LINE SEGMENT BASED STRUCTURE AND MOTION FROM TWO VIEWS
A Practical Issue

Saleh Mosaddegh, David Fofi
Le2i, UMR CNRS 5158, Université de Bourgogne, Le Creusot, France

Pascal Vasseur
MIS, Université de Picardie Jules Verne, 33 Rue Saint Leu, 80039 Amiens, France

Keywords: Motion estimation, Line segment correspondences, Two views, Optimization.

Abstract: We present an efficient measure of overlap between two co-linear segments which considerably decreases the overall computational time of a Segment-based motion estimation and reconstruction algorithm already exist in literature. We also discuss the special cases where sparse sampling of the motion space for initialization of the algorithm does not result in a good solution and suggest to use dense sampling instead to overcome the problem. Finally, we demonstrate our work on two real data sets.

1 INTRODUCTION

Using lines for estimating motion is advantageous because they are easier to extract and have less localization error than other features of interest such as points. However, one can find several works in literature in which the impossibility of motion determination from the line correspondences between only two views are mentioned (Zhang, 1995; Holt and Netravali, 1996; Netravali and Huang, 2002; Faugeras, 2001). Classical methods such as (Taylor and Kriegman, 1995; Bartoli and Sturm, 2005) which use supporting lines (geometric abstraction of straight line segments) need many line correspondences across at least three images. To our knowledge, the algorithm introduced in (Zhang, 1995) is, so far, the only work based on only two views which tries to recover motion and structure by maximizing the total overlap of line segments in correspondence. In this paper we introduce a unique and efficient way of computing overlap between two segments which considerably decreases the overall computational time of this algorithm. We improve the speed of proposed method by replacing its objective function with a less computationally expensive one without affecting the output of the algorithm. For all our data sets in hand, it was also found that the sampling strategy of the proposed method often is not dense enough to obtain a good initial guess and one need to sample the motion space with very small steps to obtain an acceptable solution. On the other hand, after densely sampling of the motion space, the best initial guess is already enough close to the good solution and further optimization does not improve the estimation considerably. This observation also motivated us to work on decreasing the time for calculating the objective function in order to be able to search for a good solution over a densely sampled motion space in a shorter time. The rest of this text is organized as follows. First a brief explanation of the Zhang method as it is introduced in (Zhang, 1995). In section 3 the new objective function is defined and its performance compared to the previous function is discussed. The results in section 5 demonstrate how this new function lead to a significant improvement in performance in terms of execution time through applying the method on a set of real data.

2 MOTION ESTIMATION BY MAXIMIZING OVERLAPS

In this section, we present a brief summary of the algorithm for solving the motion problem by maximizing the overlap of line segments introduced in(Zhang, 1995). The problem to be solved is that given the

Mosaddegh S., Fofi D. and Vasseur P..
LINE SEGMENT BASED STRUCTURE AND MOTION FROM TWO VIEWS - A Practical Issue.
DOI: 10.5220/0003357106800684
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2011), pages 680-684
ISBN: 978-989-8425-47-8

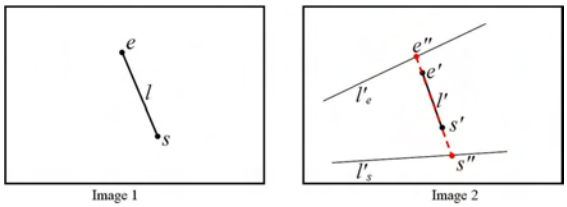

Figure 1: Overlap of two line segments in correspondence.

cameras intrinsic parameters and two sets of line segments, which are in correspondence, estimate the camera extrinsic parameters (motion R and t).

Consider the pair of line segments(l, l') in correspondence as shown in Fig. 1. The line l'_s in the second image is the epipolar line of end point s from the first image, *i.e.* $l'_s = E\widetilde{s}$, where $E = [t]_\times R$ is the essential matrix (Hartley and Zisserman, 2004). $\widetilde{s}$ is the ray which pass through end point s and center of first camera and it can be easily computed since the camera intrinsic parameters are assumed to be known. Similarly the line l'_e is the epipolar line of the other end point e. Taking cross product of each of these two epipolar lines with the segment $e's'$ results in their intersection, s" and e" with the segment. Provided that the epipolar geometry (i.e. matrix E, or the motion $(R\ ,\ t)$) between two images is correct, then s and s" correspond to a single point in space; so do e and e". Thus, the statement that two line segments l and l' share a common part of a 3D line segment is equivalent to saying that line segment s"e" and line segment $s'e'$ (*i.e.*l') overlap. The overlap length, $\mathcal{L}'$, for two line segments in correspondence can then be computed from:

