Differential motions for recovering 3D structure and motions from an unstructured environment

Jose Vicente, D. Guinea, V. Preciado

Instituto de Automtica Industrial, Spanish Council for Scientific Research
La Poveda (Arganda del Rey) - 28500 Madrid, Spain

ABSTRACT

Present work is focused to real time algorithms for estimating 3D structure and motion from an unstructured environment, we assume an uncalibrated camera. Our approach can be divided into two stages. The first is a self-calibration process, which is performed by means of a rotating camera, subsequently a reconstruction process is accomplished by translation movement. In both stages no correspondences are found in an explicit way, alternatively correspondences and transformations matrices are calculated together. To this end, we combine optical flow constraint with a projective model of a moving camera. Additionally we implement a multi-resolution approach to solve the infinitesimal nature of optical flow constraint. We also focus this work in the bottleneck time requirements existing in artificial vision systems that needs blending complex algorithms and huge amount of data. Results are presented to show the validity of the developed process

Keywords: Image Sequence, Self-calibration, Projective model

1. INTRODUCTION

One of the main goals in computer vision is the recovery of spatial information about environment from the image., It is known that process can be achieved from a sequence of images taken by a moving camera.[1] The kind of motion and additional constraints are essential to determine which level of recovery can be achieved. It become clear that with uncalibrated camera, and no additional constraints only projective reconstruction can be obtained,[2] hence the projection matrices and the reconstructed points can be obtained up to projective transformation. The link between projective reconstruction and Euclidean is to obtain the calibration matrix. Early works have fixed this problem using techniques of calibration based in employ a patron jib, and knowledge of world-points that form this jib. The Maybank and Faugeras paper [3] open a possibility of calibrating a camera by using a sequence of images, and without knowledge of scene in which the camera are situated, in this kind of techniques the calibrating process are fed with the only data from images sequence. Subsequent, works of Hartley[4] and Seo,Hong[5] have based the calibration stage by means of rotating camera and self-calibration background. In their approaches, at the first stage of calibration process matching points between images sequence are found. Later by means of these correspondences transformation matrixes are estimated, these transformations codify homologous points in images pairs. The last stage separates rotation part and calibration matrix from this transformations matrices. To achieve an Euclidean reconstruction it is necessary a calibration of a camera ,while this process can be simplified with pure rotational movements, the reconstruction of scene points isn't possible. On the other hand translation movements are suitable for reconstruction, provided that calibrations matrixes are been estimated.

Our approach can be divided into two stages. The first is a self-calibration process, which is performed by means of a rotating camera, subsequently a reconstruction process is accomplished by a translation movement. In both stages no correspondences are found in an explicit way, alternatively correspondences and transformations matrices are calculated together. To this end, we combine optical flow constraint with a projective model of a moving camera.

Author E-mail:
jvicente@iai.csic.es
domingo@iai.csic.es

2. MATHEMATICAL BACKGROUND

The projections of a scene point onto an image can be estimated using a ping hole camera, modeled by the equation $m = PM$, where $m = [x \quad y \quad 1]^T$ is an image point and $M = [X \quad Y \quad Z \quad 1]^T$ is a scene point, both expressed in homogeneous coordinates. P is a projection matrix of dimension 3x4, that map scene point onto sensor points at a fixed position of a camera. The camera projection matrix can be decomposed as

$$P = K[R| - RT] \tag{1}$$

Where T represents a translation vector and R a rotation matrix, in a rigid euclidean transformation between the camera and word coordinated system. The upper triangular matrix K encodes the intrinsic parameters of camera calibration, it can be interpreted as an afin transformation between sensor plane and camera plane.

$$K = \begin{pmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

Where f_x and f_y represent the focal length in x- and y-sensor axe, u_0 and v_0 is the principal point or projection of optical axe, and s is the skew or angle between x and y sensor axes.

When the camera moves in space, a sequence of images are taken in different camera coordinate system. Those systems are linked one another with rigid transformations. The first camera position can be selected as world reference frame. Therefore, the projection of the same space point onto different frames can be expressed as

$$m_j = P_j M = K_j[R_j| - R_j T_j]M \quad where \quad m_0 = P_0 M = K_0[I|0]M \tag{3}$$

Where the position of camera are represented with subscripts $j = 0, .., j$

The reconstructions problem consists of find the projection matrices P_j and coordinates of space point M , such that $m_j = P_j M$, where the homologous points m_j in the images are known.

