Computer Vision and Image Understanding 135 (2015) 141–156

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Metric corrections of the affine camera ☆

Adrien Bartoli *, Toby Collins, Daniel Pizarro

[a] ALCoV-ISIT, UMR 6284 CNRS/Université d'Auvergne, Clermont-Ferrand, France

ARTICLE INFO

Article history:
Received 19 August 2014
Accepted 2 March 2015
Available online 14 March 2015

Keywords:
Affine camera
Structure-from-motion
Optimal projection

ABSTRACT

Given a general affine camera, we study the problem of finding the closest metric affine camera, where the latter is one of the orthographic, weak-perspective and paraperspective projection models. This problem typically arises in stratified Structure-from-Motion methods such as factorization-based methods. For each type of metric affine camera, we give a closed-form solution and its implementation through an algebraic procedure. Using our algebraic procedure, we can then provide a complete analysis of the problem's generic ambiguity space. This also gives the means to generate the other solutions if any.

1. Introduction

We study the problems of finding the closest orthographic, weak-perspective or paraperspective projection to a general affine camera in the sense of the Frobenius norm. These form three instances of the *metric affine correction* problem class, which we called *orthographic affine correction*, *weak-perspective affine correction* and *paraperspective affine correction*, respectively. The main use of metric affine correction is in Structure-from-Motion by factorization [15,11] and alternation [10]. In the factorization algorithm, metric affine correction is the final stage of a three-stage process. In the first stage, a centered measurement matrix is factored into a joint camera matrix and a structure matrix. This factorization represents an affine 3D reconstruction and is defined up to a (3×3) matrix representing an affine change of coordinates. In the second stage, the metric structure of the affine 3D reconstruction is recovered by computing an affine-to-metric upgrade using the metric constraints from the camera model (for instance, the two rows of the orthographic camera must be orthonormal). The metric constraints are redundant, and can thus only be satisfied approximately. This means that, with noise, the upgraded camera factor is not exactly a metric camera factor. In the third stage, metric affine correction must therefore be performed for each camera in order to recover the metric cameras from the upgraded affine cameras. [15] does factorization with the orthographic camera, while [11] does factorization with the paraperspective camera, but uses a sub-optimal metric affine correction procedure, which could thus be replaced by the proposed one. In the alternation algorithm, metric affine correction is the third stage of an iterative three-stage process. The alternation algorithm requires one to provide an initial estimate of the cameras. In the first stage, the structure is computed from the current camera estimates by triangulation. In the second stage, each camera is computed from the current structure estimate by resection. Both stages amount to solve a set of small linear least squares problems. The second stage estimates affine cameras, as it leaves aside the non-linear constraints characterizing each type of metric camera model. In the third stage, metric affine correction is thus performed for each camera in order to recover the metric cameras. These three stages are repeated until convergence is reached. The third stage is fundamental in the alternation algorithm for two reasons. The first reason is that because of noise, similarly to the third stage of the factorization algorithm, the computed general affine cameras are not exactly metric cameras. The second reason is probably more important: without the third stage, an alternation algorithm would converge to an affine, and not to a metric, reconstruction. The third stage indeed introduces the metric constraints into the alternation algorithm. [10] does alternation with the weak-perspective camera, and could be readily extended to the paraperspective camera with our correction procedure.

Metric affine correction shares strong similarities with orthonormal Procrustes analysis. Inspired by the derivation of the optimal solution to orthonormal Procrustes analysis (specifically, we follow the derivation in [1]) inspired by [8,13], we solve orthographic affine correction and weak-perspective affine correction by a simple algebraic procedure, whose derivation is also fairly simple but does not seem to have appeared in the literature before. We also solve paraperspective affine correction by a simple algebraic procedure. Its derivation is however far more involved. We establish the algebraic procedures and prove their optimality. Our procedures allow us to provide an analysis of the problem's generic

☆ This paper has been recommended for acceptance by David Jacobs.

* Corresponding author.
E-mail address: Adrien.Bartoli@gmail.com (A. Bartoli).

http://dx.doi.org/10.1016/j.cviu.2015.03.001
1077-3142/

Table 1
Summary of our results on solution uniqueness. Matrix $\mathtt{P} \in \mathbb{R}^{2\times 3}$ is a known upgraded affine projection matrix whose correction into one of the three listed metric affine camera models is sought. The case $\operatorname{rank}(\mathtt{P}) = 2$ includes the two sub-cases where the singular values of $\mathtt{P}$ are distinct or equal.

