

INTEGRATING PERSPECTIVE DISTORTIONS IN STEREO IMAGE MATCHING

Damien Douxchamps, Benoît Macq

Université catholique de Louvain
douxchamps@ieee.org

ABSTRACT

This paper introduces a correlation-based method for the three-dimensional reconstruction of scenes from a multi-camera imaging system. Our technique is to cast the matching and reconstruction problems into a single 3D process that uses perspective distortions to directly retrieve, in a dense fashion, the 3D planes locally tangent to the scene. Avoiding image assumptions like fronto-parallelism, block shapes, perspective distortion models or camera models, our method only needs a local planarity hypothesis of the scene.

1. INTRODUCTION

One of the most difficult problem in 3D image reconstruction from several images is the matching of physical features between two images in order to use triangulation for the recovery of 3D coordinates. Several techniques have been developed to solve this ill-posed problem of correspondence. These are usually based on image features (e.g. points, lines) and/or image regions [1]. While most feature-based approaches yield sparse three-dimensional data, area matching has the significant advantage to provide dense results but also supposes to overcome the specific challenge of perspective distortions. Our method, based on area matching, proposes to solve for these perspective effects not by eliminating them but by using them as primary source of information.

Perspective distortions are related to the local surface orientation of the 3D object relative to the position of the cameras. These position and orientation differences have an effect on the image blocks used for correlation, as shown on fig. 1: a rectangular image block in the first image I_1 will not generally correspond to a rectangular block in the second image I_2 . Not only can the block be rotated or rescaled, but its shape will be transformed into something that depends on the camera (orientation, position and model), and the relative positions of the cameras with respect to the object. Not taking these distortions into account will inevitably lead to matching inaccuracies. Significant perspective effects can even make a block matching process close to impossible, especially in fine structured areas of the image where little changes in block shapes will prevent features to correlate.

Methods that solve the 3D reconstruction problem while taking those perspective distortions into account have started to arise in the 90s. Among others, we would like to stress the work of Devernay [2] [3] who retrieves the local tangent plane with an analysis of disparity differential, followed by an optimization based on the rectified images content. His approach elegantly fine-tunes the results from a classic correlation but does not allow large perspective distortions since the first correlation step does not take the local orientation into account. The present work can be seen as a generalization of this approach where we operate directly on the original image content, avoiding rectification or an initial classic block matching step. While we can expect the generalized approach to be more time consuming, it is also able to deal with strong perspective differences, something quite difficult for methods that use a fronto-parallel scene assumption in a first step.

2. GEOMETRIC OVERVIEW

The underlying idea of our approach is to cast all the geometric aspects in 3D, therefore avoiding any 2D hypothesis in the image plane. Let us associate an image plane I_i to the i^{th} camera of the system. Fig. 1 shows the case of a two camera system ($i = 1, 2$). For every 3D point $P = (x, y, z)$, we can define a projection line (or ray) R_i^P , for each camera, as the line passing through P and the focal point of that camera. This projection ray forms the image point $p_i^P = (m_i^P, n_i^P)$ of P at its intersection with the image plane I_i . It is important to note that since we are working in a projective space, the ray R_i^P is equivalent to its associated projected image point p_i^P [4]. In other words, each ray has only one image point, and vice-versa, so that we can write with little abuse that

$$I_i(R_i^P) \equiv I_i(p_i^P). \quad (1)$$

This relation has the interesting consequence that most of our further reasoning can be done in 3D using rays, instead of working on image coordinates. Suppose now that the two cameras of our stereo setup are calibrated so that both extrinsic and intrinsic parameters are known. Furthermore, we assume that the equation of the plane L that is locally

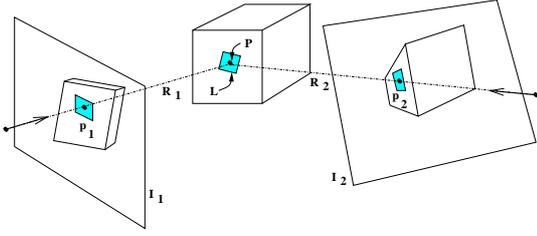


Fig. 1. Geometric aspects

tangent to the object around P is known. The intersection of the ray R_1^P with this plane will give the 3D point P . Once P is known we can exactly determine R_2^P from the *extrinsic* parameters of the camera. One can thus associate a ray $R_{2,L}^P$ for I_2 to every pair of ray R_1^P and plane L .

