

Multi-Camera Very Wide Baseline Feature Matching Based On View-Adaptive Junction Detection

Maykel Pérez, Luis Salgado, Jon Arróspide, Javier Marinas
Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid
Madrid, Spain 28040
Email: {lsa,l.salgado,jal,jmm}@gti.ssr.upm.es

Marcos Nieto
Vicomtech-IK4, Research Alliance
Donostia-San Sebastián, Spain 28009
Email: mnieto@vicomtech.org

Abstract—This paper presents a strategy for solving the feature matching problem in calibrated very wide-baseline camera settings. In this kind of settings, perspective distortion, depth discontinuities and occlusion represent enormous challenges. The proposed strategy addresses them by using geometrical information, specifically by exploiting epipolar constraints. As a result it provides a sparse number of reliable feature points for which 3D position is accurately recovered. Special features known as junctions are used for robust matching. In particular, a strategy for refinement of junction end-point matching is proposed which enhances usual junction-based approaches. This allows to compute cross-correlation between perfectly aligned plane patches in both images, thus yielding better matching results. Evaluation of experimental results proves the effectiveness of the proposed algorithm in very wide-baseline environments.

I. INTRODUCTION

The knowledge of the 3D position of relevant points in the scene is of great interest for many vision-based applications, such as object detection, tracking, and recognition. This can be obtained by establishing relationships among different elements in the scene, via feature matching strategies operating on images acquired synchronously from different viewpoints.

Feature matching strategies are usually classified into area-based and feature-based methods. In the former, the matching process is applied directly to the intensity, color or texture of the neighborhood of the candidate regions, which are typically compared through cross-correlation methods [1][2]. In contrast, feature-based methods rely on an initial extraction of relevant features, and the matching is performed upon them. Area-based methods show very good performance in short baseline settings, where illumination is similar in the two views and the corresponding regions are expected to be in a close neighborhood, while feature-based strategies are usually applied to medium or wide baseline settings. Particularly relevant is the proposal by [3], which aims to be invariant to view dependent deformations.

However, when it comes to very wide baseline settings (i.e. featuring large inter-camera distances: viewpoints change more than 60 degrees) none of the aforementioned approaches has proven to be successful. In particular, feature-based methods are prone to errors due to the severe illumination changes, perspective distortions, depth discontinuities and occlusions between views [4][5]. Among the methods proposed for this

kind of settings, the use of junctions as robust features to address matching seems to be the most promising alternative [6][7]. In particular, in [7] they propose to analyze the similarity through correlation techniques on areas around junctions. The features are previously detected using a fixed operator size. However, this is only valid when features in both views represent a 3D junction that lies at similar distance from both cameras, which is usually not the case for very wide baseline cameras. Another critical issue is the precision in the image regions used for correlation. Those are bounded by the junction edges and their corresponding end-points. Typically, the end-points delivered by the junction detector in both images are considered to be correspondent, thus neglecting the fact that they do not generally belong to the same 3D points. Hence, non-equivalent areas are used for correlation and the similarity measures are unreliable.

In this work we present a new feature matching approach that overcomes these limitations. First, the proposed strategy, which also relies on junctions as relevant features, adapts the search area of the feature detector in the images according to the analysis of the scene geometry, as opposed to typical approaches that perform independent feature extraction in the two images. As a result, the loss and mismatch of features between images is reduced. Additionally, the strategy involves a novel refinement stage that precisely computes the corresponding junction end-points. Therefore, matching of candidates, which is performed by computing appearance similarity measures in the regions defined by the junction edges, is enhanced with respect to traditional methods, since the regions for cross-correlation computation are equivalent in both images.

II. ADAPTIVE FEATURE EXTRACTION

As stated in the introduction, junctions are regarded as robust features when it comes to matching in very wide baseline settings. A junction is defined as an image point where several edges meet [8]. In other words, junctions in images occur when several nearly uniform regions join at one prominent point, (i.e., the point of junction) where the boundaries of the adjacent regions meet. A junction is thus determined by its center, the number of converging edges, and their respective orientations.

