Homography Estimation

Stéfan van der Walt

2010

1 Introduction

Consider two images, $f(x, y)$ and $f'(x, y)$, related by a geometric transformation. Given tentative point correspondences $\boldsymbol{p}_k \to \boldsymbol{p}'_k$ for $k = 0 \ldots N$, we want to estimate the transformation, T, such that

$$f(x, y) = f'(T(x, y)).$$

The function T can be very complex—the image formation process likely yields images that vary non-linearly in their geometry (radial lens distortion, for example, varies non-linearly as a function of the radius from the centre of the image). We limit ourselves to the simpler case where T is modelled as a linear coordinate transformation,

$$\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} = H \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix},$$

and the 3×3 matrix H represents a projection, encompassing rotation, scaling, skew and perspective. (Note the use of homogeneous coordinates, explained in the next section.)

2 The transformation matrix

The ability of the transformation matrix, H, to represent a projective transformation is made possible by extending the two-dimensional position vector, $[x, y]^T$, to three components. The value of the last dimension may vary, and defines an equivalence class so that

$$\begin{bmatrix} zx \\ zy \\ z \end{bmatrix} \equiv \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

Due to this property, these are known as homogeneous coordinates. Homogeneous coordinate are returned to Euclidean form simply by dividing with the last element.

The form of the transformation matrix H determines the type of geometric transformation represented. For example,

$$H = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \left[\begin{array}{c|c} R & \mathbf{0} \\ \hline \mathbf{0}^T & 1 \end{array}\right]$$

represents a rotation of angle θ. Making use of the "1" in the homogeneous coordinate, we can add translation

$$H = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & t_x \\ \sin(\theta) & \cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} = \left[\begin{array}{c|c} R & \mathbf{t} \\ \hline \mathbf{0}^T & 1 \end{array}\right],$$

and multiply the rotation matrix by s to get scaling:

$$H = \begin{bmatrix} s\cos(\theta) & -s\sin(\theta) & t_x \\ s\sin(\theta) & s\cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} = \left[\begin{array}{c|c} sR & \mathbf{t} \\ \hline \mathbf{0}^T & 1 \end{array}\right].$$

Skewing can be introduced by multiplying the x and y parameters a and b, e.g.,

$$H = \begin{bmatrix} sa\cos(\theta) & -sb\sin(\theta) & t_x \\ sa\sin(\theta) & sb\cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

while perspective is adjusted in the final row,

$$H = \begin{bmatrix} sa\cos(\theta) & -sb\sin(\theta) & t_x \\ sa\sin(\theta) & sb\cos(\theta) & t_y \\ p_0 & p_1 & 1 \end{bmatrix}.$$

In total, then, there are 8 parameters encoded in the H-matrix. Its elements are

$$H = \begin{bmatrix} H_{00} & H_{01} & H_{02} \\ H_{10} & H_{11} & H_{12} \\ H_{20} & H_{21} & 1 \end{bmatrix}$$

and, since H operates on homogeneous coordinates, it is homogeneous itself (we can always divide H by a constant without changing its function). Linear transformation matrices can be combined. For example, say we want to rotate an image around its centre at (x_c, y_c). We can express that operation as shifting the image upward and to the left, until its centre lies on the origin, rotating the image and then translating it back to its original position. The transformation matrix for this operation is

$$H = H_S^{-1} H_R H_S$$

where

$$H_S = \begin{bmatrix} 1 & 0 & -x_c \\ 0 & 1 & -y_c \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$H_{R=} \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The inverse H_S^{-1} has the same function as H_S, except that it translates in the opposite direction. Note that the order of applying transformations matters; each additional transformation must be pre-multiplied with the existing H.

3 Estimating a homography from correspondences (direct method)

We wish to estimate the 8 unknown parameters of the transformation matrix H, based on known point correspondences. The transformation of a source coordinate, $\mathbf{x}$, in Euclidean form (i.e., normalised so that $z = 1$) to a target coordinate $\mathbf{x}' = H\mathbf{x}$ yields

$$\begin{aligned} x' &= H_{00}x + H_{01}y + H_{02} \\ y' &= H_{10}x + H_{11}y + H_{12} \\ z' &= H_{20}x + H_{21}y + H_{22}. \end{aligned}$$

