Fundamental matrix and slightly overlapping views

Roman Pflugfelder
Video and Safety Technology
ARC Seibersdorf research
A-2444 Seibersdorf, Austria
roman.pflugfelder@arcs.ac.at

Horst Bischof
Institute for Computer Graphics and Vision
Graz University of Technology
A-8010 Graz, Austria
bischof@icg.tu-graz.ac.at

Abstract

This paper proposes a method to compute the fundamental matrix from slightly overlapping views. Slight overlaps are preferable to substantial overlaps in large multi-camera surveillance applications, because otherwise the number of cameras and thus the overall application costs would gravely increase. Instead of using point correspondences alone, we additionally use the infinite homography. The infinite homography can be computed from line segments when we have a Manhattan world scene and assume cameras with square pixels and the same alignment of the scene in both views. All assumptions are not too restrictive and occur frequently. The infinite homography and one of the epipoles determine the fundamental matrix. The paper demonstrates a simple solution to compute the epipoles from at least two point correspondences without requiring points uniformly distributed within the images. Experiments on synthetic images clearly show the advantages of the proposed method compared to the classical eight point algorithm. Images of an office scene show the idea on real world images.

1. Introduction

Robust and automatic methods for the computation of the fundamental matrix are well known [4]. These algorithms use putative point correspondences to compute the fundamental matrix. Such point correspondences are typically detected by matching image features - even in wide-baseline camera settings [1][9]. These methods give accurate results. However, the accuracy depends on the amount of point correspondences in uniform position. Unfortunately, cameras with only slightly overlapping views provide point correspondences in a small part of their images.

The image in Figure 1 does not provide even that, correspondences can only be established by the person in the room.

Figure 1. Two views of an office scene.

Slight overlaps are typical in large surveillance applications, because application costs limit the number of cameras. The background of this method is its application in visual surveillance, e.g. to handover objects between two views. Handovering is simplified when the epipolar geometry is known. Several methods to compute a homography induced by the ground-plane exist. All these methods assume substantially, overlapping views. Only the method of Khan and Shah [5] works with slightly overlapping views. In contrast to these works, the proposed method does not assume that any point correspondences have to lie on the ground-plane. Instead assuming a Manhattan world is not too restrictive in many corridor and room scenes. This work confines itself to man-made environments with orthogonal and parallel structures. These Manhattan worlds contain line segments which determine vanishing points of orthogonal directions in each view [2]. Corresponding vanishing points and the square pixel assumption define the infinite homography [4]. The square pixel assumption is very general and holds for most digital cameras [8]. It is shown in [4] that a homography and at least two point correspondences determine the fundamental matrix. Despite the impossibility of uniformly distributed point correspondences, this paper shows that an accurate computation of the fundamental matrix in slightly overlapping views is possible. The
infinite homography encapsulates the rotation between the two cameras and their intrinsic parameters. Line segments compensate for the lack of point correspondences in uniform position and allow the recovery of rotation and intrinsic parameters. At least two point correspondences within the overlapping image area and the infinite homography determine the epipoles. Experiments with synthetic images show that the proposed method outperforms the eight point algorithm if only a few point correspondences are known. The proposed method gives quite a reasonable fundamental matrix - even with moderate inaccuracies of the point correspondences and/or moderate lens distortion. It is simple and linear and can efficiently be implemented. The paper is organized as follows: Section 2 exploits the computation of vanishing points from line segments and the computation of the intrinsic parameters. Section 3 treats the proposed method to compute the fundamental matrix. After showing and discussing the experiments with synthetic images and the images of an office scene (section 4), the paper concludes with section 5.

2. Vanishing points and calibration

The vanishing points of the three, orthogonal directions of a Manhattan world scene can be estimated with sufficient accuracy from line segments using a non-linear, iterative maximum-likelihood method as it is shown in [7]. The concept of the image of the absolute conic and the square pixel assumption allow a calibration of the intrinsic parameters of both cameras [8]. In the remainder of the paper the intrinsic parameters are named $K_1$ and $K_2$. The vanishing points and the intrinsic parameters are simultaneously computed by a combined RANSAC/EM approach [10] which is inspired by [6]. Instead of exploiting a histogram of the line segment’s orientation, RANSAC acts as a search engine for vanishing points and groups line segments accordingly. It uses a Gaussian noise model for the end-points of the line segments. EM optimizes the grouping and the accuracy of the vanishing points by using the above mentioned maximum-likelihood method in the M-step. The correspondences between vanishing points of the same world directions in both views are manually defined. Hartley [4] and Rother [11] then show the computation of the camera’s rotation - in the remainder of the paper $R_1$ and $R_2$ - from vanishing points and $K_1$, $K_2$.

