NTNU – Trondheim
Norwegian University of
Science and Technology

Applications of single- and multiple view geometry in computer vision

Håkon Hystad

August 16, 2017

1 Image stitching using shared points

If four or more shared points between two images are known the homography between them can be calculated. The resulting homograpy can be applied on one of the images to align its shared regions with the reference image. Doing so results in a panoramic (aka mosaic) image consisting of both images. This can be repeated with the panoramic image as a reference to add even more images.

General algorithm (modifed from [1] section 8.4).

1. Pick two images with a shared region.
2. Choose one as a reference.
3. Identify common points and compute the homography mapping the other image to the reference.
4. Transform the image with the found homography and lay it over the reference.
5. Repeat with the resulting mosaic to add more images.

Note that for alignment to be proper the two image sceneries must be considered planar since the homography maps between planes. Scenery with a great distance to the camera may be approximated as a plane, otherwise the images must be taken with pure rotation around the camera center.

When doing automatic stitching the corresponding points have to be found through feature extraction and comparison. Due to likely mismatches outlier detection (e.g RANSAC) have to be applied when fitting a homography to the data. If a linear transform is used to estimate the homography it is also important to normalize the data (with a centroid at (0,0) and a mean distance of $\sqrt{2}$ to it) beforehand to get proper results. This can be done with a similarity transform.

2 Reconstruction

From [1] chapter 10.

Reconstruction is the process of recovering/estimating points in space based on point correspondences and epipolar geometry. Having the fundamental/essential matrix lets one find the camera (projection) matrices so that the corresponding 2D points of the images can be triangulated to a 3D point in space.

2.1 Finding the camera matrices

From [1] section 9.5.

Given the camera calibration matrices the camera matrices are determined by

$$\boldsymbol{P} = \boldsymbol{K}\left[\boldsymbol{I}|\boldsymbol{0}\right], \boldsymbol{P'} = \boldsymbol{K'}\left[\boldsymbol{R}|\boldsymbol{t}\right] \tag{2.1}$$

Or they can be found from the essential matrix $\boldsymbol{E} = \boldsymbol{K'^T F K}$ when $\boldsymbol{F}$ is estimated.

If the cameras are uncalibrated the matrices can still be found up to a projective transformation since $\boldsymbol{F} = \boldsymbol{K'^{-T} R K^T}[\boldsymbol{K R^T}|\boldsymbol{t}]^{\times}$. The matrices are then given a *canonical* form where the calibration matrices are set to identity. The canonical camera matrices are made by choosing camera 1 as the reference frame which gives

$$\boldsymbol{P} = [\boldsymbol{I}|\boldsymbol{0}] \Rightarrow P' = [\boldsymbol{M}, \boldsymbol{e'}]$$

Where $\boldsymbol{M}$ is the transformation between the epipoles i.e $\boldsymbol{e} = \boldsymbol{M^{-1} e'}$ and so we require a full rank for $\boldsymbol{M}$ which $\boldsymbol{F}$ does not have. Letting $\boldsymbol{S} = \boldsymbol{s}^{\times}$ be a skew symmetric matrix such that $\boldsymbol{M} = \boldsymbol{S F} = \boldsymbol{s}^{\times}$ then full rank will be satisfied as long as $\boldsymbol{s^T e'} \neq 0$. A natural selection is thus $\boldsymbol{s} = \boldsymbol{e'}$ and we can now express the canonical camera matrices given $\boldsymbol{F}$:

$$\boldsymbol{P} = [\boldsymbol{I}|\boldsymbol{0}], P' = [\boldsymbol{e'}^{\times}\boldsymbol{F}, \boldsymbol{e'}] \tag{2.2}$$

2.2 Triangulation

From [1] section 12.2.

Given two corresponding image points projecting the same point $\boldsymbol{X} = (x, y, z, 1)^T$

$$\boldsymbol{x} = \boldsymbol{PX}, \boldsymbol{x'} = \boldsymbol{P'X}$$

we can formulate $\boldsymbol{AX} = \boldsymbol{0}$ with the help of the cross product:

$$\begin{aligned}
&\boldsymbol{x} \times (\boldsymbol{PX}) = \boldsymbol{0} \\
&\Rightarrow \\
&\quad x\boldsymbol{p^{3T}}\boldsymbol{X} - \boldsymbol{p^{1T}}\boldsymbol{X} = 0 \\
&\quad y\boldsymbol{p^{3T}}\boldsymbol{X} - \boldsymbol{p^{2T}}\boldsymbol{X} = 0 \\
&\quad x\boldsymbol{p^{2T}}\boldsymbol{X} - y\boldsymbol{p^{1T}}\boldsymbol{X} = 0
\end{aligned}$$

Where $\boldsymbol{p^i}$ denote the rows of $\boldsymbol{P}$. Using the independent parts for both points 4 equations are obtained and gathered in

$$\boldsymbol{A} = \begin{bmatrix} x\boldsymbol{p^{3T}} - \boldsymbol{p^{1T}} \\ y\boldsymbol{p^{3T}} - \boldsymbol{p^{2T}} \\ x'\boldsymbol{p'^{3T}} - \boldsymbol{p'^{1T}} \\ y'\boldsymbol{p'^{3T}} - \boldsymbol{p'^{2T}} \end{bmatrix} \tag{2.3}$$

And so $\boldsymbol{X}$ can be found with singular value decomposition.

