Motion Segmentation Problem

Given a set of image points obtain:
- Number of independently moving objects
- Segmentation: object to which each point belongs
- Motion: rotation and translation of each object
- Structure: depth of each point

Three frames from the MPEG “flower garden” sequence

• Given optical flow at each point
• partition/segment the flow field into regions belonging to individual planes “layers”

Figure from “Representing Images with layers,” by J. Wang and E.H. Adelson, IEEE Transactions on Image Processing, 1994, c 1994, IEEE
Some example slides from Forsythe and Ponce. Computer Vision, A modern approach.

Model Estimation and Grouping

• Given a set of data points and a particular model
• The model parameters can be estimated by LSE fitting data to the model
• Previous model examples – essential/fundamental matrix, homographies, lines/planes
• In order to estimate the model parameters we need to know which data point belongs to which model
• Difficulty – we have multiple models – we do not know initially which data point belongs to which model and we do not the model parameters
• chicken and egg problem

Line Example
- Set of points belonging to two lines
- We need to estimate
  1. parameters of the lines slope and intercept
  2. which point belongs to which line

Solution: EM algorithm
Idea: Each of the above steps Assumes the other one is solved and iterate
EM algorithm

- Basic structure of the EM algorithm
- 1. Start with random parameter values for each model
- 2. Iterate until parameter values converge
  - E step: assign points to the model that fits best
  - M step: update the parameters of the models using only points assigned to it

Simplistic explanation here –
Theoretical foundation probabilistic (model parameters are random variables) - EM (Expectation Maximization)

---

E-Step

- Case of two lines given by slopes and intercepts \((a_1, b_1)\) and \((a_2, b_2)\)
- For each data point \(i\), estimate the residual (difference between the prediction and the model)
  \[ r_1(i) = y_i - (a_1 x_i + b_1) \]
  \[ r_2(i) = y_i - (a_2 x_i + b_2) \]
- Calculate the weights, which correspond to the probabilities of particular data point belonging to particular model
  \[ w_1(i) = \frac{e^{-\frac{r_1(i)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \]
  \[ w_2(i) = \frac{e^{-\frac{r_2(i)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \]

---

M-step

- Given the weights recalculate the parameters of the model \((a_1, b_1)\) and \((a_2, b_2)\)
- Least squares estimation of line parameters
  \[ \sum_i x_i^2 \sum_i x_i \sum_i 1 | a \quad b = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \]
- In our case we will have weighted least squares estimation of line parameters
  \[ \sum_i w_i x_i^2 \sum_i w_i x_i \sum_i w_i | a \quad b = \frac{\sum_i w_i x_i y_i}{\sum_i w_i x_i^2} \]
- Solve such estimation problem twice – once for each line

---

M-step

- Iterate until the parameters of the lines converge
- Issues with EM
  - Local maxima
    - can be a serious nuisance in some problems
    - no guarantee that we have reached the “right” maximum
- Starting if we do not know how many models we have
  - k means to cluster the points is often a good idea
Application to motion segmentation

- Consider motion model, when the flow field can be approximated by some parametric model with small number of parameters
- We can write x and y parameters of the flow field – assume that models are locally translational, i.e. we can locally approximate the model by pure translation \((\alpha, \beta)\)
- Suppose entire flow field can be explained by two translational models \((\alpha_1, \beta_1) \) \((\alpha_2, \beta_2)\)
- EM algorithm can be applied in analogous way

\[
\begin{align*}
\mathbf{r}_1(x, y) &= (\alpha_1 - x) \mathbf{r}_1(x, y) + (\beta_1 - y) \\
\mathbf{r}_2(x, y) &= (\alpha_2 - x) \mathbf{r}_2(x, y) + (\beta_2 - y)
\end{align*}
\]

\[
\begin{align*}
\mathbf{w}_1(x, y) &= \frac{\exp(-\mathbf{r}_1'(x, y))}{\sum_{i,j} \exp(-\mathbf{r}_1'(i,j))} \\
\mathbf{w}_2(x, y) &= \frac{\exp(-\mathbf{r}_2'(x, y))}{\sum_{i,j} \exp(-\mathbf{r}_2'(i,j))}
\end{align*}
\]

- M-step analogous to line fitting

\[
\begin{align*}
\sum_{i,j} \mathbf{w}_1(i, j) & 0 \left[ \begin{array}{c} \mathbf{u}_1 \\ \mathbf{v}_1 \end{array} \right] = \left[ \begin{array}{c} \sum_{i,j} \mathbf{w}_1(i, j) \mathbf{v}_2(i, j) \\ \sum_{i,j} \mathbf{w}_1(i, j) \mathbf{y}_2(i, j) \end{array} \right]
\end{align*}
\]