3. Proposed method

The fundamental matrix $F$ completely encapsulates the epipolar geometry between the images of two views. See the illustration in Figure 2. It is well known that $F$ can be computed from eight point correspondences with Hartley’s eight point algorithm. However, another less known idea exists which also works in slightly overlapping views.

Theorem 1. From [3]: One of the epipoles $e_{21}$ in image $I_1$ or $e_{12}$ in image $I_2$ and every plane induced homography such as the infinite homography $H_{\infty21}$ or its dual $H_{\infty12}$ determine $F$. More formally,

$$F \simeq [e_{21}]_x H_{\infty21} \simeq H_{\infty21}^{-T} [e_{12}]_x \simeq H_{\infty12}^{-T} [e_{12}]_x .$$

$[\cdot]_x$ is the skew-symmetric matrix operator defined in [3]. The $\simeq$ comparator denotes similarity up to a scalar factor due to the use of homogeneous coordinates. We will now show, that $H_{\infty21}$ can be computed from $R_1$, $R_2$, $K_1$ and $K_2$. $e_{21}$ is then computed with $H_{\infty21}$ and at least two point correspondences that do not necessarily have to lie in uniform position. Analogously, the same is valid for $H_{\infty12}$ and $e_{12}$.

3.1. Infinite homography

The infinite homography $H_{\infty21}$ is the homography from image $I_2$ to image $I_1$ induced by the plane at infinity $\Pi_\infty$. $H_{\infty21}$ maps a vanishing point $v_2$ in $I_2$ to a vanishing point $v_1$ in $I_1$ with $v_1 \simeq H_{\infty21} v_2$. Basically, $H_{\infty21}$ can be decomposed in a homography $H_{\infty21}^{-1}$ from $I_2$ to $\Pi_\infty$, a rotation $R$ and a second homography $H_{\infty1}$ from $\Pi_\infty$ to $I_1$. $H_{\infty1}$ and $H_{\infty2}$ map a direction in the scene to a vanishing point in $I_1$ with $v_1 \simeq H_{\infty1} d_1$ and $I_2$ with $v_2 \simeq H_{\infty2} d_2$ respectively. $d_1$ and $d_2$ are the same direction in the scene but are different coordinate vectors, because the cameras are differently aligned in the scene. If we assume an identical alignment of the scene in both views and as the alignment of the cameras is already known by $R_1$ and $R_2$, the rotation between $d_1$ and $d_2$ will be $d_1 = R d_2$ with

$$R = R_1 R_2^{-1} .$$

This assumption is valid in many scenes of a Manhattan world and can be used in many practical applications. Corollary 2 follows immediately from this assumption.
Corollary 2. If the identical alignment assumption is valid and $R_1$, $R_2$, $K_1$ and $K_2$ are known, then
\[ H_{\infty 21} \simeq H_{\infty 1} R H_{\infty 2}^{-1} \simeq K_1 R_1 R_2^{-1} K_2^{-1}. \] (3)

3.2. Epipoles

The infinite homography holds the rotational and the intrinsic information of the epipolar geometry. The only missing information is the translation between the views which is given by the epipoles. The epipole $e_{21}$ in $I_1$ is the non-trivial left null-space solution of $F^T e_{21} = 0$. Equivalently, $e_{12}$ in $I_2$ is the non-trivial right null-space solution of $F e_{12} = 0$. Consequently, $e_{21}$ must lie on every epipolar line $F x_2$ for all points $x_2$ in $I_2$. Analogously, the same applies for $e_{12}$. Conversely, at least two known epipolar lines will define the epipole. In [4] it was shown that a homography and a pair of corresponding points define the corresponding epipolar line. The reason for that is called plane induced parallax as illustrated schematically in Figure 2. Consider without loss of generality the case where this homography is $H_{\infty 21}$. The ray through one of the corresponding points maps to its corresponding epipolar line. Because all points on the ray lie on the epipolar line, the ray’s direction which is the intersection of the ray with $\Pi_\infty$ must also map onto the epipolar line.

