18CSE390T
Computer Vision

Perspective and Projective Factorization

Factorization

- When processing video sequences, we often get extended *feature track* from which it is possible to recover the structure and motion using a process called *factorization*.

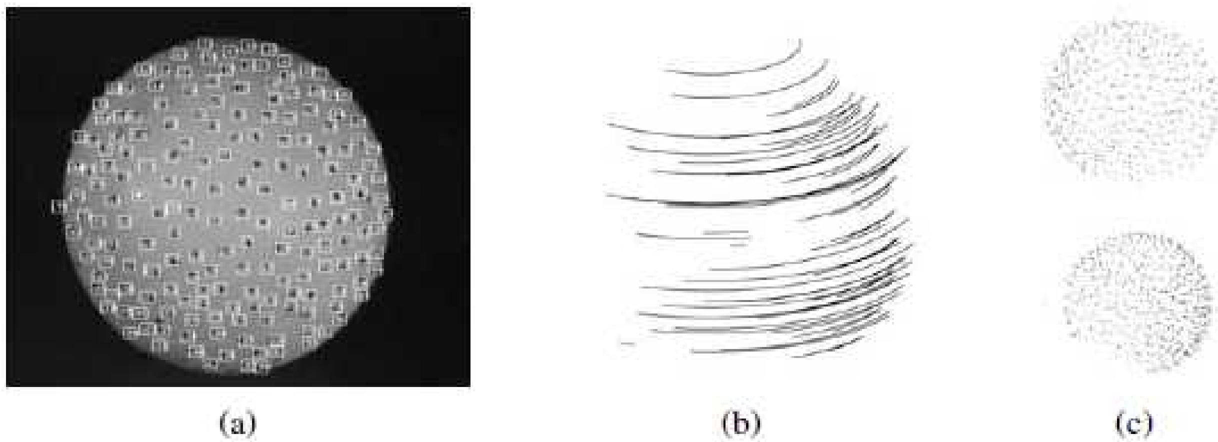

Figure: 3D reconstruction of a rotating ping pong ball using factorization (Tomasi and Kanade 1992) : (a) sample image with tracked features overlaid; (b) sub-sampled feature motion stream ; (c) two views of the reconstructed 3D model.

Factorization (cont.)

- Consider orthographic and weak perspective projection models.
- Since the last row is always [0001], there is no perspective division

x_{ij}: location of $\boldsymbol{i}^{\text{th}}$ point $\boldsymbol{x}_{ji} = \tilde{\boldsymbol{P}}_j \bar{\boldsymbol{p}}_i$,

$\tilde{\boldsymbol{P}}_j$: upper 2 4 portion of projection matrix P_j

$\bar{\boldsymbol{p}}_i$ = $(X_i, Y_i, Z_i, 1)$: augmented 3D point position

Factorization (cont.)

- Assume that every point ***i*** is visible in every frame ***j***.

  We can take the centroid (average) of the projected point locations $\boldsymbol{x}_{ij}$ in frame ***j***.

$$\bar{\boldsymbol{x}}_j = \frac{1}{N}\sum_i \boldsymbol{x}_{ji} = \tilde{\boldsymbol{P}}_j \frac{1}{N}\sum_i \bar{\boldsymbol{p}}_i = \tilde{\boldsymbol{P}}_j \bar{\boldsymbol{c}},$$

$\bar{\boldsymbol{c}} = (\bar{X}, \bar{Y}, \bar{Z}, 1)$ gmented 3D centroid of the point cloud.

- so that $\bar{X} = \bar{Y} = \bar{Z} = 0,$ $\bar{\boldsymbol{c}} = (0,0,0,1).$
- Centroid of 2D points in each frame $\bar{\boldsymbol{x}}_j$ directly gives us last element of $\tilde{\boldsymbol{P}}_j$

Factorization (cont.)

- Let $\tilde{\boldsymbol{x}}_{ji} = \boldsymbol{x}_{ji} - \bar{\boldsymbol{x}}_j$ be the 2D point locations after their image centroid has been

we can write;

$$\tilde{\boldsymbol{x}}_{ji} = \boldsymbol{M}_j \boldsymbol{p}_i,$$

$\boldsymbol{M_j}$: upper 2 by 3 portion of the projection matrix $\boldsymbol{P_j}$ and $\boldsymbol{p_i=(X_i,Y_i,Z_i)}$

- We can concatenate all of these measurements into one large matrix.

Factorization (cont.)

$$\hat{X} = \begin{bmatrix} \tilde{x}_{11} & \cdots & \tilde{x}_{1i} & \cdots & \tilde{x}_{1N} \\ \vdots & & \vdots & & \vdots \\ \tilde{x}_{j1} & \cdots & \tilde{x}_{ji} & \cdots & \tilde{x}_{jN} \\ \vdots & & \vdots & & \vdots \\ \tilde{x}_{M1} & \cdots & \tilde{x}_{Mi} & \cdots & \tilde{x}_{MN} \end{bmatrix} = \begin{bmatrix} M_1 \\ \vdots \\ M_j \\ \vdots \\ M_M \end{bmatrix} \begin{bmatrix} p_1 & \cdots & p_i & \cdots & p_N \end{bmatrix} = \hat{M}\hat{S}.$$

- $\hat{X}$: measurement matrix
- $\hat{M}$: motion matrices
- $\hat{S}$: structure matrices

Factorization (cont.)