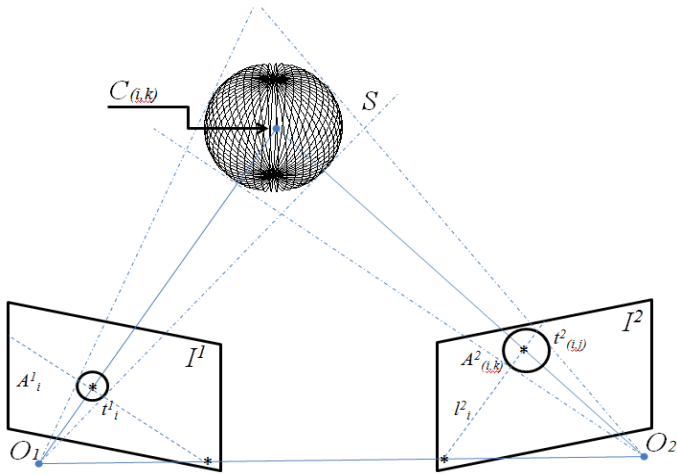


Fig. 1. Two-view geometry analysis for adaptation of λ_k^2 .

Existing detectors generally search junctions on circular areas of radius λ in the image, hence this value is of great importance for the behavior of the detector. This is the case for the detector introduced in [8], which is used for feature detection in this work. Naturally, in a stereo pair, the area of projection of a feature detected in the first image onto the second image depends on the relationship between cameras and on the 3D location of the feature with respect to them. As opposed to the traditional approach, which assumes a uniform circular search area of the same radius λ in the second image, in this section we explain how to make this area adaptive.

Consider a junction t_i^1 detected in a circular region A_i^1 in the reference image I^1 . The area A_i^1 holds the projection in the image of the 3D junction, and therefore the radius of this area defines a cone which encompasses the junction in the 3D space. Suppose the junction is centered at the point $C_{(i,k)}$ on the 3D beam defined by the camera optical center O_1 and the center of t_i^1 , then we define S as the largest sphere centered in $C_{(i,k)}$ and contained in the projection cone. The projection of this sphere will help us approximate the shape and size of the junction in I^2 . This idea is illustrated in Fig. 1. The projection of S in I^2 is given by the cone tangent to S with vertex in O_2 , which defines an elliptical region A^2 in I^2 .

In particular, suppose we have a junction centered in p_i^1 in I^1 , detected in an area A_i^1 of radius λ^1 . In order to find this junction in I^2 , we will use the idea explained above to define the search area (i.e., the radius of the junction detector) in the second image. Namely, we first make use of the epipolar geometry, which constrains the center of the junction in I^2 to be in the so called epipolar line [9], as illustrated in Fig. 2. Each point $p_{(i,k)}^2$ of the epipolar line l_i^2 constitutes a matching candidate for t_i^1 . Thus, for each pair of point correspondence candidates, $\{p_i^1, p_{(i,k)}^2\}$, a sphere $S_{(i,k)}$ will be generated centered in the intersection of the beams defined by these points, $C_{(i,k)}$. The radius of the sphere is computed as the minimum distance from its center to the beams defined by O_1 and the contour of A_i^1 .

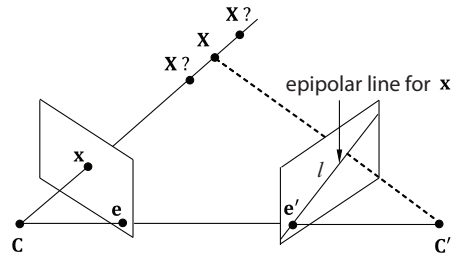


Fig. 2. Epipolar geometry: the point corresponding to x in I^1 is contained to be in the epipolar line l in I^2 .

The area $A_{(i,k)}^2$ of projection of $S_{(i,k)}$ in I^2 is obtained iteratively, since analytic derivation of $A_{(i,k)}^2$ involves the explicit knowledge of camera parameters (particularly focal length, which is usually not available). Namely, the points in the surroundings of $p_{(i,k)}^2$ are back-projected into 3D space using the projection matrix P_2 (camera calibration is available), and we check whether they intersect or not with $S_{(i,k)}$. The points that do indeed intersect with the sphere are enclosed within an elliptic pattern. In this paper, a junction detector based on circular search area is used, therefore we will approximate $A_{(i,k)}^2$ as a circular area with a radius λ_k^2 equal to the minor radius of the ellipse (this will also reduce the number of iterations, as we stop whenever one back-projected point does not intersect the sphere). The minor axis of the ellipse is taken as radius of $A_{(i,k)}^2$ so that the sensibility of the detector is higher and the probability that the junction is detected is maximized. Nevertheless, note that the whole ellipse can also be available in case a junction detector is proposed that is designed to cope with it.