We now convert $\mathbf{x}'$ back to Euclidean form by dividing with z' . Then, we move all terms to the left side:

$$\begin{aligned} \frac{x'}{z'} - \frac{H_{00}x + H_{01}y + H_{02}}{H_{20}x + H_{21}y + H_{22}} &= 0 \\ \frac{y'}{z'} - \frac{H_{10}x + H_{11}y + H_{12}}{H_{20}x + H_{21}y + H_{22}} &= 0. \end{aligned}$$

Multiplying by the denominator yields

$$\begin{aligned}\frac{x'}{z'}(H_{20}x + H_{21}y + H_{22}) - H_{00}x - H_{01}y - H_{02} &= 0\\ \frac{y'}{z'}(H_{20}x + H_{21}y + H_{22}) - H_{10}x - H_{11}y - H_{12} &= 0\end{aligned}$$

which can also be written as the system

$$A\mathbf{h} = \begin{bmatrix} -x & -y & -1 & 0 & 0 & 0 & \frac{x'x}{z'} & \frac{x'y}{z'} & \frac{x'}{z'} \\ 0 & 0 & 0 & -x & -y & -1 & \frac{y'x}{z'} & \frac{y'y}{z'} & \frac{y'}{z'} \\ & & & & \vdots & & & & \end{bmatrix} \begin{bmatrix} H_{00} \\ H_{01} \\ H_{02} \\ H_{10} \\ H_{11} \\ H_{12} \\ H_{20} \\ H_{21} \\ H_{22} \end{bmatrix} = \mathbf{0}.$$

Each feature correspondence fills two rows of A, so that n point correspondences yields a $2N \times 9$ matrix.

We now have to solve the homogeneous set of linear equations,

$$A\mathbf{h} = \mathbf{0} \quad \mathbf{h} \neq \mathbf{0}.$$

For 4 point correspondences, the solution is the one-dimensional null-space of A. For more correspondences, we seek the solution to

$$\arg \min_{\|\mathbf{h}\|=1} \|A\mathbf{h}\| = \arg \min_{\|\mathbf{h}\|=1} \mathbf{h}^T A^T A\mathbf{h} = \lambda_{min} \tag{1}$$

where λ_{min} is the smallest eigenvalue of $A^T A$. This is easily shown, given the eigenvalues λ_i and corresponding eigenvectors $\mathbf{q}_i$ of $B = A^T A$. For

$$Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \dots & \mathbf{q}_n \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

it is true that $BQ = QD$ or $B = QDQ^T$. Rewriting (1) in terms of this factorisation yields

$$\arg \min_{\|\mathbf{h}\|=1} \mathbf{h}^T QDQ^T \mathbf{h} = \arg \min_{\|\mathbf{y}\|=1} \mathbf{y}^T D^T \mathbf{y} = \arg \min_{\|\mathbf{y}\|=1} \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2.$$

With $\lambda_i = \lambda_{min}$, a minimum is achieved when all components of $\mathbf{y}$ are set to zero except for $y_i = 1$. Since $\mathbf{y} = Q^T\mathbf{h}$, we find that $\mathbf{h} = Q\mathbf{y} = \mathbf{q}_{min}$, the eigenvector of B that corresponds to its smallest eigenvalue.

To find this eigenvector, examine the structure of the singular value decomposition (SVD),

$$A = U\Sigma V^T,$$

where the columns of U contain the eigenvectors of AA^T and the columns of V the eigenvectors of A^TA corresponding to the singular values of A on the diagonal of Σ. Recall that we are interested in the eigenvectors of A^TA since the vector $\mathbf{h}$ we are solving for is one such an eigenvector with its eigenvalue closest to zero.

The SVD can be (and usually is) computed so that the singular values appear in decreasing order on the diagonal of Σ. Note that the eigenvalues of the normal matrix A^TA are the squares of the singular values of H. If the system is exactly determined (i.e., with 4 point correspondences), there will be exactly one zero singular value in the last position, which corresponds to the last column of V—our solution. For an over-determined system, we choose the solution as the last column of V, corresponding to the smallest eigenvalue of A^TA.

Often, when the skew and perspective distortion is small, we can limit H to an affine transformation,

$$H = \begin{bmatrix} H_{00} & H_{01} & H_{02} \\ H_{01} & H_{11} & H_{12} \\ 0 & 0 & 1 \end{bmatrix},$$

where we only need to solve for 6 unknown parameters.