- If SVD of $\hat{X} = U\Sigma V^T$ directly returns the matrices $\hat{M}$ and $\hat{S}$; but it does not. Instead we can write the relationship

$$\hat{M} = UQ \text{ and } \hat{X} = U\Sigma V^T = [UQ][Q^{-1}\Sigma V^T]$$

$$\hat{S} = Q^{-1}\Sigma V^T.$$

- To recover values of the 3 ✖ 3 matrix Q depends on motion model being used.
- In the case of orthographic projection, the entries in M_j are the first two rows of rotation matrices R_j.

Factorization (cont.)

- So we have

$$
\begin{aligned}
\boldsymbol{m}_{j0} \cdot \boldsymbol{m}_{j0} = &\quad \boldsymbol{u}_{2j} \boldsymbol{Q}\boldsymbol{Q}^T \boldsymbol{u}_{2j}^T &= 1, \\
\boldsymbol{m}_{j0} \cdot \boldsymbol{m}_{j1} = &\quad \boldsymbol{u}_{2j} \boldsymbol{Q}\boldsymbol{Q}^T \boldsymbol{u}_{2j+1}^T &= 0, \\
\boldsymbol{m}_{j1} \cdot \boldsymbol{m}_{j1} = &\quad \boldsymbol{u}_{2j+1} \boldsymbol{Q}\boldsymbol{Q}^T \boldsymbol{u}_{2j+1}^T &= 1,
\end{aligned}
$$

$\boldsymbol{u}_k$: 3 × 1 rows of matrix $\boldsymbol{U}$.

- This gives us a large set of equations for the entries in matrix $\boldsymbol{QQ}^T$ from which matrix $\boldsymbol{Q}$ can be recovered using matrix square root.

Perspective and Projective Factorization

- Factorization disadvantage is that it cannot deal with perspective cameras.
- Perform an initial affine (e.g.,orthographic) reconstruction and to then correct for the perspective effects in an iterative manner.
- Observe that object-centered projection model

$$y_s = s\frac{\boldsymbol{r}_y \cdot \boldsymbol{p} + t_y}{1 + \eta_z \boldsymbol{r}_z \cdot \boldsymbol{p}} + c_y$$

differ from scaled orthographic projection model $\boldsymbol{x}_{ji} = \tilde{\boldsymbol{P}}_j \bar{\boldsymbol{p}}_i,$

Perspective and Projective Factorization (cont).

$$
\begin{aligned}
x_{ji} &= s_j \frac{\boldsymbol{r}_{xj} \cdot \boldsymbol{p}_i + t_{xj}}{1 + \eta_j \boldsymbol{r}_{zj} \cdot \boldsymbol{p}_i} \\
y_{ji} &= s_j \frac{\boldsymbol{r}_{yj} \cdot \boldsymbol{p}_i + t_{yj}}{1 + \eta_j \boldsymbol{r}_{zj} \cdot \boldsymbol{p}_i}
\end{aligned}
$$

- By inclusion of denominator terms $(1 + \eta_j \boldsymbol{r}_{zj} \cdot \boldsymbol{p}_i)$.
- If we knew correct values of $\eta_j = t_{zj}^{-1}$ motion parameters R_j and p_j; we cross multiply left hand side by denominator and get correct values for which bilinear projection model is exact.

Perspective and Projective Factorization (cont).

- Once the n_j have been estimated, the feature locations can then be corrected before applying another factorization.

- Because of the initial depth reversal ambiguity both reconstructions have to be tried while computing n_j.

Perspective and Projective Factorization (cont).

- Alternative approach which does not assume calibrated cameras (known optical center, square pixels, and zero skew) is to perform *fully projective factorization.*

$$x_{ji} = \tilde{P}_j \bar{p}_i,$$

- The inclusion of third row of camera matrix

  Is equivalent to multiplying each reconstructed measurement $x_{ij}=M_j p_i$ by its inverse depth

$$\eta_{ji} = d_{ji}^{-1} = 1/(P_{j2} p_i)$$

- Or equivalently multiplying each measured position by its projective depth d_{ji}.

Perspective and Projective Factorization (cont).

$$\hat{X} = \begin{bmatrix} d_{11}\tilde{x}_{11} & \cdots & d_{1i}\tilde{x}_{1i} & \cdots & d_{1N}\tilde{x}_{1N} \\ \vdots & & \vdots & & \vdots \\ d_{j1}\tilde{x}_{j1} & \cdots & d_{ji}\tilde{x}_{ji} & \cdots & d_{jN}\tilde{x}_{jN} \\ \vdots & & \vdots & & \vdots \\ d_{M1}\tilde{x}_{M1} & \cdots & d_{Mi}\tilde{x}_{Mi} & \cdots & d_{MN}\tilde{x}_{MN} \end{bmatrix} = \hat{M}\hat{S}.$$

- Factorization method provides a "closed form" (linear) method to initialize iterative techniques such as bundle adjustment.