Image and Vision Computing 67 (2017) 16–28

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Robust epipolar geometry estimation using noisy pose priors☆

Yehonatan Goldman[a], Ehud Rivlin[a], Ilan Shimshoni[b,*]

[a] Technion - Israel Institute of Technology, Technion City, Haifa 3200003, Israel
[b] University of Haifa, 199 Aba Khoushy Ave., Mount Carmel, Haifa, Israel

ARTICLE INFO

Article history:
Received 16 February 2016
Received in revised form 19 June 2017
Accepted 9 September 2017
Available online 20 September 2017

Keywords:
Robust estimation
Epipolar geometry
Pose priors

ABSTRACT

Epipolar geometry estimation is fundamental to many computer vision algorithms. It has therefore attracted a lot of interest in recent years, yielding high quality estimation algorithms for wide baseline image pairs. Currently many types of cameras such as smartphones produce geo-tagged images containing pose and internal calibration data. These include a GPS receiver, which estimates the position, a compass, accelerometers, and gyros, which estimate the orientation, and the focal length. Exploiting this information as part of an epipolar geometry estimation algorithm may be useful but not trivial, since the pose measurement may be quite noisy. We introduce SOREPP (Soft Optimization method for Robust Estimation based on Pose Priors), a novel estimation algorithm designed to exploit pose priors naturally. It sparsely samples the pose space around the measured pose and for a few promising candidates applies a robust optimization procedure. It uses all the putative correspondences simultaneously, even though many of them are outliers, yielding a very efficient algorithm whose runtime is independent of the inlier fraction. SOREPP was extensively tested on synthetic data and on hundreds of real image pairs taken by smartphones. Its ability to handle challenging scenarios with extremely low inlier fractions of less than 10% was demonstrated. It outperforms current state-of-the-art algorithms that do not use pose priors as well as others that do.

1. Introduction

Epipolar geometry is the intrinsic projective geometry between two views, and it is encoded in the *fundamental matrix F* [1]. Its estimation is one of the core problems in computer vision and is used as a basic component for stereo matching [2], structure from motion (SfM) [3], vision-based robot navigation [4], and other applications. The epipolar geometry of two images is usually estimated by finding corresponding features in both. This is done first by detecting and matching invariant features using an algorithm such as SIFT [5], followed by the application of a robust estimation method from the RANSAC [6] family.

The main weakness of RANSAC is the necessity to sample a valid set. As the inlier fraction decreases, the probability to sample a valid set drops rapidly, increasing greatly the required number of iterations.

In recent years considerable progress has been made in developing estimation algorithms that tackle these problems. Such algorithms include LO-RANSAC (Local-Optimization RANSAC) [7], PROSAC (Progressive Sample Consensus) [8], MLESAC (Maximum Likelihood Estimation RANSAC) [9], BEEM (Balanced Exploration Exploitation Model Search) [10], BLOGS (Balanced LOcal and Global Search) [11], and recently USAC (Universal RANSAC) [12]. However, scenarios with wide baseline images or small overlapping regions between the images still challenge even the current state-of-the-art algorithms due to the low inlier fractions. Fig. 1 shows several such challenging image pairs.

The *fundamental matrix F* is constructed from the camera parameters only, independent of the scene. The intrinsic parameters are mainly the focal lengths of the cameras and their principal points, and the extrinsic parameters which describe their relative pose (camera position and orientation). Focal length data already appears in the metadata of standard cameras under the EXIF format. Thus, the internal calibration matrix K can be estimated using the focal length and by estimating the principal point as the center of the image. Image distortions even though exist are ignored. Under these assumptions algorithms that use this data exist [3,13,14].

In this case, the fundamental matrix is simplified to the *essential matrix E*. Nowadays the pose data is also becoming available as smartphones with built-in sensors such as a compass, accelerometers,

---

☆ This paper has been recommended for acceptance by Enrique Dunn.
* Corresponding author.
E-mail address: ishimshoni@is.haifa.ac.il (I. Shimshoni).

https://doi.org/10.1016/j.imavis.2017.09.006
0262-8856/

Fig. 1. Wide baseline images for which the epipolar geometry could not be estimated using current algorithms but was successfully estimated by SOREPP. Left: *Open* set. Right: *Urban* set. Top: The reference image with several control points marked. Bottom: The target image after matching with the control points and their corresponding estimated epipolar lines.

gyros and a GPS receiver become popular. Using these sensors, whose values can be readily measured, the camera pose can be estimated. There exist applications that place in the EXIF image header the position and orientation of the camera [15,16]. Thus, it is very easy to obtain this information.

In principle, knowing the pose and the internal calibration parameters would make vision-based epipolar geometry estimation methods unnecessary, since the fundamental matrix is computed from these known quantities. But this is not the case in practice since the accuracy of these sensors is low. For example, typical compass azimuth errors can exceed 7°. Therefore epipolar geometry estimation contaminated with such errors can be considered as a search problem in the parametric space of the relative pose. Yet, the noisy pose data is valuable and can be used in order to constrain the search and focus it.

This paper tackles the following problem: how to effectively exploit pose priors for epipolar geometry estimation, while taking into account the fact that the pose priors are noisy or even partly incorrect. The paper focuses on images taken by Smartphones. The images may differ greatly from one another, leading to low inlier fractions.

To handle these difficulties, we propose a novel estimation algorithm called SOREPP (Soft Optimization method for Robust Estimation based on Pose Priors). SOREPP optimizes an M-estimator cost function designed for robustness to putative outlier correspondences, while the pose priors are used to initialize the optimization and regularize the solution. Knowledge of the expected amount of noise in the pose parameters is used to limit the search to a part of the parameter space. Because, in contrast to RANSAC, explicit detection of inliers is not required, a solution can be found even in extreme conditions, typically where the inlier fraction is around 10%. The algorithm is simple yet effective. It exhibits fast runtime and does not depend on the inlier fractions, making it attractive for real-time applications. SOREPP achieves a notable improvement over the other methods both in runtime and accuracy for very challenging image pairs, such as those in Fig. 1. It could be used for example as part of a Structure from Motion (SfM) application in which a large number of images of a scene are taken and in the first part of the algorithmic pipeline are matched. Using SOREPP many more image pairs will be matched correctly and more efficiently. As a result, the recovered structure and motion are much more accurate. The implementation of SOREPP, a Matlab mex wrapper, and all the images used in the experiments are available at [17].

The paper has three main contributions. First, it addresses the problem of epipolar geometry estimation when pose prior data is available. Unlike previous works we take into account all available pose information and especially the expected noise level based on a physical model. In the paper we apply the method to smartphone based sensors but it can be easily generalized to other sensor types. The second contribution is the expansion of the robust optimization framework that is usually used to refine a solution, enabling it to be directly applied to the raw data with low inlier fractions and high levels of pose uncertainty. As far as we know this is the first time that this approach has been taken in low inlier fraction cases, instead of RANSAC. Finally, we introduce an extensive experimental setup for testing the algorithm and comparing it to the state-of-the-art. With a relatively low amount of work on single images, a large number of image pairs are generated. This enables us to test and compare different algorithms on hundreds of image pairs.

The rest of the paper is organized as follows. Section 2 reviews the main methods for epipolar geometry estimation, with or without pose prior data. In Section 3 we give details about how pose is measured by smartphones and review epipolar geometry. Section 4 introduces SOREPP, our proposed robust algorithm. Experimental results on synthetic data are described in Section 5, while results on real image pairs are presented in Section 6. Section 7 concludes the paper.