Once λ_k^2 has been computed for each candidate point, $p_{(i,k)}^2$, the junction detector is applied using the corresponding adapted radius over each $p_{(i,k)}^2$ of the epipolar line l_i^2 . As a result, a set of candidate junctions in the second image, $t_{(i,j)}^2$, is obtained for the initial junction in I^1 , t_i^1 . The subindex j is used hereafter instead of k to denote that the set of candidate features is now smaller, since only the points of the epipolar line for which the detector finds a junction are kept (if any; recall that it might not be visible due to occlusion, for instance). The intersections of the edges of a candidate junction with the contour of $A_{(i,k)}^2$ constitute the coarse end-points of the junction, $\{\tilde{e}_{(i,j)}^2\}$.

III. FEATURE REFINEMENT

A set of candidate matches in the second image are now available for each of the junctions in the first image as a result of the adaptive junction search explained in the previous section. Given these candidates, most approaches apply directly some correlation technique in order to find the final set of correspondences. However, this procedure is intrinsically inaccurate. Indeed, the end-points of the junction axes are usually non-corresponding (that is, they do not correspond to the same 3D points). Hence, when directly

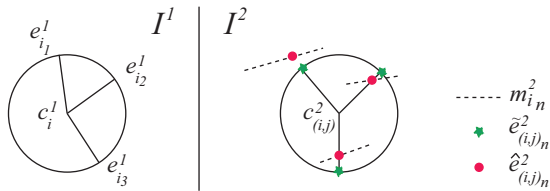


Fig. 3. Illustration of end-point refinement process.

performing matching between them, the surfaces used for correlation are not correctly delimited, thus leading to errors in feature correspondences.

Therefore, given a reference junction and a candidate one in the other image, feature refinement targets, on the one hand, to efficiently establish correspondence between the axes of these junctions, and on the other hand, to refine the position of the end-points of the candidate junction so that they match the actual end-points of the reference one. This way, cross-correlation is restricted to corresponding areas in both images, which helps to improve matching results.

Feature refinement is also addressed through epipolar geometry. Suppose that we have a junction in the first image, centered at c_i^1 , and with N edges, ending at $e_{i_1}^1, e_{i_2}^1, \dots, e_{i_N}^1$. One of the hypothesized correspondences for this junction in I_2 is centered at $c_{(i,j)}^2$, and has its coarse end-points in $\{\tilde{e}_{(i,j)_n}^2\}_{n=1}^N$, as illustrated in Fig. 3 for a junction of $N = 3$ edges. These end-points do not accurately correspond to $\{e_{i_n}^1\}_{n=1}^N$; however, we can now exploit epipolar geometry constraints to correct the end-point positions. Indeed, it is known that the point $e_{i_n}^1$ is projected into the second image to a point within the epipolar line $l_{i_n}^2$ (this can be further reduced to a segment $m_{i_n}^2$ by introducing geometrical constraints, e.g., that the junction is bounded by planes tangent to the sphere S), as illustrated by the dashed segments in Fig. 3.

Hence, the end-points $\{\tilde{e}_{(i,j)_n}^2\}$ can be corrected by finding the intersection points between the epipolar segments $m_{i_n}^2$ and the lines containing the junction edges in I_2 (see Fig. 3). As proven in [9], the line joining two points is given by the cross-product of these points in homogeneous coordinates. Dually, the intersection of two lines is the cross-product of the lines. Therefore, the refined end-point positions, $\{\hat{e}_{(i,j)_n}^2\}_{n=1}^N$ are computed as follows. First, the lines $b_{(i,j)_n}^2$ containing each of the edges of the junction in the second image are computed as the cross-product of the center and their corresponding coarse end-point, i.e., $b_{(i,j)_n}^2 = c_{(i,j)}^2 \times \tilde{e}_{(i,j)_n}^2$. Then, the lines $l_{i_n}^2$ are found using epipolar geometry as explained above. Finally, the refined end-points $\hat{e}_{(i,j)_u}^2$ of the junction in the second image are given by the intersections of these lines, $\hat{e}_{(i,j)_u}^2 = b_{(i,j)_u}^2 \times l_{i_u}^2$. In principle, an edge can intersect with the epipolar line of any of the end-points in the first image, thus, since the junction contains N edges, there are up to $N!$ possible combinations. However, if the intersection occurs out of the segment $m_{i_n}^2$ the hypothesis is disregarded, which is often the case, resulting in a smaller set of combinations (typically one). Each set of end-points, together with their



Fig. 4. Example pairs of images: above, Laboratory; below, Office.

center, constitute a candidate correspondence for the original junction in the first image. All possible correspondences are assessed through the method explained in Section IV.

