ARTICLE IN PRESS

Computer Vision and Image Understanding xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Proximal robust factorization for piecewise planar reconstruction

Wen Kou, Loong-Fah Cheong*, Zhiying Zhou

4 Engineering Drive 3, National University of Singapore, 117583 Singapore

ARTICLE INFO

Keywords:
Dense planar reconstruction
Robust factorization
Structure from motion
Non-convex optimization

ABSTRACT

In this paper, we aim to obtain a dense piecewise planar reconstruction of the scene from multiple image frames based on a factorization framework. Integrating all the relevant constraints in a global objective function, we are able to effectively leverage on the scene smoothness prior afforded by the dense formulation, as well as imposing the necessary algebraic constraints required by the shape matrix. These constraints also help to robustly decompose the measurement matrix into the underlying low-rank subspace and the sparse outlier part. Numerically, we achieve the constrained factorization and decomposition via modifying a recently proposed proximal alternating robust subspace minimization algorithm. The results show that our algorithm is effective in handling real life sequences, and outperforms other algorithms in recovering motions and dense scene estimate.

1. Introduction

3D reconstruction of scenes from motion cues is a longstanding problem in computer vision. Most of the works in this area (Agarwal et al., 2009; Frahm et al., 2010; Pollefeys et al., 2004; Snavely et al., 2006; Wu, 2013) are based on sparse features, and a post-processing step is required to obtain a dense reconstruction. In this paper, we aim to obtain directly a dense piecewise planar reconstruction based on a factorization framework, using the derivatives of dense optical flows stacked over multiple frames as inputs. Our work aims to realize a practical factorization scheme that can robustly handle the many challenges typical in practical scenarios, including the presence of significant amount of gross outliers in the input. We render the decomposition of the input matrix into the low rank part and the sparse outlier matrix more well-posed by virtue of enforcing various appropriate constraints. In particular, we impose constraints on the rank and the corrupted entries' cardinality directly in their original forms, as well as embedding additional constraints associated with the shape matrix.

Clearly, the aforementioned piecewise planar representation is a more compact and efficient representation than dense 3D point cloud. It also provides a more informative representation than sparse 3D points, especially in man-made environment. One can for instance directly link the reconstructed planes to the notion of *occupancy* for navigation purpose (Danescu et al., 2011), or more generally to the notion of *affordances*, a term coined by Gibson to denote properties of things that afford opportunities of interaction. A simple scheme to make these inferences is to regard planes whose normal orientation is vertical and height zero as navigable, and planes which are parallel to the floor as *sittable* (Holz et al., 2011) etc. Being able to make sense of the 3D structure in this manner is evidently important for autonomous robots and augmented reality applications. Despite the evident utility of such a representation, there is a paucity of works actually adopting this approach. This is despite the massive amount of works in related areas, specifically those of optical flow estimation and factorization, both with long history of research in the computer vision community. In the following, we briefly discuss some issues in these two areas which present bottlenecks to obtaining such a representation and thus motivate our research.

Optical flow is needed as input to our system and thus the quality of its estimates has a significant impact on our proposed scheme. The estimation of optical flow is indeed still a very active area of research, in no small measure due to the release of benchmark datasets (KIT; HCI; MPI). There has indeed been parametric model-based optical flow methods (Black and Anandan, 1996; Ju et al., 1996; Nir et al., 2008; Yang and Li, 2015) whose underlying model is a scene with multiple planes and thus in principle could be used for recovering these planes. Yet, while optical flow has historically been understood to be a means through which eventually 3D structure and motion (a.k.a. SFM) are obtained, there are nowadays not many works that utilize the flow to go on and tackle the latter part of the problem. The reasons for this state of affair are at least twofold. One of the reasons is simply the optical flows are not good enough. While the performance might look impressive according to the evaluation metrics used in the benchmarks, it is quite a different matter when being used for a geometrically exact process like SFM. There are also global constraints that have been largely neglected. For instance, Zelnik-Manor and Irani (2002) showed that for the

* Corresponding author.
E-mail address: eleclf@nus.edu.sg (L.-F. Cheong).

https://doi.org/10.1016/j.cviu.2017.10.002
Received 9 March 2016; Received in revised form 23 August 2017; Accepted 5 October 2017
1077-3142/

discrete case of homographies, there are hidden global constraints in the form of rank of some parameter matrix. Similar constraints exist for the continuous case of optical flow (Zelnik-Manor and Irani, 2000). Since most of the existing flow methods only consider a local smoothness prior, the resulting parametric models do not necessarily obey these global constraints, and as a consequence, problems arise during the structure recovery stage. Numerically, despite the works of Black and Anandan (1996), Ju et al. (1996) and Nir et al. (2008), the problem of estimating the parameters of the parametric model from the optical flows (either explicitly or implicitly) is still a significant challenge. Despite advances in optimization methods that permit discontinuity-preserving flow estimation, there are still errors remaining due to various practical reasons such as the need to perform relaxation (e.g. using L_1-norm in place of L_0-norm). These errors, when coupled with not knowing the number nor the boundaries of the planes, mean that there is still significant room for the parametric models to go wrong, and indeed they do go wrong.