$$\mathcal{L}' = \begin{cases} min(||e'-s'||, ||e"-s'||, ||e'-s"||, ||e"-s"||) \\ \qquad if \begin{cases} (s"-s').(e'-s") > 0 \\ or\ (s"-s').(e'-s") > 0 \\ or\ (s"-s').(e'-s") > 0 \\ or\ (s"-s').(e'-s") > 0 \end{cases} \\ -min(||e'-s"||, ||e"-s'||) \\ \qquad otherwise \end{cases} \tag{1}$$

where . stands for dot product of two vectors. The overlap length is positive if two line segments overlap, otherwise it is negative. We assume that the orientation information of a line segment is not available (*i.e.* the correspondence between end points of the segments is not known). The overlap length in the first image, denoted by $\mathcal{L}$, can be computed exactly in the same way . Since a small overlap length for a short line segment is as important as a large overlap length for a long line segment, we should use the relative overlap length, $\mathcal{L}'/||l'||$ and $\mathcal{L}/||l||$ to measure the overlap of the pair of line segments. The relative overlap length takes a value between 0 and 1 when two segments overlap; otherwise it will be negative. We define relative non-overlap length between two corresponding segments l_i and l'_i in the second image as:

$$\mathcal{H}_i = (1 - \frac{\mathcal{L}'_i}{||l'_i||}) \tag{2}$$

which is 0 when two segments completely overlap, between 0 and 1 when they partially overlap and bigger than one when there is a gap between two segments. We can now formulate the motion problem as estimating the camera motion parameters $(R\ ;\ t)$ by minimizing the following non-linear objective function:

$$\mathcal{F} = \sum_{i=1}^{n} (\mathcal{H}_i + \mathcal{H}'_i)$$

where n is the number of lines. The algorithm can be summarized as the following pseudo-code:

Sample the rotation and the translation space with sufficient steps.
for each sample $R(i)$ in the rotation space **do**
  for each $t(j)$ in the translation space **do**
    For hypothesized motion $E = [t(j)]R(i)$ calculate objective function $\mathcal{F}_0 = \sum_{k=1}^{n} (\mathcal{H}_k + \mathcal{H}'_k)$
  end for
end for
for 10 best solution in matrix $\mathcal{F}_0$ **do**
  Using downhill simplex method, minimize $\mathcal{F}$ starting with the best solution as initial guess.
end for

If sampling of motion space is with adequate small steps, at least one of the ten minimization efforts in the last loop converges to a good solution.

3 THE NEW MEASURE OF OVERLAP

In above algorithm, $\mathcal{H}'$ (or $\mathcal{H}$), the function for computing relative non-overlap length for each line correspondence is the most frequently called function and reducing computational time of this function can largely decrease the overall computational time of the algorithm. Consider the two possible configurations of two collinear line segments as shown in Fig. 2. The coordinates of two endpoints of the overlap part, $(Xmin, Ymin)$ and $(Xmax, Ymax)$ can be found by:

$X_{min} = max(min(s'_x, e'_x),\ min(s"_x, e"_x))$,
$X_{max} = min(max(s', e'),\ max(s"_x, e"_x))$,

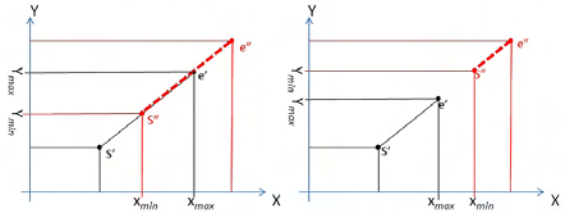

Figure 2: Two possible configurations of two collinear line segments.

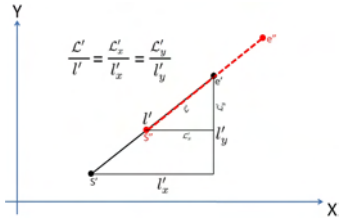

Figure 3: The relation between Cartesian components of the overlap part and the segment in the second image. The ratios of corresponding sides of two right triangles are constant.