IV. FEATURE MATCHING

The feature matching step is carried out based on a cross-correlation technic for comparing the corresponding regions around reference and candidate junctions. The correlation method used here is an adaptation of the classical Sum of Absolute Differences (SAD). The coordinates of the points belonging to regions defined by the center of the junction and a pair of axes are parameterized as in [7], in order to compare corresponding points between candidate regions. This way, the matching process adapts to the effect of projective deformation and it thus circumvents problems regarding window size and shape selection, typical for area-based methods. Therefore, the cross-correlation is measured as:

$$D = \sum_{(x^1, y^1) \in A_i^1} |I^1(x^1, y^1) - I^2(x^2, y^2)| \quad (1)$$

where $(x^2, y^2) \in A_{(i,j)}^2 = T_x(x^1, y^1)^\top$, and T_x is the parametric transformation mapping the points in A_i^1 to $A_{(i,j)}^2$. The correlation is maximum when $D = 0$. For each junction in I^1 , the candidate that minimizes D is assigned to it as long as this value is below a predefined threshold; otherwise, it is left unmatched.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Experiments have been performed on a set of stereo images featuring a wide baseline. Two pairs of this set are shown in Fig. 4, the first of them having a viewpoint difference of 90 degrees between cameras.

The effectiveness of the approach proposed for adapting the search area in the second image is shown in Fig. 5 for the 'Laboratory' example. Specifically, a detector with $\lambda^1 = 11$ causes a junction to be detected in the left image as shown in Fig. 5(a). If the same detector were applied in the right image, the junction would not be detected, since the size and relative pose of the junction edges vary (see Fig. 5(b)). In this particular example, the scale of the junction in the right image

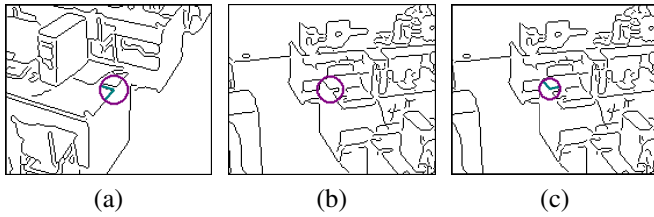


Fig. 5. Behavior junction detector regarding λ_k^2 adaptation. In (a) a reference junction is detected in the left image of the office environment with a specific λ^1 ; (b) the corresponding junction is not detected in the right image with $\lambda_k^2 = \lambda^1$; (c) it is successfully detected with an adapted $\lambda_k^2 < \lambda^1$.

is smaller, and the detector does not have enough sensibility to detect it. By using the described approach, we are able to infer the correct scale of the junction and to adapt the radius of the detector (in this case $\lambda_k^2 = 9$), thus detecting the corresponding junction, as shown in Fig. 5(c).

Feature refinement constitutes as well an important source of improvement. Indeed, once junctions are correctly assigned, the end-points of their edges do not correspond to the same 3D point, due to the change in the pose of the junction. Therefore, the regions for correlation do not fully overlap, which leads to errors in correspondences, especially if the regions present non-uniform intensity profiles.

The improvement achieved by using the feature refinement step is exemplified in Fig. 6 and Fig. 7 for the 'Office' images. In particular, Fig. 6(b) illustrates the refinement process over the initially obtained end-points of a junction in the right image. In particular, the two junction's end-points in I_1 (i.e., the intersection of the circumference and the red edges in Fig. 6(a)) define two corresponding segments in I^2 using epipolar geometry (those are painted as dashed purple lines in Fig. 6(b)). Therefore, the initial end-points in Fig. 6(b) are refined by selecting the intersection between the initial edges - painted in yellow- and the aforementioned segments. The final junction is painted in red, and corresponds to the same region as that in Fig. 6(a), as can be observed in a detail of this junction in zoomed left and right images (see Fig. 7). The areas defined by the axes of the junction detected in the left image and by its corresponding non-refined axes in the right image are painted in green. Additionally, the area associated to the refined axes is painted in yellow in the right image. Cross-correlation of the left region with both the refined and the non-refined right regions is computed, yielding values of $D = 490$ and $D = 520$. This difference is very significant taking into account the homogeneity of the regions to be correlated: regions meeting in junctions by definition always show a homogeneous texture. Hence, the capability of finding the correct correspondence is highly improved.

