Computing Similarity Transformations from Only Image Correspondences

Chris Sweeney[1], Laurent Kneip[2], Tobias Höllerer[1], Matthew Turk[1]
[1]University of California Santa Barbara. [2]Australian National University.

Computing a similarity transformation between two coordinate systems is a fundamental problem in multi-view geometry. In computer vision, similarity transformations can be useful for aligning 3D reconstructions or performing loop closure when scale drift occurs. The standard method to compute a similarity transformation is by aligning 3D points [4]; however, in the context of multi-view geometry this alignment may be sub-optimal since the error in alignment depends on the distance between 3D points. New methods have been proposed to compute a similarity transformation from 2D-3D correspondences such that reprojection error is minimized [3, 5]. These methods are much more accurate and are ideal for applications such as structure-from-motion where a minimal reprojection error is sought. These methods, though, are still overly reliant on the quality of 3D points which is known to deteriorate as the depth of the 3D point increases with respect to the cameras that observe it. Ideally, we would prefer a method that is not dependent on potentially noisy 3D points that is still able to minimize reprojection error.

In this paper, we propose a novel solution for computing the similarity transformation between two coordinate systems from only 2D-2D image correspondences. By representing each coordinate system as a generalized camera [1] solving for the similarity transformation is equivalent to solving for the relative pose and scale between the two generalized cameras. We call this new problem the generalized relative pose and scale problem. Reconciling the relative pose between two generalized cameras as well as the unknown scale is equivalent to recovering a 7 degrees-of-freedom (d.o.f.) similarity transformation. This allows for a much broader use of generalized cameras. In particular, similarity transformations can be used for loop closure in SLAM (where scale drift occurs) and for merging multiple Structure-from-Motion (SfM) reconstructions when the scale between the reconstructions is unknown. This problem arises frequently because scale cannot be explicitly recovered from images alone without metric calibration, so developing accurate, efficient, and robust methods to solve this problem is of great importance.

To solve the generalized relative pose and scale problem we utilize the generalized epipolar constraint [1] while additionally incorporating scale:

$$(f_i \times R f_i')^\top t + f_i^\top ([o_i]_\times R - Rs[o_i']_\times) f_i' = 0, \tag{1}$$

where f_i, f_i' are unit-norm pixel rays and o_i, o_i' are origins of the cameras within the generalized camera and R, t, and s are the rotation, translation, and scale of the similarity transformation. By stacking this constraint for all correspondences, we arrive at a simple linear expression for the similarity transformation:

$$M(R) \cdot \tilde{t} = 0, \text{where } \tilde{t} = \begin{pmatrix} t \\ s \\ 1 \end{pmatrix}. \tag{2}$$

The full details of how to compose matrix M from Eq. (1) are given in the paper. The generalized relative pose and scale problem has 7 d.o.f. and thus requires 7 correspondences in the minimal case. Let us consider the quaternion rotation parameterization $q = (x,\ y,\ z,\ \alpha)^\top$ such that the rotation matrix

$$R = 2(vv^\top + \alpha[v]_\times) + (\alpha^2 - 1)I, \tag{3}$$

where $v = (x,\ y,\ z)^\top$ and $[v]_x$ is the skew-symmetric cross product matrix of v. Thus, M is quadratic in the quaternion parameters and the generalized epipolar constraint of Eq. (2) is a 4-parameter Quadratic Eigenvalue Problem (QEP). No methods currently exist to directly solve a 4-parameter QEP

To make Eq. (2) solvable, we consider a simplified problem by assuming that the vertical direction has been aligned between the two generalized

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

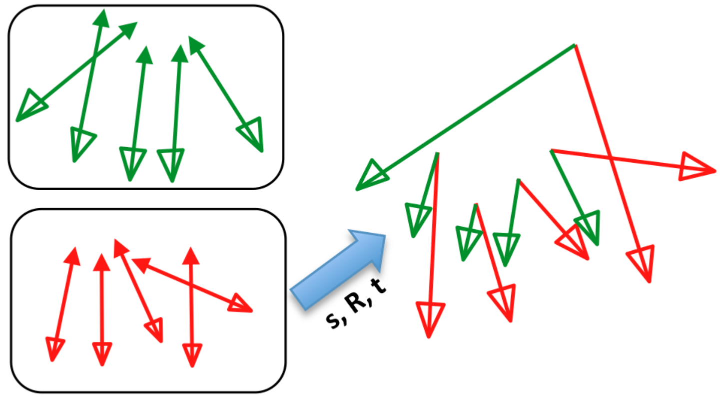

Figure 1: We present a method to solve the generalized relative pose and scale problem. We first align the generalized cameras to a common vertical direction then use image rays obtained from 5 2D-2D correspondences to solve for the remaining degrees of freedom. Solving this problem is equivalent to computing a similarity transformation

cameras. In the quaternion parameterization, this means that $v = (0,\ 1,\ 0)^\top$ and we are left with solving for the unknown parameter α which is related to the rotation angle about the axis v [2]. When considered in the context of the generalized relative pose and scale problem, the intractable 4-parameter QEP from Eq. 2 has now been reduced to a single unknown parameter α in matrix M:

$$(\alpha^2 A + \alpha B + C) \cdot \tilde{t} = 0, \tag{4}$$

where A, B, and C are 5×5 matrices formed from matrix M in Eq. (2). The quadratic eigenvalue problem is simple to solve. The full solution details, including how to robustly align the vertical direction, are provided in the full paper.

In contrast to alternative similarity transformation methods, our approach uses 2D-2D image correspondences thus is not subject to the depth uncertainty that often arises with 3D points. To our knowledge, this is the first method for computing similarity transformations that does not require any 3D information. Our experiments on synthetic and real data demonstrate that this leads to improved performance compared to methods that use 3D-3D or 2D-3D correspondences, especially as the depth of the scene increases.

[1] Robert Pless. Using many cameras as one. In *Proc. IEEE Conference on Conference on Computer Vision and Pattern Recognition*, volume 2, pages II–587. IEEE, 2003.

[2] Chris Sweeney, John Flynn, and Matthew Turk. Solving for relative pose with a partially known rotation is a quadratic eigenvalue problem. In *Proc. of the International Conference on 3D Vision*. IEEE, 2014.

[3] Chris Sweeney, Victor Fragoso, Tobias Hollerer, and Matthew Turk. gdls: A scalable solution to the generalized pose and scale problem. In *European Conference on Computer Vision*, volume 8692, pages 16–31. Springer, 2014.

[4] Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, 1991.

[5] Jonathan Ventura, Clemens Arth, Gerhard Reitmayr, and Dieter Schmalstieg. A minimal solution to the generalized pose-and-scale problem. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014.