As always it is important to normalize the coordinates before performing the SVD. Since the points are measurements prone to error it is best to fit the corresponding point to the closest epipolar line ensuring a compatibility with the epipolar geometry (here in the form of $\boldsymbol{P}$ and $\boldsymbol{P'}$). See algorithm 12.1 in Hartley-Zisserman.

2.3 Finding the plane induced homography

From [1] section 13.2.

Each plane in 3D has a homography which maps points via that plane. It is possible to calculate this homography given the three points defining the plane.

The camera matrices

$$\boldsymbol{P} = [\boldsymbol{I}|\boldsymbol{0}], \boldsymbol{P'} = [\boldsymbol{A}|\boldsymbol{e'}]$$

where $\boldsymbol{A} = \boldsymbol{e'}^{\times}\boldsymbol{F}$ projects points according to

$$\boldsymbol{x} = \boldsymbol{PX} = (\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3})^{\boldsymbol{T}}, \quad \boldsymbol{x'} = \boldsymbol{P'X} = [\boldsymbol{A}|\boldsymbol{e'}]\boldsymbol{X} = \boldsymbol{Ax} + \boldsymbol{e'}X_4$$

The world point is in a plane $\boldsymbol{\pi} = (\boldsymbol{v^T}, 1)^T$ and so $\boldsymbol{\pi^T X} = 0$ which determines the homogneous coordinate from $X_4 = -v^T x$. The equations become

$$\boldsymbol{x} = \boldsymbol{PX} = (\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3})^{\boldsymbol{T}}, \quad \boldsymbol{x'} = \boldsymbol{P'X} = [\boldsymbol{A}|\boldsymbol{e'}]\boldsymbol{X} = (\boldsymbol{A} - \boldsymbol{e'v^T})\boldsymbol{x}$$

And so the homography $\boldsymbol{x'} = \boldsymbol{Hx}$ is given by

$$\boldsymbol{H} = \boldsymbol{A} - \boldsymbol{e'v^T} \tag{2.4}$$

Since a vector crossed with it self is zero the equation $\boldsymbol{x'} \times \boldsymbol{Hx} = \boldsymbol{0}$ implies that

$$\boldsymbol{x_i'} \times (\boldsymbol{A}\boldsymbol{x_i} - \boldsymbol{e'}(\boldsymbol{v^T}\boldsymbol{x_i})) = \boldsymbol{x_i'} \times \boldsymbol{A}\boldsymbol{x_i} - \boldsymbol{x_i'} \times \boldsymbol{e'}(\boldsymbol{v^T}\boldsymbol{x_i}) = \boldsymbol{0}$$
$$\Rightarrow \quad \boldsymbol{x_i^T}\boldsymbol{v} = \frac{(\boldsymbol{x_i'} \times \boldsymbol{A}\boldsymbol{x_i})^{\boldsymbol{T}}(\boldsymbol{x_i'} \times \boldsymbol{e'})}{(\boldsymbol{x_i'} \times \boldsymbol{e'})^{\boldsymbol{T}}(\boldsymbol{x_i'} \times \boldsymbol{e'})} = b_i$$

Given three point correspondences it is therefore possible to determine the plane which forms $\boldsymbol{M}\boldsymbol{v} = \boldsymbol{b} \Rightarrow \boldsymbol{v} = \boldsymbol{M}^{-\boldsymbol{1}}\boldsymbol{b}$, where $\boldsymbol{M}$ is made of $\boldsymbol{x_i^T}$. The final expression for a plane induced homography given 3 point correspondences is therefore

$$\boldsymbol{H} = \boldsymbol{e'}^{\times}\boldsymbol{F} - \boldsymbol{e'}(\boldsymbol{M}^{-\boldsymbol{1}}\boldsymbol{b})^{\boldsymbol{T}} \tag{2.5}$$

2.4 Projective reconstruction

Without calibration or scene knowledge the best we can do is a projective reconstruction by triangulating with the canonical camera matrices.

2.5 Affine reconstruction

From [1] section 10.4.

