Table of Contents

Homographies

In epipolar geometry the coplanar constraint $(\vec{P}_l-\vec{T})^T\vec{T}\times\vec{P}_l=0$ is employed to derive the essential matrix which is part of the epipolar constraint equation $\vec{P}_r^{\,T}E\vec{P}_l=0$ where

$$E=RS=\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}\begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

is the matrix of the extrinsic parameters of stereoscopic systems. In cases when translation is null ($\vec{T}=\vec{0}$) then the essential matrix is undefined. Instead, the equation specializes to

$$H\vec{p}_l=s\vec{p}_r$$

where $\vec{p}_l=(x_l,y_l,1)^T$ and $\vec{p}_r=(x_r,y_r,1)^T$. This is a linear relation between $\vec{p}_r$ and $\vec{p}_l$ which holds up to a scale factor s . The matrix H is a homography and represents a 2D projective transformation.

The first question to consider is how many correspondences are needed to solve for the projective transformation H . This matrix has nine elements but is defined only up to scale, yielding 8 degrees of freedom. Since each correspondence provides two constraints, 4 correspondences are necessary to estimate H . The minimal data set for a solution can be used in a RanSaC framework in order to provide an estimate that is robust to noise.

The Direct Linear Transformation Algorithm

Given a set of $i=1,\ldots,4$ image correspondences, the transformation matrix is given by the equation $s\vec{p}_{ri}=H\vec{p}_{li}$. This equation involves homogeneous vectors and thus $\vec{p}_{ri}$ and $H\vec{p}_{li}$ are not strictly equal. They have the same direction but may differ in magnitude. If we use a geometric constraint such as $\vec{p}_{ri}\times H\vec{p}_{li}=0$

instead, we obtain a simpler solution for H . If the j^{th} row of H is denoted by $\vec{h}_j^T$, then

$$H\vec{p}_{li}=\begin{bmatrix}\vec{h}_1^T\vec{p}_{li}\\ \vec{h}_2^T\vec{p}_{li}\\ \vec{h}_3^T\vec{p}_{li}\end{bmatrix}$$

If we write $\vec{p}_{ri}=(x_{ri},y_{ri},\omega_{ri})^T$ the cross product can then be written as

$$\vec{p}_{ri}\times H\vec{p}_{li}=\begin{bmatrix}y_{ri}\vec{h}_3^T\vec{p}_{li}-\omega_{ri}\vec{h}_2^T\vec{p}_{li}\\ \omega_{ri}\vec{h}_1^T\vec{p}_{li}-x_{ri}\vec{h}_3^T\vec{p}_{li}\\ x_{ri}\vec{h}_2^T\vec{p}_{li}-y_{ri}\vec{h}_1^T\vec{p}_{li}\end{bmatrix}$$

Using $\vec{h}_j^T\vec{p}_{li} = \vec{p}_{li}^T\vec{h}_j$, we can write a set of three equations such as

$$\begin{bmatrix}\vec{0}^T & -\omega_{ri}\vec{p}_{li}^T & y_{ri}\vec{p}_{li}^T\\ \omega_{ri}\vec{p}_{li}^T & \vec{0}^T & -x_{ri}\vec{p}_{li}^T\\ -y_{ri}\vec{p}_{li}^T & x_{ri}\vec{p}_{li}^T & \vec{0}^T\end{bmatrix}\begin{bmatrix}\vec{h}_1\\ \vec{h}_2\\ \vec{h}_3\end{bmatrix} = \vec{0}$$

These equations have the form $A_i\vec{h}=\vec{0}$, where A_i is a 3 by 9 matrix and $\vec{h}$ is a vector containing the 9 elements of H in the following way:

$$H=\begin{bmatrix}\vec{h}_1^T\\ \vec{h}_2^T\\ \vec{h}_3^T\end{bmatrix}$$

The equation $A_i\vec{h}=\vec{0}$ is linear in $\vec{h}$ and quadratic in the known coordinates of the correspondence points.

Only 2 of the 3 equations in $A_i\vec{h}=\vec{0}$ are linearly independent and hence each correspondence gives two equations in the entries of H . It is thus usual to omit the third equation (it is a linear combination of the first two) and write:

$$A_i\vec{h} = \begin{bmatrix}\vec{0}^T & -\omega_{ri}\vec{p}_{li}^T & y_{ri}\vec{p}_{li}^T\\ \omega_{ri}\vec{p}_{li}^T & \vec{0}^T & -x_{ri}\vec{p}_{li}^T\end{bmatrix}\begin{bmatrix}\vec{h}_1\\ \vec{h}_2\\ \vec{h}_3\end{bmatrix} = \vec{0}$$

where A_i is now a 2 by 9 matrix. A number of observations are worth a note:

- These equations hold for any homogeneous coordinate representation $(x_{ri}, y_{ri}, \omega_{ri})^T$ of $\vec{p}_{ri}$. If $\omega_{ri}=1$ then (x_{ri}, y_{ri}) are measured in image coordinates.
- Since H is generally determined up to scale, $\vec{h}$ actually gives the required matrix H . A scale may be arbitrarily chosen, such as $\|\vec{h}\|=1$, for instance.