Our method for recovering L is based on a modified optical flow, block-matching algorithm (BMA). We wish to compare a region around p_1^P with image points in I_2 , but without imposing any region shape or scale in I_2 . Let M be an estimate for plane L . We can project on plane M a neighborhood of n image points of I_1

$$p_{1,j} \equiv R_{1,j}, j = 1 \dots n, \quad (2)$$

as shown on fig. 2, (1). The intersection of this bundle of rays with the plane M yields a set of 3D points $P_{1,j,M}$. Using the second camera extrinsic calibration data we can obtain the rays $R_{2,j,M}^P$ (fig. 2, (2)). Comparing the pixel intensities of these two neighborhoods $R_{1,j}$ and $R_{2,j,M}^P$ using a sum of absolute differences estimator (SAD) provides a correlation information associated to plane M for an initial image point p_1 :

$$SAD(p_1, M) = \sum_{j=0}^n |I_1(R_{1,j}) - I_2(R_{2,j,M}^P)|. \quad (3)$$

Finding the minimum of the SAD for M by varying the plane parameters yields an estimation of the locally tangent plane L (fig. 2, (3)). Note that this problem is significantly more complex than the classical stereo: M has three degrees of freedom while there is only one in classical stereo.

3. ALGORITHM

The locally tangent plane L has a 3D equation that can be written as:

$$L \equiv ax + by + cz + d = 0. \quad (4)$$

This equation is homogenous and has four degrees of freedom that can thus be reduced to three without loss of generality by adding the following constraint:

$$\| (a, b, c) \| = 1. \quad (5)$$

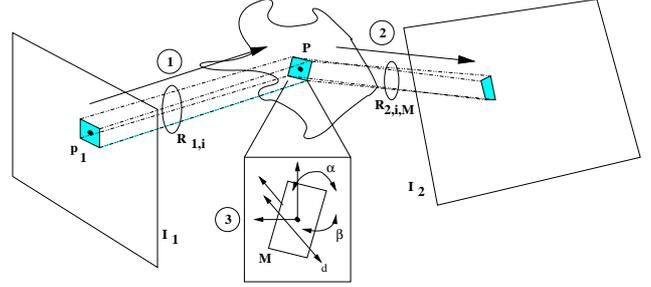


Fig. 2. Projection of the neighborhood and optimization of the tangent plane

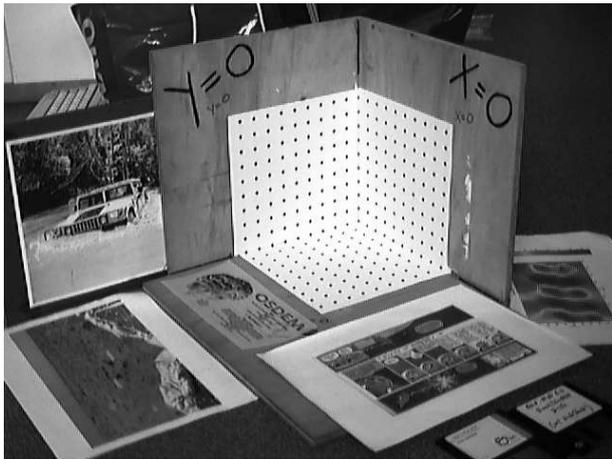
The triplet (a, b, c) being normalized, we can express it in polar coordinates as $(\alpha, \beta, 1)$, yielding the parameters triplet of the tangent plane L : (α, β, d) . This formulation separates the plane parameters in two classes: orientation and offset. As comparison, a classic block matching process for stereo would only optimizing for d , leaving the orientation fixed and equal to that of the image plane.

This is the basis of a two-step optimization: first find P with a classic single degree of freedom optimization along the ray R_1 using d , then find the final plane equation by optimizing on the three parameters (α, β, d) . Because L is at first unknown, finding P with a classic BMA process (like Devernay) will not work in regions of intense perspective distortions. This is precisely where we aim to provide better results so that we must somehow integrate M in the first look for P . This chicken-and-egg problem can be solved with an hierarchical approach [5]: plane orientations obtained at lower resolutions are used for a better search along the ray R_1 . We thus shift the plane estimate along the ray R_1 and look for the best correlation (α and β fixed estimates - d varying). Once this first estimation of P is found we start a full optimization on the three plane parameters.