An affine reconstruction is possible if extra knowledge is provided the image (e.g pure translation or parallel lines). Finding 3 sets of parallel lines lets one find the plane at infinity from their vanishing points. The plane at infinity can be used to recover the affine properties analog to the 2D case with the line at infinity.

Having the plane at infinity means that the infinite homography is given. The infinite homography transfers points between the two images via the plane at infinity. If two points

$$\boldsymbol{x} = \boldsymbol{M}\tilde{\boldsymbol{X}}, \boldsymbol{x'} = \boldsymbol{M'}\tilde{\boldsymbol{X}}$$

are the projection of a space point $\boldsymbol{X} = (\tilde{\boldsymbol{X}}^{\boldsymbol{T}}, 0)^T$ at infinity. Then

$$\boldsymbol{x'} = \boldsymbol{M'}\tilde{\boldsymbol{X}} = \boldsymbol{M'}\boldsymbol{M}^{-1}\boldsymbol{x} = \boldsymbol{H}_\infty \boldsymbol{x}$$

which makes the canonical camera matrices

$$\boldsymbol{P} = [\boldsymbol{I}|\boldsymbol{0}], \boldsymbol{P'} = [\boldsymbol{H}_\infty|\boldsymbol{e'}] \tag{2.6}$$

Using these matrices for triangulation results in affine recovery.

2.6 Metric reconstruction

Orthogonality constraint or calibration.

3 Rectification

From [2].

Since a point in 1 image will lie on an epipolar line in the other it is sufficient to search along this line to find a match. This search would be dramatically simplified if the epipolar lines where parallel so a search along the lines equated to searching along the horizontal pixel rows. It is possible to move the epipoles to infinity by two homographies to achieve this since the epipolar constraint is preserved:

$$\begin{aligned} \boldsymbol{\bar{x}} = \boldsymbol{Hx}, \boldsymbol{\bar{x}'} = \boldsymbol{H'x'} & \qquad (3.1) \\ \Rightarrow \quad \boldsymbol{\bar{x}'^T} \boldsymbol{\bar{F}} \boldsymbol{\bar{x}} = \boldsymbol{\bar{x}'^T} \boldsymbol{H'^T} \boldsymbol{\bar{F}} \boldsymbol{Hx} = 0 & \qquad (3.2) \end{aligned}$$

Where $\boldsymbol{F} = \boldsymbol{H'^T} \boldsymbol{\bar{F}} \boldsymbol{H}$. The homographies moving the epipoles to infinity are not unique and so we are free to choose $\boldsymbol{H}$. The desired form of the epipoles is $\boldsymbol{e} = (e_u, 0, 0)^T$ since the v-direction will also be aligned. The form can be achieved with

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{e_v}{e_u} & 1 & 0 \\ -\frac{1}{e_u} & 0 & 1 \end{bmatrix} \qquad (3.3)$$

To inspect the behavior of the applied homography the Jacobi determinant of $\boldsymbol{x'} = \boldsymbol{Hx}$ is analyzed

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} \end{bmatrix}$$

$$\det \boldsymbol{J} = \frac{e_u}{e_u - x}$$

We see that for points close to e_u, Hx is badly behaved and that $e_u >> x$ is needed to avoid severe distortions ($\det J < 1$ equals destruction of pixels, $\det J > 1$ equals creation). In practise this means that the two image planes should be pretty parallel to begin with to keep the epipoles out of the images.

The corresponding fundamental matrix of this epipole form is

$$\bar{\boldsymbol{F}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

The expressions for $\boldsymbol{H}$ and $\bar{\boldsymbol{F}}$ used in (3.1) leaves

$$\begin{bmatrix} h_{21}h'_{31} - h_31h'_{21} & h'_{31} & -h'_{21} \\ h_{21}h'_{32} - h_31h'_{22} & h'_{32} & -h'_{22} \\ h_{21}h'_{33} - h_31h'_{23} & h'_{33} & -h'_{23} \end{bmatrix} = \boldsymbol{F} \tag{3.4}$$

