Tomasi-Kanade Factorization Method

CMPE 264: Image Analysis and Computer Vision

Hai Tao

3D reconstruction with the orthographic camera model

- Problem statement
  - Assumption: an orthographic camera
  - Given: point correspondences across multiple views
  - Recover: camera motion and 3D coordinates for feature points

- We will describe a method proposed in Tomasi and Kanade'92. This method has been extended to deal with other affine cameras such as weak perspective cameras and paraperspective cameras. Based on the same principal, algorithms have also been developed for 3D reconstruction with perspective cameras and scenes with non-rigid motions

- Reference:
  - Forsyth and Ponce book: Section 14.3
  - Carlo Tomasi and Takeo Kanade, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, 9(2):137-154, November 1992.

Orthographic camera model

■ Recall that for an orthographic camera, the projection matrix is

$$
M = \begin{bmatrix} s_x & 0 & u_0 \\ 0 & s_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1^T & -\mathbf{r}_1^T T \\ \mathbf{r}_2^T & -\mathbf{r}_2^T T \\ 0 & 1 \end{bmatrix}
$$

Since the last row is [0,0,1], the projection process can be simplified as

$$
p = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \left(\begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \end{bmatrix} P + \begin{bmatrix} -\mathbf{r}_1^T T \\ -\mathbf{r}_2^T T \end{bmatrix} \right) + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}
$$

where $p=[x,y]^T$ and $P=[X,Y,Z]^T$. If we assume the camera intrinsic parameters (s_x,s_y,u_0,v_0) are known, then the image coordinates can be converted into the image plane coordinates system as

$$
p' = (p - [u_0,v_0]^T) \begin{bmatrix} 1/s_x & 0 \\ 0 & 1/s_y \end{bmatrix}
$$

For the ith feature point in the jth image, we will use $p_i^{(j)} = [x_i^{(j)}, y_i^{(j)}]^T$ to denote its image plane coordinates

Registered measurement matrix – remove the centroid

- P_i and it projection in the jth image $p_i^{(j)}$ are then related by

$$p_i^{(j)} = \begin{bmatrix} \mathbf{r}_1^{(j)T} \\ \mathbf{r}_2^{(j)T} \end{bmatrix} P_i + \begin{bmatrix} -\mathbf{r}_1^{(j)T} T^{(j)} \\ -\mathbf{r}_2^{(j)T} T^{(j)} \end{bmatrix}$$

- We can put image plane coordinates of M feature points together as

$$\begin{bmatrix} p_1^{(j)} & \cdots & p_M^{(j)} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^{(j)T} \\ \mathbf{r}_2^{(j)T} \end{bmatrix} \begin{bmatrix} P_1 & \cdots & P_M \end{bmatrix} + \begin{bmatrix} -\mathbf{r}_1^{(j)T} T^{(j)} & \cdots & -\mathbf{r}_1^{(j)T} T^{(j)} \\ -\mathbf{r}_2^{(j)T} T^{(j)} & \cdots & -\mathbf{r}_2^{(j)T} T^{(j)} \end{bmatrix} \quad (1)$$

Since for orthographic projection, the centroid $\overline{P}$ of the N scene points is mapped to the centroid of the N image feature points $\overline{p}$. Therefore,

$$\overline{p}^{(j)} = \begin{bmatrix} \mathbf{r}_1^{(j)T} \\ \mathbf{r}_2^{(j)T} \end{bmatrix} \overline{P} + \begin{bmatrix} -\mathbf{r}_1^{(j)T} T^{(j)} \\ -\mathbf{r}_2^{(j)T} T^{(j)} \end{bmatrix} \quad (2)$$

Rank theorem

- As the result,

$$\left[p_1^{(j)} - \overline{p}^{(j)} \quad \ldots \quad p_M^{(j)} - \overline{p}^{(j)} \right] = \begin{bmatrix} \mathbf{r}_1^{(j)T} \\ \mathbf{r}_2^{(j)T} \end{bmatrix} \left[P_1 - \overline{P} \quad \ldots \quad P_M - \overline{P} \right]$$

- W can combine image correspondences from *N* images to form the *registered measurement matrix W*

$$W = \begin{bmatrix} p_1^{(1)} - \overline{p}^{(1)} & \ldots & p_M^{(1)} - \overline{p}^{(1)} \\ & \ldots & \\ p_1^{(N)} - \overline{p}^{(N)} & \ldots & p_M^{(N)} - \overline{p}^{(N)} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^{(1)T} \\ \mathbf{r}_2^{(1)T} \\ \ldots \\ \mathbf{r}_1^{(N)T} \\ \mathbf{r}_2^{(N)T} \end{bmatrix} \left[P_1 - \overline{P} \quad \ldots \quad P_M - \overline{P} \right] = RS$$