Definition 1. Let \( \{(x^i_1, x^i_2) \mid 1 \leq i \leq N, N \geq 2\} \) be tuples of corresponding points. Let all $x^i_2$ map to $x^i_{21}$ in $I_1$ with $x^i_{21} \simeq H_{\infty 21} x^i_2$. A measurement matrix $M$ can be constructed with
\[ M = \left( x^1_1 \times x^1_{21} \times \cdots \times x^N_1 \times x^N_{21} \right)^T. \] (4)

The result is that $M$ can be used to compute the epipoles.

Lemma 3. The epipole $e_{21}$ is the non-trivial solution of the right null-space of $Me_{21} = 0$.

The problem of estimating the intersection of $N$ lines is a least-squares problem and can be solved using Singular Value Decomposition. The problem is shown with proofs in more detail in [3] and [4]. Analogously, the same applies for $e_{12}$. Instead of $H_{\infty 21}$ we use $H_{\infty 12} = H_{\infty 1}^{-1}$. Alternatively, $e_{12}$ is also given by equation (1) with $[e_{12}]_x = H_{\infty 21}^T F$.

4. Experiments

In order to have ground truth we tested a manually designed scene and computed the ground truth of the epipolar geometry from the design parameters (Figure 3a). The images were moderately distorted with a radial, first-order lens distortion model [3] with the radial center in the image center and a distortion parameter of 0.1. The points were further disturbed with Gaussian noise (zero mean and a variance $\sigma_{\text{noise}}$ of 0.01pixel). To compare fundamental matrices with each other we used the error measure of Zhang with the same evaluation as described in [3]. Figure 3b shows the epipolar geometry computed from eight point correspondences (circles) with the proposed method. The error of the fundamental matrix is 6.88pixel. The result of the eight point algorithm with the same eight point correspondences is shown in Figure 3c. Here, the error is 128.25pixel. Compared with the ground truth the result of the proposed method is reasonable with respect to [3, p. 339-340] and the poor result of the eight point algorithm can obviously be seen. Figure 4 shows the error for 8, 10, 15, 20, 25 and 30 point correspondences. The point correspondences are randomly drawn within a certain volume (4a and 4b). The images were not distorted. The Gaussian noise was set to zero mean with a $\sigma_{\text{noise}}^2$ of 0.01pixel (4c) and 0.25pixel (4d). Table 1 and Table 2 show the errors of the fundamental matrix. The proposed method outperforms the eight point algorithm in our experiment if 20
point correspondences or less are available - a frequent case in many surveillance scenes. In contrast to the eight point algorithm, the errors for the proposed method are comparable and of the same order of magnitude - even under different noise levels. These errors mostly reflect the error in the prior estimation of the infinite homography. These results confirm that the eight point algorithm is improper to estimate an accurate fundamental matrix with a few, clustered point correspondences. However, the proposed method is limited by an accurate estimation of the infinite homography. Hence, the error of the eight point algorithm improves with the number of point correspondences while the error in the proposed method remains roughly the same. The proposed method has advantages in situations like the real office scene where only a few points on a moving person are available. These points and their corresponding epipolar lines are shown in Figure 1. A comparable fundamental matrix was computed with the self-calibration method discussed in [12] which uses hundreds of point correspondences produced by a moving light source. The difference of our fundamental matrix compared to the result of the self-calibration method was only 6.42 pixel.

5. Conclusion and future work

This paper presented a simple method to compute the fundamental matrix in slightly overlapping views. The idea is to compute the infinite homography from line segments and the epipoles from point correspondences not necessarily in uniform position. The necessary assumptions are not too restrictive and are valid in many real scenes. Experiments with synthetic images and a real office scene support the idea and show reasonable results. The proposed method outperforms the eight point algorithm with a few, clustered point correspondences. The experiments further confirm the knowledge that the eight point algorithm should not be applied in such situations. Unfortunately, the method discussed in section 3.2 is not accurate enough in case of many point correspondences. Our future work will concentrate on a more accurate estimation. For the applicability in wide-area surveillance we further concentrate on detecting point correspondences automatically on moving objects, e.g. humans, vehicles.

References


