Pattern Recognition 73 (2018) 158–171

Contents lists available at ScienceDirect

Pattern Recognition

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

Mapping and localization from planar markers

Rafael Muñoz-Salinas*, Manuel J. Marín-Jimenez, Enrique Yeguas-Bolivar, R. Medina-Carnicer

Computing and Numerical Analysis Department, Edificio Einstein. Campus de Rabanales, Córdoba University, Córdoba 14071, Spain

ARTICLE INFO

Article history:
Received 30 January 2017
Revised 13 July 2017
Accepted 6 August 2017
Available online 18 August 2017

Keywords:
Fiducial markers
Marker mapping
SLAM
SfM

ABSTRACT

Squared planar markers are a popular tool for fast, accurate and robust camera localization, but its use is frequently limited to a single marker, or at most, to a small set of them for which their relative pose is known beforehand. Mapping and localization from a large set of planar markers is yet a scarcely treated problem in favour of keypoint-based approaches. However, while keypoint detectors are not robust to rapid motion, large changes in viewpoint, or significant changes in appearance, fiducial markers can be robustly detected under a wider range of conditions. This paper proposes a novel method to simultaneously solve the problems of mapping and localization from a set of squared planar markers. First, a quiver of pairwise relative marker poses is created, from which an initial pose graph is obtained. The pose graph may contain small pairwise pose errors, that when propagated, leads to large errors. Thus, we distribute the rotational and translational error along the basis cycles of the graph so as to obtain a corrected pose graph. Finally, we perform a global pose optimization by minimizing the reprojection errors of the planar markers in all observed frames. The experiments conducted show that our method performs better than Structure from Motion and visual SLAM techniques.

1. Introduction

Camera pose estimation is a common problem in several applications such as robot navigation [1,2] or augmented reality [3–5]. The goal of camera pose estimation is to determine the three-dimensional position of a camera w.r.t. a known reference system.

To solve that problem, a great part of the research focuses on using natural landmarks, being Structure from Motion (SfM) and Simultaneous Localization and Mapping (SLAM), the two main approaches. Both methods rely on keypoints [6–8], which detect distinctive features of the environment. However, keypoint matching has a rather limited invariability to scale, rotation and scale, which in many cases makes them incapable of identifying a scene under different viewpoints. Thus, mapping an environment for tracking purposes under unconstrained movements requires a very exhaustive exploration. Otherwise, localization will fail from locations different from these employed for mapping. Take as example Fig. 1, where two images of the same scene are shown from different viewpoints and the SURF [6] keypoint matcher is applied, showing as coloured lines the detected matches. Only two correct matches are obtained in this scene.

Squared planar markers, however, are designed to be easily detected from a wider range of locations [4,9–12]. Most frequently, squared markers use an external (easily detectable) black border and an inner binary code for identification, error detection and correction. A single marker provides four correspondence points which can be localized with subpixel precision to obtain an accurate camera pose estimation. The scene in Fig. 1 contains a set of planar markers which have been properly detected and identified despite the viewpoint changes. However, camera localization from a planar marker suffers from the ambiguity problem [13], which makes it impossible to reliably distinguish the true camera location in some occasions (see Sections 2.4 and 4.5 for further details).

Despite their advantages, large-scale mapping and localization from planar markers is a problem scarcely studied in the literature in favour of keypoint-based approaches. While it is true that some environments cannot be modified, in many occasions it is possible to place as many markers as desired. In these cases, a large-scale and cost-effective localization system can be done using planar markers exclusively. Additionally, in many indoor environments, such as labs or corridors, there are frequently large untextured regions from which keypoints can not be detected. If the environment must be texturized, then, it would be preferable to do

* Corresponding author..
E-mail addresses: rmsalinas@uco.es (R. Muñoz-Salinas), mjmarin@uco.es (M.J. Marín-Jimenez), eyeguas@uco.es (E. Yeguas-Bolivar), rmedina@uco.es (R. Medina-Carnicer).

