
<td>4.77</td>
<td>1190.55</td>
<td>120.12</td>
<td>L</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16.10</td>
<td>3.20</td>
<td>1093.66</td>
<td>77.72</td>
<td>R</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12.87</td>
<td>3.64</td>
<td>1694.07</td>
<td>195.19</td>
<td>L</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13.08</td>
<td>3.31</td>
<td>1260.41</td>
<td>55.55</td>
<td>R</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>18.19</td>
<td>4.67</td>
<td>511.52</td>
<td>154.67</td>
<td>L</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>17.83</td>
<td>3.65</td>
<td>480.07</td>
<td>161.78</td>
<td>R</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>13.67</td>
<td>4.42</td>
<td>735.52</td>
<td>221.31</td>
<td>L</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>15.00</td>
<td>3.02</td>
<td>524.62</td>
<td>152.01</td>
<td>R</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Finally, table 4 presents the results obtained with the procedure proposed in this article.

Table 4. Estimated velocity with the proposed methodology.

<table>
<thead>
<tr>
<th>AV (km/h)</th>
<th>SD (km/h)</th>
<th>Nr. of vectors</th>
<th>SD (Vectors)</th>
<th>Cam</th>
<th>T</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.06</td>
<td>3.26</td>
<td>949.17</td>
<td>290.94</td>
<td>L</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>18.78</td>
<td>1.95</td>
<td>809.59</td>
<td>263.71</td>
<td>R</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14.53</td>
<td>3.27</td>
<td>1243.86</td>
<td>376.10</td>
<td>L</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>14.34</td>
<td>2.83</td>
<td>839.24</td>
<td>252.34</td>
<td>R</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>19.67</td>
<td>2.96</td>
<td>510.83</td>
<td>153.35</td>
<td>L</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>19.90</td>
<td>1.96</td>
<td>482.14</td>
<td>162.35</td>
<td>R</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>15.88</td>
<td>3.96</td>
<td>735.31</td>
<td>222.83</td>
<td>L</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>16.51</td>
<td>1.68</td>
<td>520.10</td>
<td>152.77</td>
<td>R</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

The filtering methodology reduces the amount of vectors and it improves the velocity estimation in both stages. The amount of vectors in the second stage is lesser than that of the first stage so the estimation is even better. This suggests that the remaining vectors are those that actually contain the implicit information related to the vehicle displacement and also that the eliminated vectors are those that really disturbed the estimation process.

The average velocity is computed from the left and the right image velocities and it is attributed to all images of the sequence. Then the approximate relative position of every image is computed based on the constant sampling rate (30 fps) or 1/30th of the second.

In the first image pair, a set of 17 homologous points was manually measured and the photocoordinates were determined. Together with an arbitrary position and orientation for the first pair, the object coordinates of the 17 points were estimated by least square photogrammetric intersection. A forward correlation between two images at t_1 (1st basis) and t_2 (2nd basis) for both left and right sequences (figure 2) is computed. p(t_1) is one of the 17 image points determined in the first pair; p(t_2) is the correspondent point in the second pair of the sequence. The position of p(t_2) in the image plane is given by the collinearity equations, the correspondent object coordinates of P, and the approximate orientation of the second pair. The perspective center coordinates of the second pair are estimated from the displacement given by the use of the estimated velocity and the 1/30 time interval. The target window is centered at p(t_1) and the search window at p(t_2). The best correspondence indicates the definitive image point p(t_2) that corresponds to p(t_1).
The results that follow were obtained with seven bases (or stereo pairs) which mean fourteen images. The bases were taken every 5/30 s. With all the seven bases the equation system had 476 observation equations, 51 ground coordinates, 7 basis constraints, and 6 velocity constraints. Of course, the adjustment model was composed by the linearized and the constraint equations. The basis constraint used the base length equals to 0.94m and the velocity constraint used the estimated velocity values. Figure 4 shows the sigma zero values accordingly to the experiments.

Table 6 shows the results related to the adjusted bases and velocity (m/s). The computed velocity does not apply to the first basis.

### Table 6. Adjusted bases and velocities.

<table>
<thead>
<tr>
<th>Bases</th>
<th>(m)</th>
<th>(m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.93996</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.93997</td>
<td>5.48980</td>
</tr>
<tr>
<td>3</td>
<td>0.94009</td>
<td>5.49010</td>
</tr>
<tr>
<td>4</td>
<td>0.93998</td>
<td>5.48999</td>
</tr>
<tr>
<td>5</td>
<td>0.93976</td>
<td>5.49000</td>
</tr>
<tr>
<td>6</td>
<td>0.93996</td>
<td>5.48999</td>
</tr>
<tr>
<td>7</td>
<td>0.93996</td>
<td>5.48999</td>
</tr>
</tbody>
</table>

### 6. CONCLUSION

The goal of this work was to estimate the positioning and orientation of a road image sequence of stereoscopic pairs without using external sensors or control points. Monocular optical flow was used to compute the velocity and then the displacement between every two consecutive bases. The filtering applied to select only the radial pattern vectors and also those with origins at pixels detected by the Canny algorithm was a strategy that produced a good estimation for the velocity. The basis and the velocity constraints provided improved result as well as the adjustment. In future work it is intended to eliminate the manual measuring of the homologous points in the first basis. The idea is to use a sample of the filtered points. This will represent an attempt to a totally automated methodology.

### REFERENCES


ACKNOWLEDGEMENT

The authors thank FAPESP for the financial support related to this research project (Process FAPESP 03/00552-1).