Structure From Motion: Tomasi-Kanade Factorization

1 The Orthographic Camera Model and the Low Rank Result

P points are tracked along F frames. Let u_{fp} and v_{fp} denote the 2D image location of the point p in frame f. Let X_p, Y_p and Z_p denote the 3D coordinates of point p. Let $U = (u_{fp})$, $V = (v_{fp})$ and $W = (w_{ij})$ where $w_{i,j} = u_{ij}$ and $w_{i+F,j} = v_{ij}$ for $1 \leq i \leq F$, i.e. $W = \begin{bmatrix} U \\ V \end{bmatrix}$.

In the orthographic camera model, points in the $3D$ world are projected in parallel onto the image plane. For example, if the camera's optical center is in the origin (w.r.t $3D$ coordinate system), and its u, v axes coincide with X, Y axes in the $3D$ world, then taking a picture is a simple projection: $\begin{pmatrix} u \\ v \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$. The depth, Z, has no influence on the image.

In this model, a camera can undergo rotation, translation, or a combination of the two:

$$\begin{pmatrix} u_{fp} \\ v_{fp} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \left(R'_f \begin{bmatrix} X_p \\ Y_p \\ Z_p \end{bmatrix} + t_f \right) \tag{1}$$

where R'_f is the 3×3 matrix that describes the rotation of the f frame and t_f is the 3×1 vector that describes the translation of frame f. Define $R_f = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} R'_f$ (R_f is a 2×3 matrix) then equation (1) becomes:

$$\begin{pmatrix} u_{fp} \\ v_{fp} \end{pmatrix} = R_f \begin{bmatrix} X_p \\ Y_p \\ Z_p \end{bmatrix} + t_f \tag{2}$$

Stacking together the projections of all the points at time (frame) f gives:

$$\begin{pmatrix} u_{f1} & \cdots & u_{fP} \\ v_{f1} & \cdots & v_{fP} \end{pmatrix} = R_f \begin{bmatrix} X_1 & \cdots & X_P \\ Y_1 & \cdots & Y_P \\ Z_1 & \cdots & Z_P \end{bmatrix} + \begin{pmatrix} t_f & \cdots & t_f \end{pmatrix} \tag{3}$$

Stacking the projections of all points along the entire sequence together yields:

$$\begin{pmatrix} u_{f1} & \cdots & u_{fP} \\ u_{21} & \cdots & u_{2P} \\ \cdots & & \\ u_{F1} & \cdots & u_{FP} \\ v_{11} & \cdots & v_{1P} \\ v_{F1} & \cdots & v_{FP} \end{pmatrix}_{2F \times P} = \begin{bmatrix} i_1^T \\ i_2^T \\ \vdots \\ i_F^T \\ j_1^T \\ j_2^T \\ \vdots \\ j_F^T \end{bmatrix}_{2F \times 3} \begin{bmatrix} X_1 & \cdots & X_P \\ Y_1 & \cdots & Y_P \\ Z_1 & \cdots & Z_P \end{bmatrix}_{3 \times P} + \begin{pmatrix} t_1^1 & \cdots & t_1^1 \\ t_2^1 & \cdots & t_2^1 \\ \vdots & & \\ t_F^1 & \cdots & t_F^1 \\ t_1^2 & \cdots & t_1^2 \\ t_2^2 & \cdots & t_1^1 \\ \vdots & & \\ t_F^2 & \cdots & t_F^2 \end{pmatrix}_{2F \times P} \tag{4}$$

where $R_f = \begin{bmatrix} i_f^T \\ j_f^T \end{bmatrix}$. Let $W = \begin{pmatrix} u_{f1} & \cdots & u_{fP} \\ u_{21} & \cdots & u_{2P} \\ \cdots & & \\ u_{F1} & \cdots & u_{FP} \\ v_{11} & \cdots & v_{1P} \\ v_{F1} & \cdots & v_{FP} \end{pmatrix}_{2F \times P}$, $M = \begin{bmatrix} i_1^T \\ i_2^T \\ \vdots \\ i_F^T \\ j_1^T \\ j_2^T \\ \vdots \\ j_F^T \end{bmatrix}_{2F \times 3}$ $S = \begin{bmatrix} X_1 & \cdots & X_P \\ Y_1 & \cdots & Y_P \\ Z_1 & \cdots & Z_P \end{bmatrix}_{3 \times P}$ and $T = \begin{pmatrix} t_1^1 & \cdots & t_1^1 \\ t_2^1 & \cdots & t_2^1 \\ \vdots & & \\ t_F^1 & \cdots & t_F^1 \\ t_1^2 & \cdots & t_1^2 \\ t_2^2 & \cdots & t_1^1 \\ \vdots & & \\ t_F^2 & \cdots & t_F^2 \end{pmatrix}_{2F \times P}$.