Estimation of Homographies with RanSaC

Because of inherent noise, it is imperative to solve for H in a robust way. To this end, we can use RanSaC in this manner:

- For as long as necessary do
  - Select 4 correspondences at random
  - Compute H directly
  - Find inliers as $\|\vec{p}_{ri}, H\vec{p}_{li}\|<\epsilon$
- Identify H that leads to the largest set of inliers
- Recompute H with the largest set of inliers using Least-Squares

If we know the proportion of inliers π from the set of all correspondences, and given that a minimum of $n=4$ correspondences are needed for a solution, then the probability that we randomly choose $n=4$ inlier correspondences is given by π^n . Consequently, the probability that we have not picked a set of inliers after N iterations is given by $(1-\pi^n)^N$.

Consider a proportion of inliers $\pi=0.3$ (so mostly noise). The probability of randomly picking inliers during one RanSaC iteration is thus $0.3^4=0.0081$ or a 0.8% chance. If we use 100 iterations, the probability that we did not pick inliers once is $(1-0.3^4)^{100}=0.44$ or 44%. However if we use 1000 iterations, the probability of not having picked inliers once fall to one in 3400.

Overdetermined Solution

If more than 4 correspondences are available, then the equations provided by $A_i\vec{h}=\vec{0}$ are overdetermined. If all the correspondences are correct, then A will have rank 8 and $\vec{h}$ will be the correct solution. If, on the other hand, some correspondences are inaccurate (noise), then the only exact solution to $A_i\vec{h}=\vec{0}$ is $\vec{h}=\vec{0}$. To avoid this trivial solution, an additional constrain may be imposed such as $\|\vec{h}\|=1$. Instead of an exact solution we can require that $\vec{h}$ minimizes the norm $\|A\vec{h}\|$ subject to $\|\vec{h}\|=1$. The solution is then the unit eigenvector of

$A^T A$ associated with the smallest eigenvalue.

The direct linear transformation algorithm implements this approximative solution:

- Given $n \geq 4$ image correspondences $\{\vec{p}_{li}, \vec{p}_{ri}\}$, compute H such that $\vec{p}_{ri} \times H\,\vec{p}_{li} = 0$
- For each correspondence $\{\vec{p}_{li}, \vec{p}_{ri}\}$, compute the 2×9 matrix A_i
- Assemble the n 2 by 9 matrices A_i into a single $2n \times 9$ matrix A
- Perform the SVD of A. The unit singular vector corresponding to the smallest singular value is the solution $\vec{h}$.