Each of such relationships can be regrouped with image in first position, the result are transformations relating correspondence points in first image and j. Noting that the scene point take part in those relations, expressed by the inverse of Z coordinate

$$\begin{pmatrix} x_j \\ y_j \\ z_j \end{pmatrix} = \underbrace{K_j R_j K_0^{-1}}_{Infinite-homography \quad H_{\infty 0j}} \begin{pmatrix} x_0 \\ y_0 \\ 1 \end{pmatrix} - \frac{1}{Z} \underbrace{K_0 R_j T_j}_{Epipole \quad E_{0j}} \tag{4}$$

Where $H_{\infty 0j}$ is a matrix 3x3 that represents the infinite homography i.e. homography that map corresponding point from view 0 to j through the infinite plane, and E_j is a vector 1x3 that represent the epipole, i.e. the projection of the optical center of view 0 onto image j. Therefore, in this case the reconstruction problem consists of find the infinite homography matrices $H_{\infty 0j}$, epipole vector E_{0j} and inverse of Z coordinate, such that $m_j = H_{\infty 0j} m_0 - \frac{1}{Z} E_{0j}$. Where the homologous points m_0 and m_j have been previously determinated.

Between all possible movements, It can be remarked two particulars movements: Pure rotations and pure Translations.

In the case that camera undergoing rotations:

$$m_j = H_{\infty 0j} m_0 \tag{5}$$

The infinite homography can be computed directly from images correspondences. This featured can be used to self-calibration methods.

3. DIFERENTIAL MOTIONS

This paper is focused to small movements of homologous pixels between images. This can be archived with short rotations, around twenty degrees. In the case of translation its depends on relation between Z coordinate of space points and translation magnitude. In the experiments presented in this paper have been handled displacement around 40 pixels in 385 x 288 resolutions images.

We can express the previous relationship for small displacement of pixels images.

$$\begin{pmatrix} x_j \\ y_j \\ z_j \end{pmatrix} = [1 + \Delta H_{\infty 0j}] \begin{pmatrix} x_0 \\ y_0 \\ 1 \end{pmatrix} - \frac{1}{Z} \Delta E_{0j} \tag{6}$$

3.1. Optical flow

Let $I = I(x, y, t)$ denote the time-varying image intensity function and let (u, v) denote the $x-$ $y-$components of the instantaneous optical flow. The computation of the optical field using the classical image motion constraint equation $I_x u + I_y v + I_t = 0$ is difficult to owing to the aperture problem an unstable, instead we adopted a spatial support adopting the approach of Lucas y Kanade,[6]

The gradient-based local optimization method obtains the necessary additional constraints from a finite support and combines them by weight, linear least squared. Specifically finds the pair that minimizes

$$E(u, v) = \sum_{i,j} w_{i,j} [I(x + i + u, y + j + v, t + 1) - I(x + i, y + j, t)]^2 \tag{7}$$

Where the indices i, j take a range over small support centered on (x, y). Substituting a first order approximation and expanding the squared yield

$$E(u, v) = \sum_{i,j} w_{i,j} [I_x^2 u^2 + 2uvI_x I_y + I_y^2 v^2 + 2uI_u I_t + 2vI_v I_t] \tag{8}$$

where I_x and I_y are the components of image gradients at points inside the support, located in the reference frame, and I_t is the discrete temporal derivative inside the same region. The weight sum over the support can been substituted by a gaussian convolution. To obtain the minimum the function is Differentiated with respect (u, v) and set to zero. The follow linear system is obtained to solve in unknown u, v.

$$\begin{pmatrix} G \otimes I_x^2 & G \otimes I_x I_y \\ G \otimes I_x I_y & G \otimes I_y^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} G \otimes I_x I_t \\ G \otimes I_y I_t \end{pmatrix} = 0 \tag{9}$$