This is the solution for the last 2 rows of $\boldsymbol{H'}$ and we see that is overconstrained, note that a least square solution is preferable to direct inspection when $\boldsymbol{F}$ is imperfect. The first row is still undetermined and can initially be set to $(1, 0, 0)$ without affecting the epipolar constraint. The homographies are not unique and the desired ones are some that minimize distortion, both homographies are free to change as long as (3.1) holds. I.e $\boldsymbol{AHe} = \boldsymbol{He}$ which gives $\boldsymbol{A}$ the form

$$\boldsymbol{A} = \begin{bmatrix} & \boldsymbol{a^T} & \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Let $\boldsymbol{K} = \boldsymbol{AH}$ and $\boldsymbol{K'} = \boldsymbol{A'H'}$ be the new homographies

$$K = \begin{bmatrix} \boldsymbol{a^T h_1} & \boldsymbol{a^T h_2} & \boldsymbol{a^T h_3} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

where $\boldsymbol{h_i}$ are the column vectors of $\boldsymbol{H}$. It is desired (as previously discussed) to get the Jacobian determinant close to 1 which is equivalent to finding the singular values of the Jacobian close to 1[1]. The Jacobian for $\boldsymbol{Kx}, \boldsymbol{x} = (x, y, 1)^T$ is

$$\boldsymbol{J} = \frac{1}{d_3^2} \begin{bmatrix} \boldsymbol{a^T h_1} d_3 - h_{31} d_1 & \boldsymbol{a^T h_2} d_3 - h_{32} d_1 \\ h_{21} d_3 - h_{31} d_2 & h_{22} d_3 - h_{32} d_2 \end{bmatrix}$$

where $\boldsymbol{k_i^T}$ is the rows of $\boldsymbol{K}$ and $d_i = \boldsymbol{k_i^T x}$. It is seen that a_3 only adds a translation in the x-direction to the homography and does not affect distortion, this is confirmed by inspecting the Jacobian determinant:

$$\det \boldsymbol{J} = \frac{1}{\boldsymbol{\bar{h}_3^T x}} \left(-\boldsymbol{\bar{h}_2^T x}(a_1 h_{32} + a_2(h_{21} h_{32} - h_{22} h_{31})) + a_1(h_{22} h_{33} - h_{23} h_{32}) \right) \tag{3.5}$$

where $\boldsymbol{\bar{h}_i^T}$ is the *rows* of $\boldsymbol{H}$.

Therefore a_3 can be chosen freely. The cost function to minimize is then dependent on a_1 and a_2 as

$$\sum_i^n \left[(\sigma_{1,i} - 1)^2 + (\sigma_{2,i} - 1)^2 \right] \tag{3.6}$$

where $\sigma_{j,i}$ are the singular values $j = 1, 2$ of the Jacobian at some chosen points i (e.g corners of the image). The minimization may for instance be performed by a Nelder-Mead simplex search.

[1] $\det \boldsymbol{J} = \det \boldsymbol{U \Sigma V^T} = \det \boldsymbol{\Sigma}$

4 Projective depth

From [1] section 13.3.

Let $\boldsymbol{x_\pi}$ be the intersection of the ray $\vec{\boldsymbol{CX}}$ and a plane $\boldsymbol{\pi}$ where $\boldsymbol{X}$ denotes a world point and $\boldsymbol{C}$ the camera center. When $\boldsymbol{x_\pi}$ is depicted as $\bar{\boldsymbol{x}}'$ in another image with camera center $\boldsymbol{C'} \neq \boldsymbol{C}$ it will differ from the depicted world point $\boldsymbol{x'}$ as long as $\boldsymbol{X}$ is not on $\boldsymbol{\pi}$.

Let $\boldsymbol{x}$ be the projection of $\boldsymbol{x_\pi}$ on image $\boldsymbol{C}$ then its epipolar line in image $\boldsymbol{C'}$ is the projection of $\vec{\boldsymbol{CX}}$. This means that the epipole $\boldsymbol{e'}$, the image point $\bar{\boldsymbol{x}}'$ and the image $\boldsymbol{x'}$ are all on the same line. Further more the homography $\boldsymbol{H_\pi}$ induced by the plane $\boldsymbol{\pi}$ will map $\bar{\boldsymbol{x}}' = \boldsymbol{H_\pi x}$ and so

$$\boldsymbol{x'} = \boldsymbol{H_\pi x} + \rho \boldsymbol{e'} \tag{4.1}$$

Where ρ is the so called virtual parallax, the distance between $\boldsymbol{x'}$ and $\bar{\boldsymbol{x}}'$ along $\boldsymbol{e'}$.

This parallax can be thought of as projective depth. Although distance in the projective space has no real meaning the sign of the parallax will give information on what side of the plane $\boldsymbol{X}$ is. The projective depth is then given as

$$\frac{(\boldsymbol{x_{C'}} - \boldsymbol{H x_C})^{\boldsymbol{T}} \boldsymbol{e'}}{||\boldsymbol{e'}||} \tag{4.2}$$

Note that projective depth are not invariant under projective transformation, and so the side of a plane as well as the parallax to it are only meaningful for a specific projection.

References

[1] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision.* Cambridge University Press, Cambridge, 2nd ed. edition, 2003.

[2] John Mallon and Paul F. Whelan. Projective rectification from the fundamental matrix. *Image and Vision Computing*, 23(7):643–650, 2005.