Camera model	$\operatorname{rank}(\mathtt{P}) = 2$	$\operatorname{rank}(\mathtt{P}) = 1$	$\operatorname{rank}(\mathtt{P}) = 0$
Orthographic			
Rotation	Unique	$\mathbb{SO}_2$ ambiguous	Unrecoverable
Weak-perspective			
Scale	Unique	Unique	Unique
Rotation	Unique	$\mathbb{SO}_2$ ambiguous	Unrecoverable
Paraperspective			
Scale	Unique	Unique	Unique
Rotation	Unique	$\mathbb{SO}_2$ ambiguous	Unrecoverable

ambiguities. These are generic in the sense that they apply to any solution algorithm. Our analysis thus determines cases for which the problem has a unique solution, and cases for which it does not. For the latter, we provide a characterization of the solution space[1] and a means to generate all solutions. Our results on the solution ambiguities are summarized in Table 1. Metric affine correction is a set of constrained polynomial optimization problems, to which polynomial optimization methods could be applied. This would however be computationally more expensive by several orders of magnitude than our analytical solutions and would not reveal the problems' intrinsic structure and degenerate cases.

Our input data is an affine projection matrix written as $\mathtt{P} \in \mathbb{R}^{2\times 3}$ (and the direction of projection in the paraperspective case). Our goal is to perform metric affine correction on $\mathtt{P}$. For the orthographic camera, this means finding the camera's rotation, and for the weak-perspective and paraperspective cameras, this means finding the camera's rotation and scale factor. The rank of $\mathtt{P}$ must be two [5]. A rank of one would mean that all 3D space points would be projected to a single image line; a rank of zero would mean that they would be projected to a single image point. Even if these are not proper projections, the rank of matrix $\mathtt{P}$ may drop to one or zero for near degenerate geometries under the effect of noise. For instance, a rank of one may happen when viewing an object with a strong tilt, while a rank of zero may happen when viewing an object at a distance with a narrow field of view. Analyzing degenerate cases thus tells us what may happen in near degenerate cases. We established that, excluding the case where $\mathtt{P}$ vanishes (which is equivalent to it having a zero rank), the weak-perspective and paraperspective scale is always uniquely recoverable. However, for the three metric affine cameras, the rotation is uniquely recoverable only if $\mathtt{P}$ has full rank, otherwise it has an ambiguity in $\mathbb{SO}_2$.

We first give our notation and background in Section 2. We then solve the metric correction problem for the orthographic, weak-perspective and paraperspective cameras in Sections 3–5 respectively. For each camera model, we first give the correction's cost function and pseudo-code. We then derive the correction procedure based on the Singular Value Decomposition (SVD) and analyze the correction's ambiguities. The details of our analysis of the correction's ambiguities for the paraperspective camera are deferred to Appendix A. We finally give experimental results in Section 6 and conclude in Section 7.

2. Notation and background

2.1. Notation

General notation. We use italics for scalar (such as a and α), bold fonts for vector (such as $\mathbf{v}$) and typewriter fonts for matrices (such as $\mathtt{A}$). The entries of a vector or matrix are written as in $\mathtt{A} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$. We use diag to create (block) diagonal matrices. We use double bar fonts for sets (such as $\mathbb{R}$). We use $\mathbb{B}$ to denote a generic binary set with $|\mathbb{B}| = 2$. We have for instance $\{-1, 1\} \equiv \mathbb{B}$. We write vector two-norm as in $\|\mathbf{v}\|_2$ and matrix Frobenius norm as in $\|\mathtt{A}\|_{\mathcal{F}}$. We define $[a, b]_\times \stackrel{\text{def}}{=} \begin{bmatrix} a & b \\ b & a \end{bmatrix}$ and $\odot$ as the Hadamard (element-wise) product.