The interesting point is that can be obtained the follow lines in projective space

$$s_x = \begin{pmatrix} G \otimes I_x^2 \\ G \otimes I_x I_y \\ G \otimes I_x I_t - xG \otimes I_x^2 - yG \otimes I_x I_y \end{pmatrix}^T \tag{10}$$

$$s_y = \begin{pmatrix} G \otimes I_x I_y \\ G \otimes I_y^2 \\ G \otimes I_y I_t - xG \otimes I_x I_y - yG \otimes I_y^2 \end{pmatrix}^T \tag{11}$$

Noting the important fact that if $m_0 = [x_0\ y_0\ 1]^T$ and $m_j = [x_j\ y_j\ z_j]^T$ are homologous points in different view, this lines are coincident with homologous points m_j. That is

$$s_x m_j = 0 \qquad s_y m_j = 0 \tag{12}$$

This lines encoding spatial-temporal gradient in reference image. We will combine this lines in projetive plane, with a projective model of moving camera in order to calibrate and reconstruct the Euclidean space.

3.1.1. Pure Rotation

We can apply previous equations in the case of pure Rotation. The followed linear system is obtained in the unknown $H_{\infty 0j}$ parameter components

$$s_x H_{\infty 0j} m_0 = 0 \tag{13}$$

$$s_y H_{\infty 0j} m_0 = 0 \tag{14}$$

Removing the scale factor inherent in homogeneous coordinates, we parameterize the infinite homography with eight degrees of freedom,

$$H_{\infty 0j} = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, 1] \tag{15}$$

Thus, each image point gives two constraints on a eight degree of freedom, the system is over-determitated, it can by solved by Linear Least Squares. Once the infinite homographys have been estimated, it can be used to camera calibration.

3.1.2. Pure Translation

We can apply previous equations in the case of pure translations, For each pair of images the followed equations can be obtained, Note that this expressions later on will be used for reconstruction.

$$s_x H_{\infty 0j} m_0 - \frac{1}{Z} s_x E_{0j} = 0 \tag{16}$$

$$s_y H_{\infty 0j} m_0 - \frac{1}{Z} s_y E_{0j} = 0 \tag{17}$$

4. CAMERA CALIBRATION

We employ rotational movements to calibrate the camera, in this case the projection matrix P_j depends on calibration matrix K_j and rotation matrix R_j encoding the orientation of the camera

$$m_j = P_j M = K_j [R_j | 0] M \tag{18}$$

Select one image as reference, that is which camera position is coincident with world reference frame, it is possible completely eliminate the scene point.

$$m_j = K_j R_j K_0^{-1} m_0 = H_{\infty 0j} m_0 \tag{19}$$

Its important remark that with rotation movement, the infinite homagraphy between images is directly observable through image matching point. In this paper this homographys $H_{\infty 0j}$ have been estimated by means of equations (14) and (15), hence this estimation is direct, without need to search matching point in an explicit way.

Noting that this homographys verifies the important relationships

$$H_{\infty 0j} K_0 K_0^T H_{\infty 0j}^T = K_j K_j^T \tag{20}$$

$$H_{\infty 0j}^{-T} K_0^{-T} K_0^{-1} H_{\infty 0j}^{-1} = K_j^{-T} K_j^{-1} \tag{21}$$

Both of them (21)(22) are suitable for camera calibration, And more important think, even if intrinsic parameter varying between images, provided that suitable assumptions can be done over K_j, that is K_j has zero skew, .

It become clear that if $K_0 K_0^T$ or $K_0^{-T} K_0^{-1}$ are known, the matrix K_0 is also known by Choleski factorization

In that equations (21)(22), the matrix $H_{\infty 0j}$ are known, and left to determine any of the symmetrical matrices $K_0 K_0^T$ or $K_0^{-T} K_0^{-1}$

Note that equations (21)(22) can be interpretated in the follow sense: Matrix $K_j^{-T} K_j^{-1}$ is not computed directly, instead it imposes restrictionts over $K_0^{-T} K_0^{-1}$

