This lecture was given by Yair Weiss.

- Wertheimer principles of grouping
- Grouping based on motion
- Horn & Schunk algorithm
- Layered approach to grouping
- Conclusions

1 Wertheimer's principles of grouping

Recall from Wertheimer's principles of grouping,[1] the principle factors which lead to grouping include:

- proximity
- similarity (brightness, color, shape, texture, motion, etc.)
- good continuation of boundary contours
- closure
- symmetry and parallelism
- familiar configuration

2 Grouping based on motion

Grouping based on motion is difficult because of the *aperture problem*. Figure 1 gives a graphical representation of the aperture problem. When we only see the local motion of a contour, we can only see the component of the motion orthogonal to the local contour. That is, if we see the edge moving to the right, we have no way of knowing whether the motion of the edge has any vertical component.

The aperture problem can be described mathematically by the equation

$$I_x v_x + I_y v_y + I_t = 0 \tag{1}$$

[1]See the paper "Laws of organization in perceptual forms" in `http://www.yorku.ca/dept/psych/classics/Wertheimer/Forms/forms.htm`

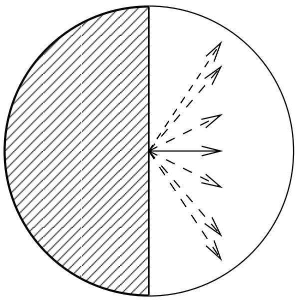

Figure 1: Graphical deptiction of the aperture problem

Given an image brightness pattern sequenced in time, it is a good approximation that locally the overall brightness stays constant over small intervals of time. Denoting the image sequence as $I(x,y,t)$, the approximation says that

$$\begin{aligned} 0 &= \frac{d}{dt} I(x,y,t) \\ &= \frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dx}{dt} + \frac{\partial I}{\partial t} \\ &=: I_x v_x + I_y v_y + I_t \end{aligned}$$

Equation (1) comes directly from the constant brightness condition. $(I_x, I_y) \in \mathbf{R}^2$ is the gradient of the image brightness and $(v_x, v_y) \in \mathbf{R}^2$ is the velocity of the point (x,y). In the example illustrated in Figure 1 the brightness gradient of the center pixel is porportional to $(I_x, I_y) = (1,0)$. Notice that for this image brightness gradient, equation (1) does not constrain v_y. That means for a given horizontal motion v_x any vertical component of motion is consistent with the constraint (1). The aperture problem was demonstrated by a video of the two squares moving diagonally, where we only had local views of edges of the squares.

3 Horn & Schunk algorithm

Horn & Schunk (1982) proposed an algorithm to compute the image velocity vector field by combining the local velocity measurements with a smoothness cost. The algorithm incorporates the constraint in equation (1) with a cost for non-smooth vector fields, thus attempting to regularize the aperture problem by combining local measurements. The cost function considered in the Horn & Schunk algorithm is:

$$\begin{aligned} J(v) &= \sum_{(x,y)\in I} (I_x v_x + I_y v_y + I_t)^2 + && (2) \\ & \quad \gamma_1 \|v(x,y) - v(x-1,y)\|^2 + \gamma_2 \|v(x,y) - v(x,y-1)\|^2 && (3) \end{aligned}$$

Then a probability distribution on velocity vector fields is given by

$$p(v) = \frac{1}{\lambda} e^{-J(v)} \tag{4}$$

Equation (4) is a special case of a Markov Random Field (MRF) since the velocity field only depends on its nearest neighbors. In the cost function $J(v)$, the term (2) can be thought of as the likelihood term, and the term (3) which is a smoothness cost can be thought of as the prior.

There are several criticisms of the Horn & Schunk algorithm including

- Is this the right prior to use?
- There's no notion of grouping
- Fails if there is more than one moving object

In the cases that there is more than one moving object, the smoothness constraint would actually be detrimental. In such cases, the constraint should really be a piecewise smoothness constraint, with possible discontinuities at the boundaries of the moving objects. One approach to incorporate this idea would be to add a line process to the cost function.

However, there is evidence to suggest that Horn & Schunk's algorithm is not the right model for the way the human visual system works. To demonstrate this, there was a demo video of two ellipses rotating. The thin ellips looked as if it were one rigid body rotating, but the fat ellipse looked as if it were deforming. Figure 2 should remind the reader of the video.