http://dx.doi.org/10.1016/j.patcog.2017.08.010
0031-3203/

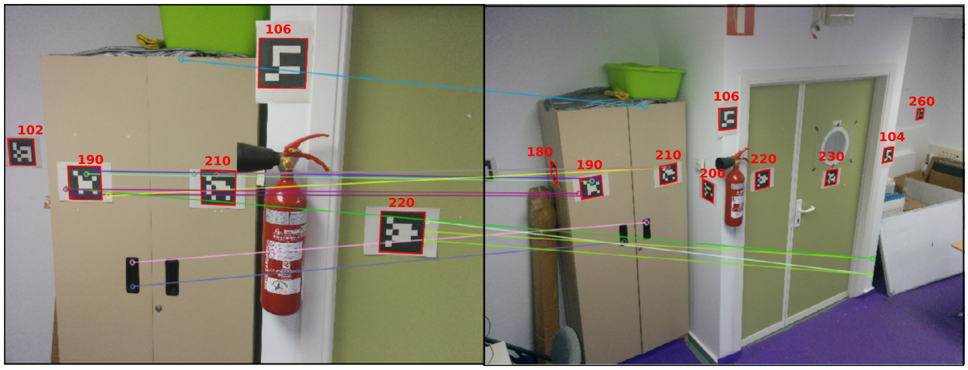

Fig. 1. Example showing the matching capabilities of keypoints versus fiducial markers systems. Coloured lines show the best matches obtained by the SURF keypoint detector. Red rectangles show the markers detected along with its identification. Despite large viewpoint changes, fiducial markers are correctly localized and identified. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

it with fiducial markers, since they can be identified from a wider range of viewpoints than keypoints.

This work proposes a solution to the problem of mapping and localization from planar markers. The contribution of this work is three-fold. First, we propose to tackle the marker mapping problem as a variant of the Sparse Bundle Adjustment problem, but considering that the four corners of a marker must be optimized jointly. As a consequence, our approach reduces the number of variables to be optimized and ensures that the true distance between corners is enforced during optimization. Second, we propose a graph-based method to obtain the initial map of markers dealing with the ambiguity problem. To that end, we first create a quiver of poses from which an initial pose graph is obtained which is then optimized distributing the rotational and translational errors along its cycles. Third, we propose a localization method considering all visible markers, which is able to cope with the ambiguity problem.

In order to validate our proposal, it has been evaluated against two SfM and two SLAM state-of-the-art methods, and the results show that our proposal improves them.

The rest of this paper is structured as follows. Section 2 explains the related works, while Section 3 presents some initial concepts and definitions. Later, Section 4 explains our proposal and Section 5 the experiments conducted. Finally, Section 6 draws some conclusions.

2. Related works

This section provides an overview of the main research related to ours.

2.1. Structure from motion

Structure from Motion techniques take as input a collection of images of the scene to be reconstructed, from which keypoints are detected so as to create a connection graph. From the set of image matches, the relative position of the cameras is obtained by either an incremental or a global approach. Incremental approaches [14,15] select an initial good two-view reconstruction, and images are repeatedly added along with their triangulated matched keypoints. At each iteration, bundle adjustment is applied to adjust both structure and motion. Global approaches [16–18], however, create a pose graph by computing pairwise view poses. In a first step, they compute the global rotation of the views, and in a second step the camera translations. All cycles of the graph impose multi-view constrains that when enforced reduce the risk of drifting occurring in incremental methods. Both incremental and global approaches end with a bundle adjustment process to jointly optimize the motion and structure components.

In order to compute the relative pose between two views it is necessary to assume that the scene is locally planar [19], so that the homography can be computed [20], or compute the essential matrix, which can model both planar and general scenes using the five-point algorithm [21]. However, in most cases, a relatively large number of matches between image pairs is required in order to obtain reliable solutions.

At the core of the SfM techniques there is the need to solve the so called “motion averaging” problem. It refers to a set of methods employed to estimate the poses of a camera observing a common scene from multiple viewpoints. In general, the problem is decomposed into “rotation averaging” and “translation averaging”, since they can be dealt independently. Chatterjee and Govindu [22] proposed a method with the main advantage of not requiring to explicitly compute the cycles in which outlier edges are handled by an iterative non linear refinement. Zach et al. [23] proposed a method to identify outlier edges in a relative rotation graph, which in the case of SfM is a hard problem. Özyesil and Singer [24] proposed a method for robust “translation averaging”. Their main contribution is a method robust to outliers in point correspondences between image pairs, and a convex optimization method to maintain robustness to outlier directions. Govindu also proposed [25] methods to linearly solve for consistent global motion models using a highly redundant set of constraints by using all possible algebraic constraints. Martinec and Pajdla [26] proposed a method to estimate both rotation and translations using a standard technique based on Second Order Cone Programming. Robustness is achieved by using only a subset of points according to a criterion that diminishes the risk of choosing a mismatch. Tron et al. [27,28] proposed distributed algorithms for estimating the average pose of an object viewed by a localized network of cameras. They also show that generalizations of Euclidean consensus algorithms fail to converge to corrects solutions. Sharp et al. [29], proposed methods to solve both the rotation and motion averaging problem. In their approach, outliers are not considered and only the problem of motion averaging is solved, by separately distributing rotation and translation errors along the basis cycles of a graph. This method adapts well to our problem in which no outliers are present since all established correspondences are known to be correct. The main advantage of their approach is that since there is not outliers to be considered, the motion averaging problem is simplified allowing a fast implementation. For a deeper introduction to this problem, the interested reader is referred to Hartley's work [30].