In order to clarifying, the expression for symmetrical $K_j^{-T} K_j^{-1}$ matrix is a six unknowns parameters matrix

$$K_j^{-T}K_j^{-1} = \begin{pmatrix} \frac{1}{f_x^2} & -\frac{s}{f_x^2 f_y} & \frac{sv_0}{f_x^2 f_y} - \frac{u_0}{f_x^2} \\ \cdot & \frac{s^2}{f_x^2 f_y^2} + \frac{1}{f_y^2} & -\frac{s^2 v_0}{f_x^2 f_y^2} + \frac{su_0}{f_x^2 f_y} - \frac{v_0}{f_y^2} \\ \cdot & \cdot & 1 + (\frac{sv_0}{f_x f_y} - \frac{u_0}{f_x})^2 + \frac{v_0^2}{f_y} \end{pmatrix} = \begin{pmatrix} k_1 & k_2 & k_3 \\ k_2 & k_4 & k_5 \\ k_3 & k_5 & k_6 \end{pmatrix} \tag{22}$$

If the assumption is zero skew causes that parameter $k_2 = 0$, this impose a linear constraint over the remaining five unknown parameters of $K_0^{-T}K_0^{-1}$ When also $f_x = f_y$ is added cause that $k_1 = k_4$, and this impose two additional constraint over the remaining four unknown parameters In this paper the restrictions imposed are zero skew and $f_x = f_y$

5. EUCLIDEAN RECONSTRUCTION

Once the camera is calibrated, we employ translational movements to Euclidean space reconstruction. To achieve this objective a sequence of three views is used. As previously, in order to denote the position of the camera, the indices $0, 1, 3$ are used.

Three views are known to produce four trilinear forms, whose coefficients are arranged in a tensor representing a bilinear funcitions of cameras matrices. The tensor acts on a triplet of point-line-line matching in the three view. That is, if points m_0, m_1, and m_2 are matching point in the three view, cause that m_0, s^1, s^2 contract the tensor to zero. Where s^1 and s^2 are projective lines coincident with points m_1 and m_2 respectively.

Denoting the tensor by T_i^{jk} this can be expressed in tensorial notation as:

$$s_j^1 s_k^2 m_0^i T_i^{jk} = 0 \tag{23}$$

In the case of pure translation, and considerating that intrinsic parameter are constant or known in the three views, the following equations can be deduced

$$s_k^1 I m_0^i - \frac{1}{Z} s_k^1 E_{01} = 0 \tag{24}$$

$$s_j^2 I m_0^i - \frac{1}{Z} s_j^2 E_{02} = 0 \tag{25}$$

Eliminating the inverse of depth, the tensor can be easily reduced to the follow simple form

$$s_k^1 m_0^i s_j^2 E_{02} - s_j^2 m_0^i s_k^1 E_{01} = 0 \tag{26}$$

Thus the tensor only depend on epipoles E_{02} and E_{02}, we can linearity solve the unknows E_{02} and E_{02} applying the next projective lines

$$s_x^1 = \begin{pmatrix} G \otimes I_x^2 \\ G \otimes I_x I_y \\ G \otimes I_x I_t(1) - xG \otimes I_x^2 - yG \otimes I_x I_y \end{pmatrix}^T \quad s_y^1 = \begin{pmatrix} G \otimes I_x^2 \\ G \otimes I_x I_y \\ G \otimes I_x I_t(1) - xG \otimes I_x^2 - yG \otimes I_x I_y \end{pmatrix}^T \tag{27}$$

$$s_x^2 = \begin{pmatrix} G \otimes I_x I_y \\ G \otimes I_y^2 \\ G \otimes I_y I_t(2) - xG \otimes I_x I_y - yG \otimes I_y^2 \end{pmatrix}^T \quad s_y^2 = \begin{pmatrix} G \otimes I_x I_y \\ G \otimes I_y^2 \\ G \otimes I_y I_t(2) - xG \otimes I_x I_y - yG \otimes I_y^2 \end{pmatrix}^T \tag{28}$$

Thus, each image location gives four constraints on the six degree of freedom, the system is over-determined an is solved using Linear Least Squares. It's important remark that a control over equations (10) is maked, In image points that has full rank causes that four constraints are introduced, when the rank is one bring about two lines are used, in the special case that the system is ill conditioned, no constraint are applied. This control is maked by eigenvalues magnitude.