Using these notations, equation (4) is:

$$W = MS + T \tag{5}$$

The matrix M has 3 columns, hence $rank(M) \leq 3$. The matrix S has 3 rows, hence $rank(S) \leq 3$. All the columns of the matrix T are identical, hence $rank(T) \leq 1$. This implies $rank(W) \leq 4$.

1.1 Eliminating the Translation

Define a $2F \times P$ matrix $\tilde{W}$ by subtracting the mean of each row i of W from each of the entries of the ith row of W:

$$\tilde{W} = W \left(I - \frac{1}{P} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & & & \\ 1 & 1 & \cdots & 1 \end{pmatrix} \right) \tag{6}$$

Substituting equation (5) into equation (6) yields:

$$\tilde{W} = W \left(I - \frac{1}{P} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & & & \\ 1 & 1 & \cdots & 1 \end{pmatrix} \right) = \tag{7}$$

$$(MS + T)(I - \frac{1}{P} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & & & \\ 1 & 1 & \cdots & 1 \end{pmatrix}) = \tag{8}$$

$$MS(I - \frac{1}{P} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & & & \\ 1 & 1 & \cdots & 1 \end{pmatrix}) + T(I - \frac{1}{P} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & & & \\ 1 & 1 & \cdots & 1 \end{pmatrix}) = \tag{9}$$

$$M\hat{S} \tag{10}$$

where $\hat{S}$ is a the matrix of 3D coordinates centred around the origin (subtracting the mean of each coordinate from all the points). The last equation is true because subtracting the average of each row from the matrix T results in the matrix of all zeros.

1.2 Factorizing the matrix $\tilde{W}$

Our goal is to factorize the matrix $\tilde{W}$ to $M_{2F\times3}$ and $\hat{S}_{3\times P}$ such that $\tilde{W} = M\hat{S}$. The Singular Value Decomposition (SVD) of $\tilde{W}$ is:

$$\tilde{W} = UDV^T \tag{11}$$

where $U_{2F\times2F}$ and $V_{P\times P}$ are orthonormal matrices and the matrix D is diagonal (with non negative entries on its diagonal sorted in decreasing order). If there is no noise, the matrix $\tilde{W}$ is a rank 3 matrix. Hence only the first 3 entries on the diagonal of D are non-zeros. Equation 11 can be written as:

$$\tilde{W} = U'D'V'^T \tag{12}$$

where U' and V'^T are obtained by taking the first 3 columns of the matrices U and V respectively and the matrix D' is the principal 3×3 block of the matrix D.

Define $\tilde{M} = U'\sqrt{D'}$ and $\tilde{S} = \sqrt{D'}V'^T$ and we have factorized the matrix $\tilde{W}$ to matrices of the required dimensions. However, the factorization is not unique, and the matrices we have obtained are likely not be the matrices M and $\hat{S}$

1.3 Ambiguities and the Metric Constraints

The factorization of $\tilde{W}$ is not unique. For every 3×3 invertible matrix A,

$$\tilde{W} = \tilde{M}\tilde{S} = \tilde{M}AA^{-1}\tilde{S} = M\hat{S} \tag{13}$$

Our goal is to find an invertible A such that $M = \tilde{M}A$ and $S = A^{-1}\tilde{S}$.

To find such an A we will use the metric constraints:

$$i_f^T i_f = 1 \Rightarrow \tilde{i}_f^T AA^T \tilde{i}_f = 1 \tag{14}$$

$$j_f^T j_f = 1 \Rightarrow \tilde{j}_f^T AA^T \tilde{j}_f = 1 \tag{15}$$

$$i_f^T j_f = j_f^T i_f = 0 \Rightarrow \tilde{i}_f^T AA^T \tilde{j}_f = 0 \tag{16}$$

These equations are quadratic in A. To solve them, we define $C = AA^T$ and then solve linear equations for C (6 unknowns as C is a 3×3 symmetric matrix). Then we factorize C (using Cholesky or SVD) to obtain A.

After finding A, there still remains orthographic ambiguity (if Q is orthogonal and M satisfies the metric constraints so does MQ). this ambiguity corresponds to the specific definition of the 3D coordinate system.