- Rank Theorem: Without the noise, the registered measure matrix W is at most rank three (Because the *motion matrix R* and the *structure matrix S* are rank three)

Reconstruction up to a 3 by 3 transformation

- With noise, the rank of the registered measurement matrix may not be three. This problem, however, can be solved by keeping only the first three greatest eigenvalues in the singular value decomposition of W. More specifically, if the SVD of W is $W=UDV^T$ and

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_M & \\ & 0 & & \end{bmatrix}$$

then $W'=U'D'V'^T$, or

$$W' = \begin{bmatrix} U_1 & U_2 & U_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_3 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \\ V_3^T \end{bmatrix}$$

We assign $\hat{R}=U'D'$ and $\hat{S}=V'$

This solution is up to a 3 by 3 rank three matrix Q, because $\hat{R}'=\hat{R}Q$ and $\hat{S}'=Q^{-1}\hat{S}$ is also a solution

The metric constraints

■ From the structure of the *R* matrix, we notice that addition conditions need to be satisfied. These conditions are

$$\hat{\mathbf{r}}'^{T}_{2i-1}\hat{\mathbf{r}}'_{2i-1} = \hat{\mathbf{r}}^{T}_{2i-1}QQ^{T}\hat{\mathbf{r}}_{2i-1} = 1$$

$$\hat{\mathbf{r}}'^{T}_{2i}\hat{\mathbf{r}}'_{2i} = \hat{\mathbf{r}}^{T}_{2i}QQ^{T}\hat{\mathbf{r}}_{2i} = 1$$

$$\hat{\mathbf{r}}'^{T}_{2i}\hat{\mathbf{r}}'_{2i-1} = \hat{\mathbf{r}}^{T}_{2i}QQ^{T}\hat{\mathbf{r}}_{2i-1} = 0$$

where *i=1,…,N*, $\hat{\mathbf{r}}_k$ is the *k*th row of $\hat{R}$

To solve for *Q*, there are two methods

- Solve for *B* first, where $B = QQ^{T}$, using a linear method. The *Q* can be recovered via the Cholesky decomposition. One problem of this method is that the linear solution of *B* may not be positive definite.
- Directly solve these equations via non-linear least squares methods.

The rotation ambiguity

- Once Q is estimated, we obtained the solution $\hat{R}' = \hat{R}Q$ and $\hat{S}' = Q^{-1}\hat{S}$ where $\hat{R}'$ satisfies the metric constraints

- The solution is still up to a rotation matrix because, for any rotation matrix R_0, $\hat{R}'' = \hat{R}'R_0$ and $\hat{S}'' = R_0^T\hat{S}'$ are also solutions. To eliminate this ambiguity, we can determine R_0 by forcing the first motion matrix as

$$\begin{bmatrix} \hat{\mathbf{r}}''^{(1)T}_1 \\ \hat{\mathbf{r}}''^{(2)T}_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$R_0 = \begin{bmatrix} \hat{\mathbf{r}}'^{(1)}_1 & \hat{\mathbf{r}}'^{(1)}_2 & \hat{\mathbf{r}}'^{(1)}_1 \times \hat{\mathbf{r}}'^{(1)}_2 \end{bmatrix}$$

therefore

$$W \approx \hat{R}''\hat{S}''$$

- Finally, we obtain the solution

The structure and the motion

■ The recovered structure is represented in the coordinate system with the centroid as the origin

■ The rotation matrix of the jth camera is

$$R^{(j)} = \begin{bmatrix} \mathbf{r''}_1^{(j)} & \mathbf{r''}_2^{(j)} & \mathbf{r''}_3^{(j)} \end{bmatrix}$$

■ To obtain the translation of the jth camera, observe that

$$\overline{p}^{(j)} = \begin{bmatrix} \mathbf{r''}_1^{(j)T} \\ \mathbf{r''}_2^{(j)T} \end{bmatrix} [0] + \begin{bmatrix} -\mathbf{r''}_1^{(j)T} T^{(j)} \\ -\mathbf{r''}_2^{(j)T} T^{(j)} \end{bmatrix} = \begin{bmatrix} -\mathbf{r''}_1^{(j)T} T^{(j)} \\ -\mathbf{r''}_2^{(j)T} T^{(j)} \end{bmatrix}$$

which yields $T^{(j)} = \begin{bmatrix} \mathbf{r''}_1^{(j)} & \mathbf{r''}_2^{(j)} & \mathbf{r''}_3^{(j)} \end{bmatrix} \begin{bmatrix} \overline{p}^{(j)} \\ \alpha \end{bmatrix}$