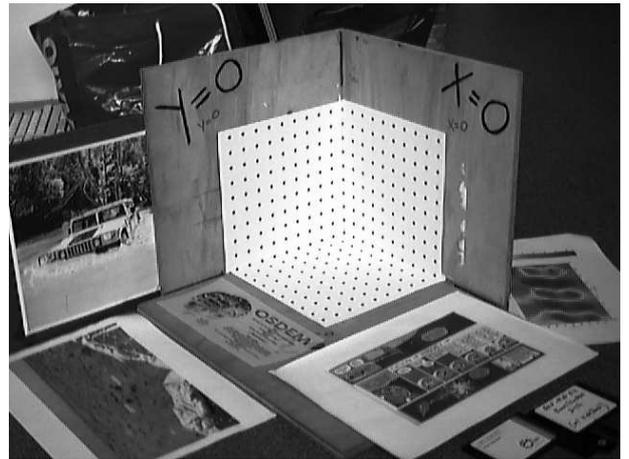
We found that a better stability can be achieved by testing several orientations during this first search. The processing time becoming quite large in that case, we process the image pair at each resolution with a region-growing BMA in order to propagate the first search result to neighbors that will likely have the same local orientation. The criterion used for region growing is that L should remain approximately constant across a region so that we can directly optimize on the three plane parameters with a Nelder-Mead technique [6], [7].

4. RESULTS

The system has been tested on image pairs like the one shown on fig. 3a-b. This image pair contains the calibration target consisting in a trihedral with calibration dots on each of the principal planes $X = 0, Y = 0$ and $Z = 0$. Two dimensional features have been added on (or close to) these



(a)



(b)



(c)



(d)



(e)



(f)

Fig. 3. Results from a test image pair

principal planes. Let us represent the results as four images for the four coefficients a , b , c and d of (4). In the area of the image corresponding to $X = 0$, which should theoretically be as $(a, b, c, d) = (1, 0, 0, 0)$, we expect to see high values (white) for the a coefficient image, while other images should remain relatively dark in that zone (intensity values near zero). $Y = 0$ and $Z = 0$ are described in the same way.

Fig. 3c through 3e show the three components (a, b, c) of the recovered planes, at a reduced resolution. As expected these images show almost only black or white pixels since the scene consists almost only in principal planes. These have been correctly extracted, and their edges are in general well defined. Sometimes, however, the region growing leads to incorrect results, like plane $X = 0$ which extends in a featureless zone.

Regions defined by the region-growing BMA are shown on fig.3f. In general, the algorithm spreads correctly on plane surfaces, gaining much processing time and stability.

5. CONCLUSIONS

We have presented a method for recovering locally tangent planes of a scene from a multi-camera setup. The first implementation produced some encouraging results, although it should still be seen a basic platform for further experiments. It is clear at this point that two major drawbacks must be overcome: the computational power required is too large and the overall quality of the results must be enhanced. In that perspective, we are now working on using image properties that are not subject to perspective distortions in order to guide the matching process.

6. REFERENCES

- [1] "Special issue on 3-d and stereoscopic visual communications," *IEEE Signal Proc. Mag.*, vol. 16, no. 3, May 1999.
- [2] F. Devernay, "Computing differential properties of 3-d shapes from stereoscopic images without 3-d models," Tech. Rep. RR-2304, INRIA, Sophia Antipolis, 1994.
- [3] F. Devernay, *Vision Stéréoscopique et propriétés différentielles des surfaces*, Ph.D. thesis, INRIA, Sophia Antipolis, 1997.
- [4] Olivier Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, Cambridge, Massachusetts, 1993.
- [5] L. Cohen, L. Vinet, P.T. Sander, and A. Gagalowicz, "Hierarchical region based stereo matching," in *Proc. of the CVPR Conference*, 1989, pp. 416–421.
- [6] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Comp. Jour.*, vol. 7, pp. 308–313, 1963.
- [7] Brian D. Bunday, *Basic Optimization Methods*, Edward Arnold, 1984.