Orthonormal matrices. We use $\mathbb{O}_d$ for the Lie group of orthonormal matrices[2] and $\mathbb{SO}_d \subset \mathbb{O}_d$ for the Lie group of special orthonormal matrices, with $d \in \{2, 3\}$. For $\mathtt{A} \in \mathbb{O}_d, \det(\mathtt{A}) = \pm 1$; for $\mathtt{A} \in \mathbb{SO}_d, \det(\mathtt{A}) = 1$. We thus have $\mathbb{O}_d \equiv \mathbb{SO}_d \times \mathbb{B}$. Elements of $\mathbb{SO}_2$ may be parameterized as $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ for $\theta \in \mathbb{R}$. Elements of $\mathbb{O}_2$ may be parameterized as $\begin{bmatrix} a\cos\theta & -\sin\theta \\ a\sin\theta & \cos\theta \end{bmatrix}$ for $\theta \in \mathbb{R}$ and $a \in \{-1, 1\}$. This is equivalent to having $\begin{bmatrix} b\cos\mu & -b\sin\mu \\ \sin\mu & \cos\mu \end{bmatrix}$ for some $\mu \in \mathbb{R}$ and $b \in \{-1, 1\}$. For $\mathtt{A} \in \mathbb{O}_2, \det(\mathtt{A}) = \det(-\mathtt{A})$, and the variable a is thus required to specify whether $\mathtt{A} \in \mathbb{SO}_2$ (for $a = 1$) or $\mathtt{A} \in \mathbb{O}_2 \setminus \mathbb{SO}_2$ (for $a = -1$). For $\mathtt{A} \in \mathbb{O}_3$ however, $\det(\mathtt{A}) = -\det(-\mathtt{A})$, and $\mathtt{A} \in \mathbb{SO}_3$ can thus be switched to $\mathbb{O}_3 \setminus \mathbb{SO}_3$ by simply negating its entries. We write $\mathbb{P}_2$ for the space of (2×2) permutation matrices defined as $\mathbb{P}_2 \stackrel{\text{def}}{=} \{\mathtt{I}, \tilde{\mathtt{I}}\}$ with $\tilde{\mathtt{I}} \stackrel{\text{def}}{=} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. We have that $\mathbb{P}_2 \subset \mathbb{O}_2$ and $\mathbb{P}_2 \equiv \mathbb{B}$,

Sub-Stiefel matrices. A sub-Stiefel set $\mathbb{SS}_{r\times c}, 1 \leqslant r \leqslant 3, 1 \leqslant c \leqslant 3$, is formed as the set of $r \times c$ blocks taken from all orthonormal matrices in $\mathbb{O}_3$. Consequently, the Frobenius norm of any element of $\mathbb{SS}_{r\times c}$ is bounded by 1. For instance, $n \in \mathbb{SS}_{1\times 1} \subset \mathbb{R}$ is a scalar such that $|n| \leqslant 1$ and $\mathbf{n} \in \mathbb{SS}_{2\times 1} \subset \mathbb{R}^{2\times 1}$ is a vector such that $\|\mathbf{n}\|_2 \leqslant 1$.

2.2. Metric affine camera models

The affine camera is simply defined as a projection which preserves parallelism. The general affine camera is thus represented by a matrix $\mathtt{A} \in \mathbb{R}^{2\times 3}$ for the rotational part and a vector $\mathbf{t} \in \mathbb{R}^{2\times 1}$ for the translational part. More specifically, a point with world coordinates $\mathbf{Q} \in \mathbb{R}^{3\times 1}$ is projected to image coordinates $\mathbf{q} \in \mathbb{R}^{2\times 1}$ as $\mathbf{q} = \mathtt{A}\mathbf{Q} + \mathbf{t}$. We use *metric affine camera* to mean an affine camera which satisfies some additional constraints called the metric constraints. Metric affine cameras are important: they form the basis of many *Shape-from-X* methods, such as Photometric Stereo [17] and Shape-from-Shading [7], to name a few. The metric affine cameras may be derived from the perspective camera in two ways. First, by increasing the focal length to infinity while back-tracking along the principal ray [5]. Second, by approximating perspective projection to some order [3].

The orthographic camera is the simplest metric affine camera. An affine camera is orthographic if $\mathtt{A} = \bar{\mathtt{R}}$ with $\bar{\mathtt{R}} \in \mathbb{SS}_{2\times 3}$. In other words, $\mathtt{A}$ must be the leading two rows of a 3D rotation matrix. Geometrically, it rotates the object's points and simply projects them orthographically to the camera's retina. The weak-perspective camera is a zeroth order approximation of the perspective camera. An affine camera is weak-perspective if $\mathtt{A} = \alpha\bar{\mathtt{R}}$ with $\alpha \in \mathbb{R}^+$ and $\bar{\mathtt{R}} \in \mathbb{SS}_{2\times 3}$. In other words, $\mathtt{A}$ must be the leading two rows of a 3D rotation matrix up to a positive rescaling.

[1] A space is a set (a simple collection of objects) with some added structure, such as a norm.

[2] A group is a set associated with an operation called the group law. The set and group law must satisfy closure and associativity, and there must be an identity element and an inverse element for each member of the group. For instance, the group law of $\mathbb{O}_d$ and $\mathbb{SO}_d$ is matrix multiplication, the identity element is the identity matrix in $\mathbb{R}^{d\times d}$ and the inverse is element is obtained by matrix transposition.