$Y_{min} = max(min(s'_y, e'_y),\ min(s"_y, e"_y))$,
$Y_{max} = min(max(s'_y, e'_y),\ max(s"_y, e"_y))$,
therefore, overlap length can be expressed by its Cartesian length by:

$$\mathcal{L}'_i = \mathcal{L}'_{ix} + \mathcal{L}'_{iy}$$

where

$$\mathcal{L}'_{ix} = (X_{max} - X_{min}), \quad \mathcal{L}'_{iy} = (Y_{max} - Y_{min})$$

Note that the output of this new measure of overlap length is exactly equal to that of Equ. 1 but without need for a $if-then$ construct with four OR conditions. While computing the relative non-overlap length, computational time can further be reduced by half by considering only one of the Cartesian components of the overlap part and the segment in the second image (we chose x component) based on the relation shown in Fig. 3:

$$\mathcal{H}_i = (1 - \frac{\mathcal{L}'_{ix}}{||l'_{ix}||}) \tag{3}$$

However care should be taken when the segment is vertical where y components should be used to avoid undefined division $0/0$. In order to have a very accurate comparison between two measures, we carefully counted the number of CPU cycles needed to run the assembly instructions for the new non-overlap length measure as defined by Equ. 3 versus the measure defined by Equ. 2 compiled using a C compiler on an Intel Pentium machine. Our new non-overlap length measure needs 302 clockticks (including the *conditional if* for proper treating of vertical segments). The original measure of non-overlap needs a minimum of 720 clockticks (if the first inequality condition among four inequalities in Equ. 1 is satisfied) and a maximum of 858 clockticks (if none of four inequalities are satisfied). We can not use clockticks average since for the majority of samples in motion space and except for some random lines, the rest of lines do not exhibit an overlap therefore the computation of the objective function for these samples requires maximum number of clockticks. This means our new measure can be computed on average slightly less than 858/302 = 2.841 times faster than the measure introduced in (Zhang, 1995), assuming that the variation of the line segments is random. Refer to the result section for a comparison using real data.

4 RECOVERING MOTION BASED ON DENSE SAMPLING

Sparse sampling of the 5 dimensional motion space (3 for the rotation and 2 for the translation) followed by refinement of the best samples as suggested in (Zhang, 1995) is problem-in-hand dependent and depending on how far the best initial guesses are to the global minimum, the optimization stage can use considerable iterations to converge to a good solution or it may not be able to converge at all. Thanks to our faster method for calculating the objective function, we are able to sample the motion space more densely, resulting in a better initial guess closer to the global minimum with less time required by optimization algorithm to converge to a good solution. The results in the next section demonstrate how this new approach can help to recover the motion for the examples where the sparse sampling followed by a refinement of the best samples cannot converge to a good solution.

5 RESULTS

We have already shown the efficiency of the new objective function in the terms of execution cycles. In this section, however, we give the results on two real data sets where for the last set the original algorithm fails to recover the motion due to sparse sampling of the motion space. The first set of real data is an image pair of a bakery (Fig. 4). The position and rotation of the second camera with respect to the first one was obtained through a very careful setup and use of a gyroscope:

$R = [-0.0073, -0.3049, -0.0036]$,

$T_{tra} = [0.9318, -0.0123, 0.3629]$,

where the translation T_{tra} is normalized and the rotation R is represented by a 3D vector whose direction is that of the rotation axis and whose norm is equal to the rotation angle. The segments which are aligned with the epipolar lines are neglected during computing total overlap and later for the scene reconstruction since in this case computed intersections, s and s" are instable and the result overlap can be irrationally big, resulting in eliminating a good solution.

For speed comparison, we applied the algorithm with both original and new objective functions on this data. We extracted and matched 85 lines between two views manually. Through searching for the initial motion estimation by the sampling strategy as described in the original algorithm (*i.e.* sampling the range $[-\frac{\pi}{4}, \frac{\pi}{4}]$ with steps equal to $\frac{\pi}{8}$ for the rotation and 40 uniform sampling of a Gauss hemisphere based on the icosahedron for the translation), only 1 of 10 best samples converged to the good solution. The result of the best solution is shown in Fig. 4. The whole process (excluding the time for line extraction and matching) took 3159 seconds composed of 1479 seconds for evaluating the objective function over the motion space and 1680 seconds for optimization of 10 best initial guesses (both algorithm were implemented in Matlab and were executed on the same computer).

Figure 4: Top row: Two images of a bakery with 85 matched line segments superimposed on the images (in green) . Bottom row: 3D reconstruction of the bakery by the structure from motion technique described in (Zhang, 1995). The right image corresponds to the top view.

Though replacing the function for computing relative non-overlap with our function does not alter the output of the new algorithm, however it reduces the overall computational time to 1215 seconds (around 2.6 times faster including all overhead computations). The error in the translation direction is 1.0847°. The error in the rotation angle is 0.3087° and the error in the rotation axis is 1.9147°.