Iterate until convergence

Application for Motion Segmentation

- Model image pair (or video sequence) as consisting of regions of parametric motion
  - affine motion – commonly used –
  - Approximates locally motion of the planar surface

\[
\begin{bmatrix}
\mathbf{v}_x(x, y) \\
\mathbf{v}_y(x, y)
\end{bmatrix} = \begin{bmatrix}
\alpha_{11} & \alpha_{12} \\
\alpha_{21} & \alpha_{22}
\end{bmatrix} \begin{bmatrix}
x \\
y
\end{bmatrix} + \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix}
\]

- Now we need to
  - determine which pixels belong to which region
  - estimate parameters

Three frames from the MPEG “flower garden” sequence

Figure from “Representing Images with layers,” by J. Wang and E.H. Adelson, IEEE Transactions on Image Processing, 1994, c 1994, IEEE
Grey level shows region no. with highest probability

Segments and motion fields associated with them

Figure from "Representing Images with layers," by J. Wang and E.H. Adelson, IEEE Transactions on Image Processing, 1994, c 1994, IEEE

Other examples

- Segmentation
  - a segment is a gaussian that emits feature vectors (which could contain color; or color and position; or color, texture and position).
  - segment parameters are mean and (perhaps) covariance
  - if we knew which segment each point belonged to, estimating these parameters would be easy

Some generalities

- Many, but not all problems that can be attacked with EM can also be attacked with RANSAC
- need to be able to get a parameter estimate with a manageably small number of random choices.
- RANSAC is usually better
- We wish to choose a model to fit to data
  - e.g. is it a line or a circle?
  - e.g. is this a perspective or orthographic camera?
  - e.g. is there an aeroplane there or is it noise?
- In general, models with more parameters will fit a dataset better, but are poorer at prediction

Top is not necessarily a better fit than bottom
(actually, almost always worse)
Motion Segmentation Problem

Given a set of image points obtain:

- Number of independently moving objects
- Segmentation: object to which each point belongs
- Motion: rotation and translation of each object
- Structure: depth of each point

Alternative Algebraic approach GPCA

- Generalized principal component analysis
- Main idea – see hand written notes
- Automatic determination of the number of models
- Model parameter estimation, via polynomial factorization

Single Motion Epipolar Constraint

\[ x = [x, y, z]^T \in \mathbb{R}^3 \]
\[ x_0^T \bar{R} x_1 = \mathbf{C} \]
\[ F = \bar{T} \in \mathbb{R}^{3 \times 2} \]
\[ x_0^T F x_1 = \mathbf{C} \]

\( e \equiv T = \) epipole or translation
\( e^T \bar{F} = 0 \) lies in the nullspace of \( \bar{F} \)

Multibody Epipolar Constraint

\[ \{ (R_i, T_i) \}_{i=1}^n \]
\[ \{ F_i \equiv T_i R_i \}_{i=1}^n \]
\[ \exists F_i : x_0^T F_i x_1 = \mathbf{C} \]
\[ \prod_{i=1}^n (x_0^T F_i x_1) = \mathbf{C} \]
Multibody Fundamental Matrix

\[ \prod_{i=1}^{n} (x_i^T F x_i) = 0 \]

\[ \nu_n(x_i)^T F v_n(x_i) = 0 \]

\[ F = \sum_{\sigma \in \mathcal{B}_n} F_{\sigma(1)} \otimes \cdots \otimes F_{\sigma(n)} \]

Embedding – Veronese map

Veronese map \( \nu_n = [x_1, x_2, \ldots, x_k]^T \rightarrow [\ldots, x^n, \ldots] \)

ranges over all monomials of degree \( n \) sorted in lexicographical order.

Number of such monomials is

\[ C_n = \binom{n+k-1}{n} \]

Example \( k = 3 \) and \( n = 2 \)

\[ C_2 = \begin{pmatrix} n + 3 - 1 \\ 3 - 1 \end{pmatrix} = \frac{(n+1)(n+2)}{2} \]

\[ \nu_2(x) = [x^2, xy, xz, yz, y^2, z^2] \]

Number of Independent Motions

Theorem: Given image points corresponding to \( N \) motions, if at least 8 points correspond to each object, then

\[ \text{Minimum number of points} \]

\[ \begin{array}{cccc}
1 & 2 & 3 & 4 \\
1 & 35 & 99 & 225 \\
\end{array} \]
Decomposition of Multibody $F$

- $n$ motions generate $n$ groups of epipolar lines
- Each group will intersect at the epipole $e_i$
- Details Chapter 7