For the second related area of factorization, the literature is also immense, though very much dominated by the discrete feature-based formulation (Dai et al., 2013). Continuous flow-based factorization works are few and far between (Irani, 2002). As a consequence, useful scene constraints such as scene smoothness and orthogonality of planes are seldom brought to bear on most feature-based factorization approaches to SFM. Indeed, as far as we can ascertain, there is no concrete practical factorization formulation for those approaches based on parametric models (be it from discrete feature or continuous flow). Zelnik-Manor and Irani (2000); 2002) only gave theoretical formulation, whereas practical implementation is fraught with difficulties. These difficulties are manifold. Firstly, due to errors in the optical flow or the parametric model estimation, the input matrix to the factorization problem contains a significant number of outliers, which must be dealt with using appropriate robust factorization algorithms. Secondly, constraints of various forms should be imposed on the problems to improve the quality of the SFM solutions. These include the following: (1) the rank constraint that comes with the factorization formulation, (2) what we called the structural constraints that preserve the structured patterns of the shape matrix (governed by the underlying physical model), and finally (3) scene constraints such as piecewise smoothness of surfaces or orthogonality between planes. Incorporating all these constraints make the factorization problem much harder to optimise. A straightforward robust implementation of the Alternating Least Squares scheme (e.g. Okatani and Deguchi, 2007) ignores the constraints first, and partly as a result, it does not work well (as shown in Wang et al., 2015); it is prohibitively costly too. Deferring the structural constraints to a post-factorization rectification step might sidestep some of the optimization difficulties but such an approach is sub-optimal as the constraint is not imposed during the minimization. In our experience, such a sub-optimal approach breaks down in the face of inevitable noise present in our problems.

Our work proposes a parametric flow-based factorization formulation that deals with all the aforementioned challenges. Fig. 1 shows the overview of our method. Optical flow is first estimated with modern optical flow technique that can handle large displacement and incorporates various best practices such as multi-scale implementation. Parameters of the affine flow that characterizes the local plane in a superpixel are then estimated. Stacking the affine parameters from all the local planes and from all views into a huge matrix, we present a robust version of factorization algorithm which factors the input matrix into a motion matrix and a shape matrix with inner dimension of six, as well as removing outliers in the form of a sparse outlier matrix. Our robust factorization is based on the proximal alternating robust subspace minimization algorithm known as PARSuMi (Wang et al., 2015) but modified to incorporate the additional constraints mentioned in the preceding paragraph. The advantage of the PARSuMi approach is that it has demonstrated significantly better performance on real practical problems with corruptions compared to other methods such as GRASTA (He et al., 2012), Wiberg L_1 (Okatani and Deguchi, 2007) and BALM (Del Bue et al., 2012). Readers are referred to Wang et al. (2015) for details about why the PARSuMi's subspace approach to bilinear problems generally yields the best results. The PARSuMi approach also does not seek convex relaxation of any form, but rather constrains the rank and the corrupted entries' cardinality directly in their original forms. Such faithful representation of the original problem in PARSuMi accounts for its success in solving real problems. Our modification of PARSuMi allows us to embed the aforementioned structural and scene constraints integrally into the optimization process. Specifically, we impose structural constraints on the factorized shape matrix (e.g. equality of some matrix elements) so that the shape matrix has a physical interpretation. We also enforce scene constraint such as smoothness on the resultant dense planar structure estimates. This latter constraint helps to reduce the uncertainty in decomposing the input affine parameter matrix into the low rank part and the sparse outlier matrix.

The structure of this paper is as follows: in Section 2 we introduce the related work; in Section 3, we present the first-order optical flow associated with a plane and explicate the rank constraints present in the affine parameter matrix; in Section 4, the proposed robust factorization algorithm is developed in details. In Section 5, our proposed algorithm is evaluated based on both synthetic data and real world image sequences. The results show that our algorithm outperforms other factorization methods that do not incorporate the constraints integrally into the optimization process.

2. Related work

Most factorization-based SFM works are based on sparse feature points, starting from the seminal work of Tomasi and Kanade (1992) to the latest work such as Dai et al. (2013), with the exception of few works such as Zelnik-Manor and Irani (2000) and Irani (2002) that are based on dense optical flow. It is beyond the scope of this paper to offer a comprehensive review of the immense factorization literature; we only highlight the more related works here.

Irani (2002) is the first work that deals with the continuous case of optical flow; it derives various rank constraints for a variety of imaging models, scene models, and motion models. It focuses on theoretical understanding but it does not address the actual problems encountered in practical scenarios, unlike our paper. There have been quite a few factorization works that are based on a piecewise planar scene model. For the discrete formulation, Zelnik-Manor and Irani (2002) derived the rank constraints on homographies across multiple views based on multiple planar surfaces. Later, Chen and Suter (2009) further refined the rank four constraint over two views for practical implementation. For the continuous flow formulation, Zelnik-Manor and Irani (2000) and Irani (2002) showed that the parameter matrix for the planar flow is of rank six at most. Our basic model is along similar lines, although we omit the second-order effects as these effects cannot be reliably estimated in practice, and more importantly, as mentioned above, we also incorporate various other constraints rather than just considering low rank constraint in isolation. There are other different formulations

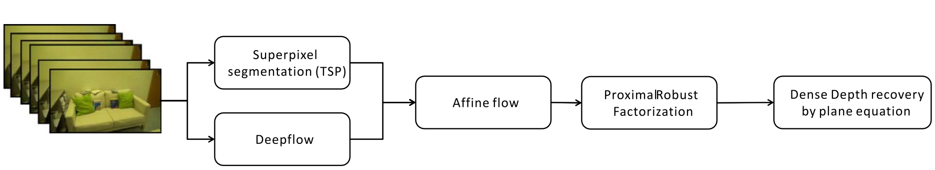

Fig. 1. Algorithm pipeline.