Fig. 5 shows the second pair of images taken from the real stereo image data set available at (INRIA,). The transformation from the first camera to the second camera is:

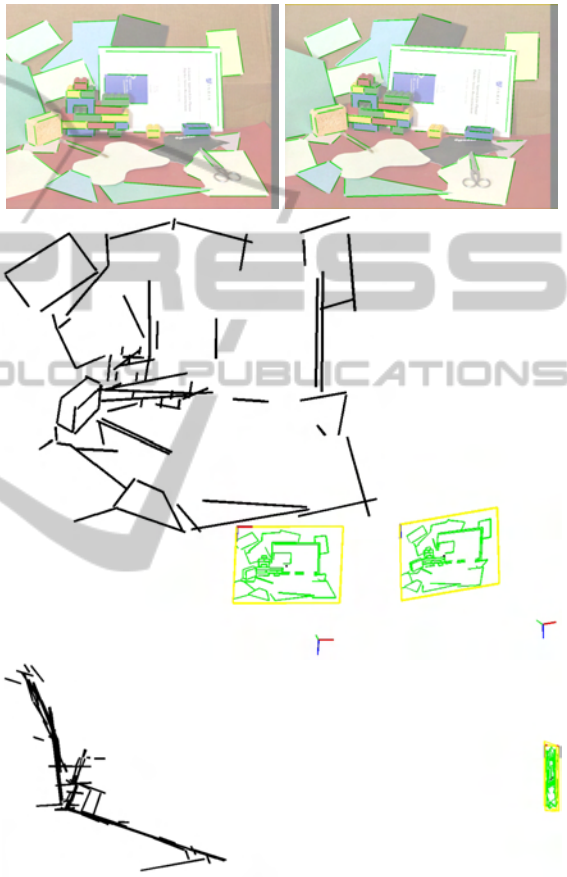

Figure 5: Top row: A stereo pair with 104 matched line segments superimposed on the images (in green) . Middle row: A perspective view of the 3D reconstruction by classical stereo including two camera image planes. Bottom row: a view from the side.

$R = [-0.0004, 0.3133, 0.0717]$,

$T_{tra} = [-0.9859, -0.0441, 0.1617]$,

We extracted and matched 104 lines between two views manually. Through searching for the initial motion estimation by original sampling, none of the best samples converged to a good solution. The best solution reconstruction corresponding to the initial guess with the smallest value of objective function is shown in Fig. 6.

As a matter of fact, this is an example of a scene

Figure 6: 3D reconstruction of the scene by the best solution of the structure from motion technique described in (Zhang, 1995). The bottom image corresponds to a side view. Apparently this is not a good solution.

Figure 7: 3D reconstruction of the scene corresponding to the sample with minimum value of objective function from densely partitioned motion space.

where it can be shown that the global minimum is closely located to many local minima and only a fine sampling of the motion space can result in a good solution. Therefore we applied our dense sampling strategy by 90 sampling of translation space and 1330 sampling of rotation space. The evaluation of the objective function for all these samples takes around 3120 seconds. The best solution's reconstruction is shown in Fig. 7. The error in the translation direction is 1.9°. The error in the rotation angle and rotation axis are 0.94° and 1.986° respectively. This result is already very good and further optimization is not necessary. One can notice that even though we are benefiting from a faster objective function, however the evaluation of all samples in the dense motion space is quite time consuming and except inevitable evaluation of all these samples, there is not any other deterministic alternative approach for such particular examples.

6 CONCLUSIONS

We introduce a new measure of overlap which increases the speed of the calculating the overlap between two line segments in correspondence. It also allows a denser sampling of the motion space for finding initial guesses for optimization of the non-linear objective function for recovering motion based on line segment correspondences and therefor facilitating the search for a good solution where due to the nature of the scene, sparse sampling followed by optimization does not converge to a good solution. We demonstrated this situation by giving the results on two real data sets including the scene where the original algorithm fails to recover the motion.

REFERENCES

Bartoli, A. and Sturm, P. (2005). Structure-from-motion using lines: Representation, triangulation, and bundle adjustment. *CVIU*, 100(3):416–441.

Faugeras, O. (2001). *Three-dimensional computer vision : a geometric viewpoint.* MIT Press, Cambridge, Mass.

Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Holt, R. and Netravali, A. (1996). Uniqueness of solutions to structure and motion from combinations of point and line correspondences. In *JVCIR*, volume 7, pages 126–136.

INRIA. Syntim stereo images. http://perso.lcpc.fr/tarel.jean-philippe/syntim/paires.html.

Netravali, A. and Huang, T. (2002). Motion and structure from feature correspondences: A review. In *AIPU02*, pages 331–348.

Taylor, C. and Kriegman, D. (1995). Structure and motion from line segments in multiple images. *PAMI*, 17(11):1021–1032.

Zhang, Z. (1995). Estimating motion and structure from correspondences of line segments between two perspective images. In *ICCV95*, pages 257–262.