6. EXPERIMENTS

In the experiment a multi-scale of four levels have been created. With the displacement that have been managed, around 30 pixels, this have been enough to validate the differential equations introduced by the defined proyective lines. Fig 1 shows the image sequence that have been used in order to reconstruct the scene. The movement of camera between the three images have been translations in the same directions, it is important remark that in this ill conditioned case, the algorithm have taken proper performance. Fig 2 shows the recovered depth map at different levels of resolutions.

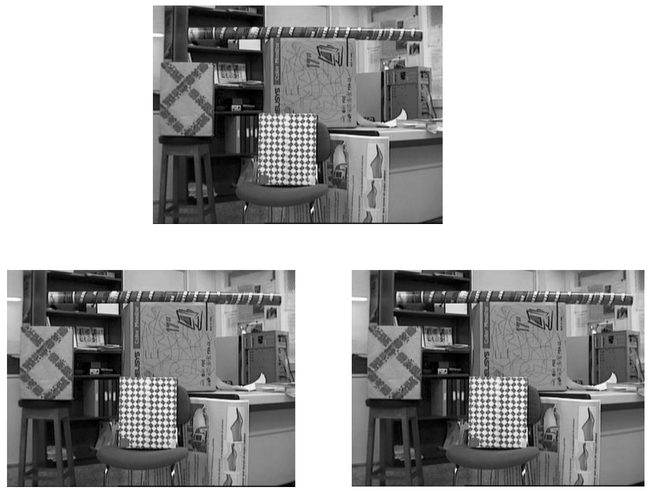


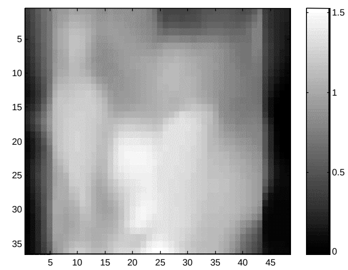
Figure 1. Image sequence of three view

ACKNOWLEDGMENTS

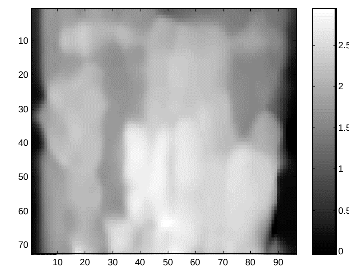
This work has been founded by the Spanish CICYT project SIVA (Integrated System for Active Vision). We thanks also our partners of Institute of Microelectronics of Sevilla-CSIC and the Institute of Optics Daza de Valdes-CSIC

REFERENCES

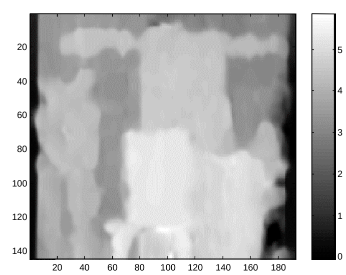
1. R. Hartley, "Euclidean reconstructions from uncalibrated views," *AICV* , pp. 237–256, 1994.
2. R. Hartley, "Estimation of relative camera positions for uncalibrated cameras," *Proc. ECCV* , pp. 579–587, 1992.
3. S. Maybank and O. Faugeras, "A theory of self-calibration of moving camera," *International Journal of computer Vision* , pp. 123–151, 1992.
4. R. Hartley, "Self-calibration of stationary cameras, international journal of computer vision," *International Journal of computer Vision* **22**, pp. 5–23, 1997.
5. Y. Seo and K. S. Hong, "Auto-calibration of rotating and zooming camera," *IAPR Workshop on Machine Vision Applications* , 1998.
6. B. Lucas and T. Kanade, "An iterative image registrations technique with an application to stereo visions," *DARPA Image Understanding Workshop* , pp. 121–130, 1981.



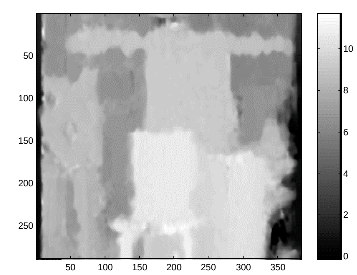
Level 4



Level 3



Level 2



Level 1

Figure 2. Depth map at four level of resolution