VI. CONCLUSION

The exploitation of epipolar geometry has been proven to constitute a suitable approach to address feature matching in very wide baseline stereo images, where traditionally used approaches fail. This approach allows i) to obtain a potential 3D junction location corresponding to two features from different images, ii) to adapt the parameters of the detector

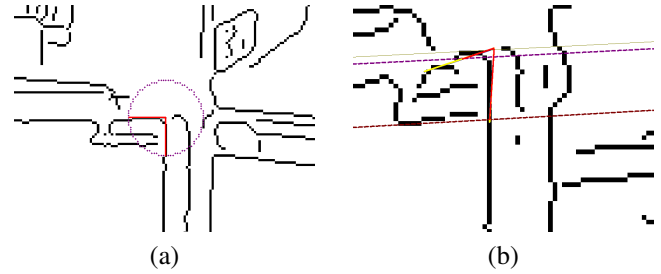


Fig. 6. Refinement step for a candidate junction: (a) reference detected junction; (b) candidate junction refinement.

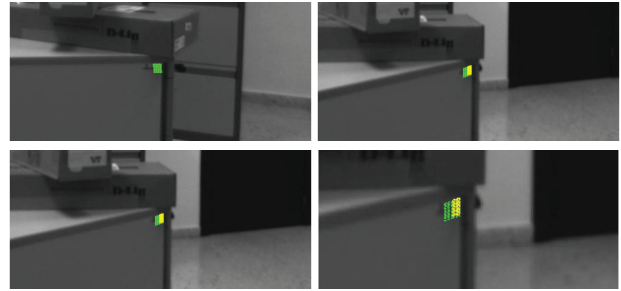


Fig. 7. Regions covered by refined and non-refined axes in Fig. 6. The lower row is a zoom of the upper row for better observation of details. In the left image, the area in green corresponds to the two edges of the reference junction. In the right image, the area in green is defined by the non-refined edges of the corresponding junction, while the area in yellow is defined by the refined axis. The refined (yellow) region matches the reference area better.

for finding candidate junctions given a reference one, and iii) to refine the candidate junctions end-points for improving the accuracy of the matching step. Thus, the strategy proposed in this paper results in an overall improvement of the feature matching process in very wide baseline settings.

ACKNOWLEDGMENT

This work was supported in part by the Ministerio de Ciencia e Innovación of the Spanish Government under projects TEC2010-20412 (Enhanced 3DTV) and TEC2007-67764 (SmartVision).

REFERENCES

- [1] H. Hirschmüller, P. R. Innocent, and J. Garibaldi, "Real-time Correlation-based Stereo Vision with Reduced Border Errors," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 229-246, 2002.
- [2] R. Laganière and F. Labonté, "Stereokineopsis: A survey," Tech. Rep. GRPR-RT-9603, 1996.
- [3] H. Hirschmüller, P. R. Innocent, and J. Garibaldi, "Real-time Correlation-based Stereo Vision with Reduced Border Errors," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 229-246, 2002.
- [4] P. Moreels and P. Perona, "Evaluation of Features Detectors and Descriptors Based on 3D Objects," *International Journal of Computer Vision*, vol. 73, no. 3, pp. 263-284, 2007.
- [5] C. Ancuti *et al.*, "An Efficient Two Steps Algorithm for Wide Baseline Image Matching," *The Visual Computer: International Journal of Computer Graphics*, vol. 25, no. 5-7, pp. 677-686, 2009.
- [6] E. Vincent and R. Laganière, "Junction Matching and Fundamental Matrix Recovery in Widely Separated Views," in *British Machine Vision Conference*, London, UK, Sept. 7-9, 2004, pp. 77-86.
- [7] S. Kamel, M. Moharram, and R. Elias, "Wide baseline stereo matching through junction parametric and polar warping," in *International Conference on Computer Graphics and Imaging*, pp. 59-64.
- [8] R. Laganière and R. Elias, "The detection of junction features in images," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montréal, Canada, May 17-21, 2004, pp. 573-576.
- [9] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2nd ed., 2004.