where α is an arbitrary real number

Summary of the algorithm

- Tomasi-Kanade factorization algorithm
  - Obtain the registered measurement matrix W
  - Compute the SVD of W as $W=UDV^T$ and keep the first three largest eigenvalues and obtain $W'=U'D'V'^T$. Define $\hat{R}=U'D'$ and $\hat{S}=V'$
  - Compute Q so that $\hat{R}'=\hat{R}Q$ satisfies the following metric constrains and define $\hat{S}'=Q^{-1}\hat{S}$

  $$\hat{\mathbf{r}}^T_{2i-1}QQ^T\hat{\mathbf{r}}_{2i-1}=1 \qquad \hat{\mathbf{r}}^T_{2i}QQ^T\hat{\mathbf{r}}_{2i}=1 \qquad \hat{\mathbf{r}}^T_{2i}QQ^T\hat{\mathbf{r}}_{2i-1}=0$$

  - Compute $R_0 = \begin{bmatrix} \hat{\mathbf{r}}'^{(1)}_1 & \hat{\mathbf{r}}'^{(1)}_2 & \hat{\mathbf{r}}'^{(1)}_1\times\hat{\mathbf{r}}'^{(1)}_2 \end{bmatrix}$ and define $\hat{R}''=\hat{R}'R_0$ and $\hat{S}''=R_0^T\hat{S}'$
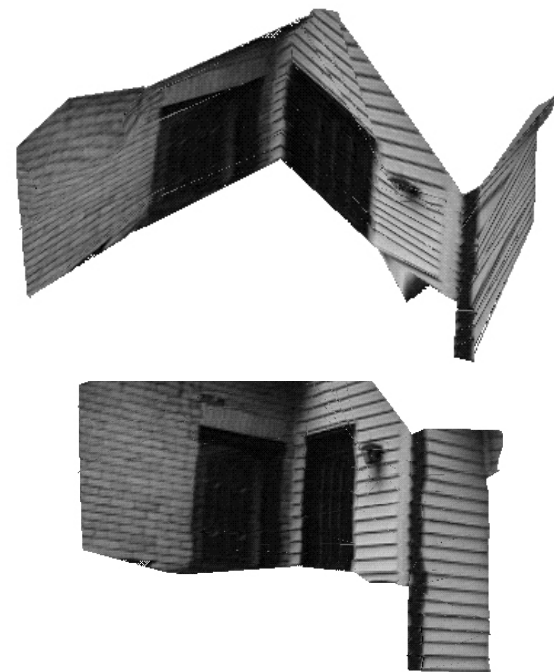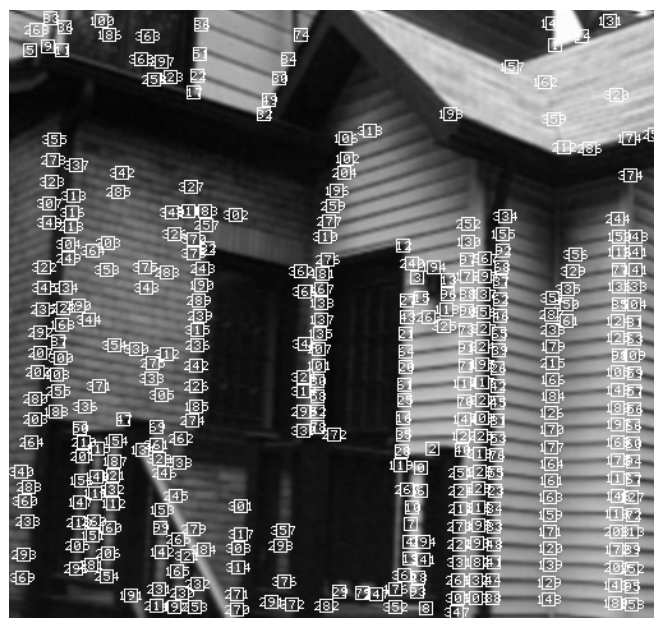 - Output structure as $\hat{S}''$ and motion for the jth camera as

  $$R^{(j)} = \begin{bmatrix} \mathbf{r}''^{(j)}_1 & \mathbf{r}''^{(j)}_2 & \mathbf{r}''^{(j)}_3 \end{bmatrix}$$

  $$T^{(j)} = \begin{bmatrix} \mathbf{r}''^{(j)}_1 & \mathbf{r}''^{(j)}_2 & \mathbf{r}''^{(j)}_3 \end{bmatrix}\begin{bmatrix} \overline{p}^{(j)} \\ \alpha \end{bmatrix}$$

  where α is an arbitrary real number

Experimental results

Reprinted from Tomasi and Kanade 1992