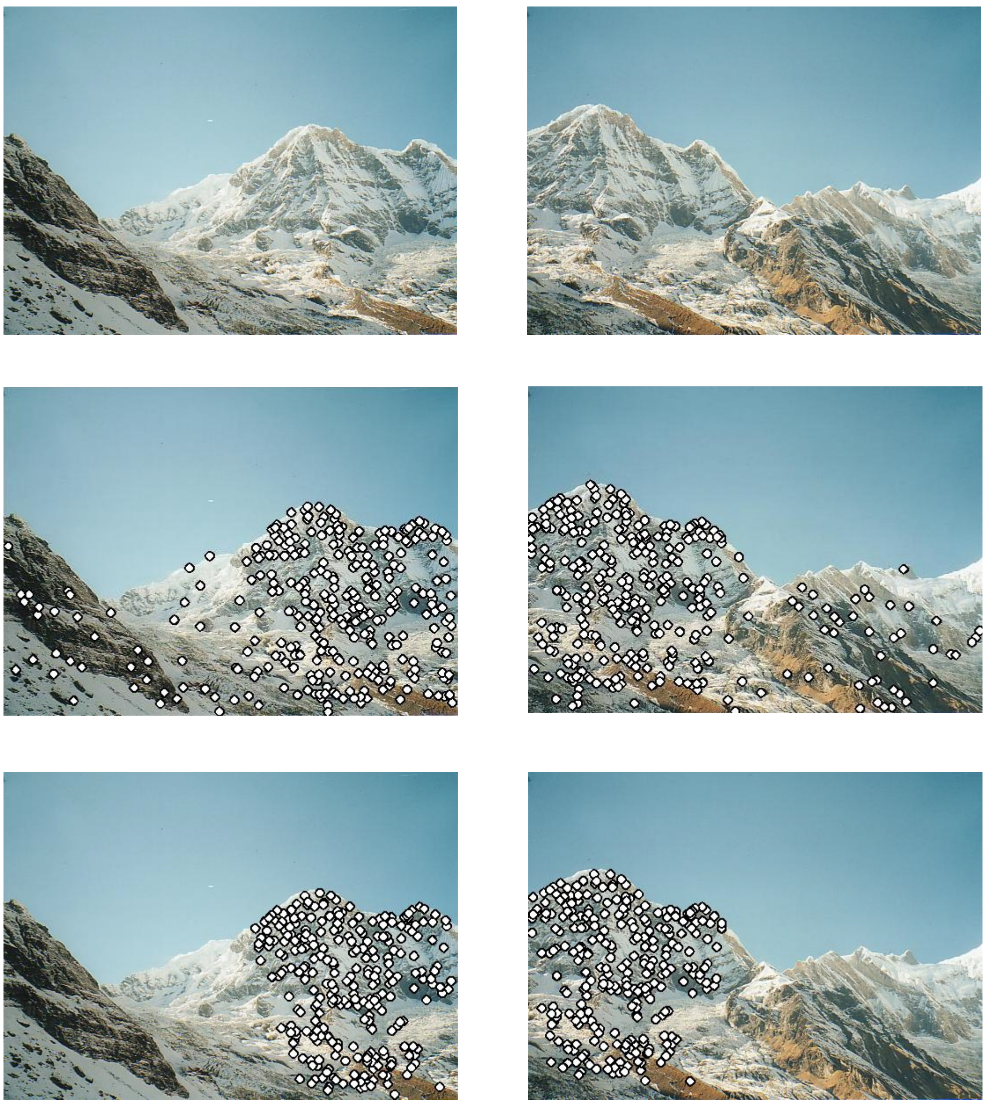

Illustration 1: The resulting stitched image

Solutions from Other Entities

Homographies can be computed from other image primitives than point correspondences. For instance, line correspondences can be used for the same purpose. With a line transformation $l_r = H^T l_l$, a matrix $A\vec{h} = \vec{0}$ can be derived with a minimal solution requiring 4 lines. Homographies can also be computed from conic correspondences, and so on.

Conditions for Homographies

Two images are related by a homography if and only if

- Both images are viewing the same plane from a different angle
- Both images are taken from the same camera but from a different angle
- Camera is rotated about its center of projection without any translation

Any homography relationship is independent of scene structure. In order to see this, consider that when to cameras have a non-zero baseline, epipolar geometry holds and thus there exists a transformation from a point in one image to a line in the other, with the actual mapping depending on the depth of the point in the scene. As the distance separating the two cameras goes to zero, so does the baseline. Triangulation (recovering depth from image disparity) is not feasible in this case. As a consequence, the homography maps a point in one image to another point in the other image in a way that is independent of depth (or 3D structure).

Known Rotation and Intrinsics

If the matrix of intrinsics M is known for the camera, and if somehow the rotation about its projection centre R can also be known, then obtaining the homography H is straightforward and performed in the following way:

$$H = MRM^{-1}$$

The matrix of the intrinsics is required here because we are working with pixel coordinates. In general, when one uses a camera and rotates around to take pictures, both the matrix of intrinsics and the rotation are unknown and require estimation through the identification and quantification of 2D image correspondences within overlapping image regions. When a photographer rotates around a point to take panoramic pictures, it is normally a fact that the optical center of the camera does not exactly coincide with the rotational point of the photographer. However, if this shift in rotational centers is small in comparison to the distance of objects being imaged by the camera then the resulting error is negligible.

Panoramic Imaging

- Higher resolution photographs, stitched from partly overlapping images
- Capture scenes that cannot be captured in one frame
- Cheaply and easily achieve effects that used to cost a lot of money
- Use computational methods to go beyond the physical limitations of the camera

Panoramic imaging, when it can be performed with computational techniques provides way to augment the capabilities of the physical sensors that are being used.

Illustration 2: Panoramic view created with automatic stitching from estimated homographies

Capturing panoramic images poses a number of challenges and one is constrained in a number of ways:

- Tripod versus hand-held
- Consistent exposure between frames yields smooth transitions
- Manual exposure makes consistent exposure of dynamic scenes easier
- Scenes do not have constant intensity everywhere
- Caution with distortion caused by lens
- Polarizing filters can cause problems
- Sharpness in image overlap region
- Image sequence requires a reasonable overlap (at least 15-30%) such that good features can be matched

Image Blending in Overlapping Regions

Given a feature match between two images that partially overlap (as it should for image stitching), the match in one image will be closer to its image borders than its converse in the other image (note that it can be closer to one border and further to an other border, with respect to its match). From this observation, a simple weighing function can be devised to give more importance to the image gain for the match that is further to its image boundaries. Other techniques for blending have been devised, some of them very advanced, involving Laplacian pyramids, wavelets, and so on.

When moving objects are present while taking a number of pictures to create a panorama, there is the possibility that moving objects will be present in the overlapping regions of images. The blending of these regions usually result in what is called ghosting. That is to say, the moving object will present itself twice, in different locations, in the fused part of the images. This is a problem that can be alleviated by various means, such as finding matches within overlapping regions of images that indicate motion and blending accordingly, such that only one instance of the moving object appears.