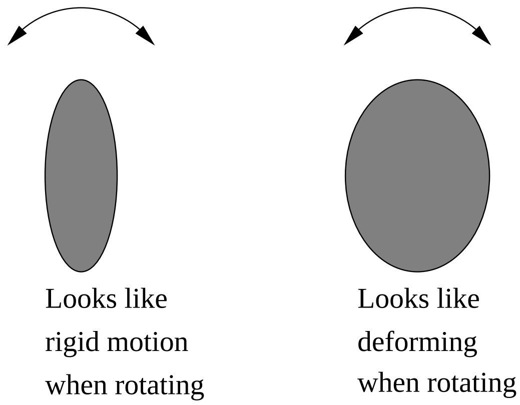

Figure 2: Two ellipses rotating

Also, there are experiments that show that people are willing to group objects across motion discontinuities. For example, there was a demo video of an ellipse with 4 satallite points that were rotating together on top of a static textured background. The fact that there was static texture between the ellipse and the satellite points does not stop people from grouping the ellipse and the satellite points.

To paraphrase a quote the Italian Gestalt psychologist Matteli (since I didn't catch the exact wording):

> We don't measure the motion of points, but the motion of the group to which the points belong.

4 Layered Approach to Grouping

An approach to explaining why humans group objects across motion discontinuities is the approach of requiring smoothness in layers. Within each layer, there should be some smoothness in the velocity vector field. This layered approach would explain the fact people are willing to group object across motion discontinuities.

For illustration purposes, we were shown an example that looked roughly like Figure 3.

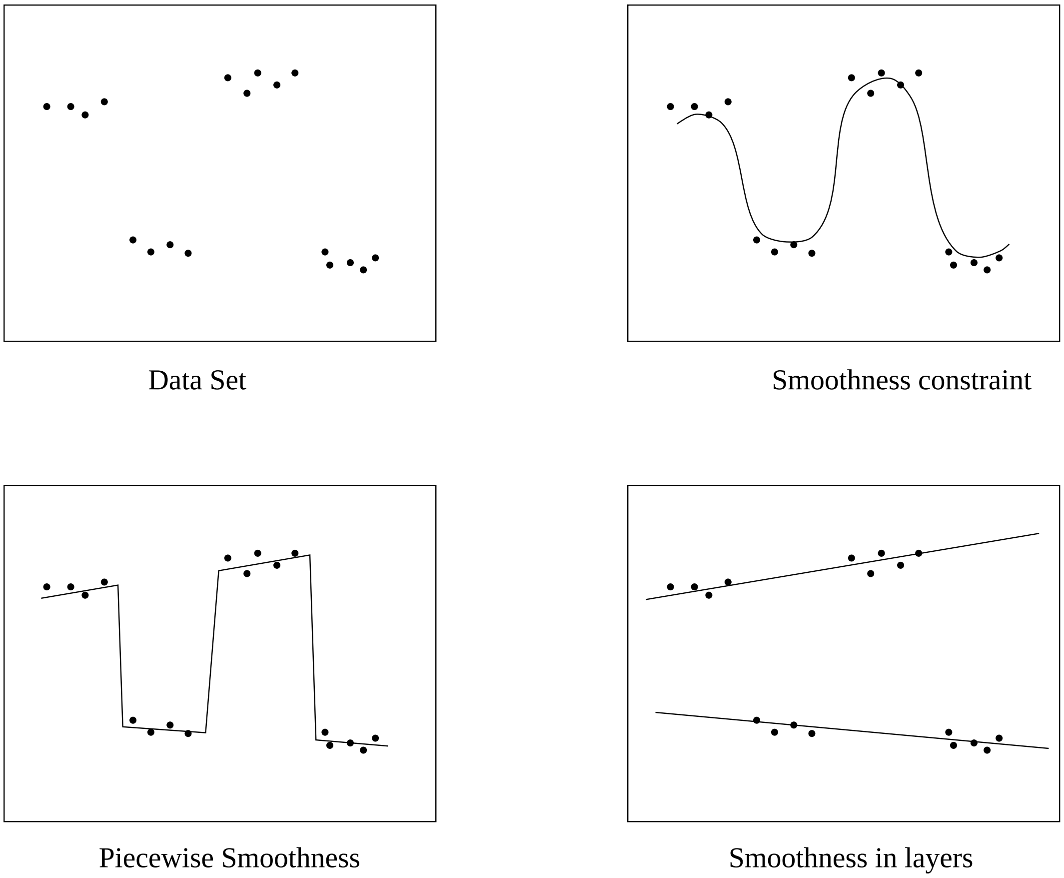

Figure 3: Example of smoothness in layers

Next we briefly reviewed the Expectation Maximization algorithm. Recall that the EM algorithm is an example of mixture estimation where the hidden variables we are trying to estimate are the group to which the data points belong. We were shown a demonstration of fitting two lines to a given data set. We also saw a demo of curve fitting for a group of lines

4.1 Parametric vs. Non-parametric motion estimation

Line estimation in EM is easy because we are only trying to estimate 2 parameters (slope, etc). People have also tried to do parametric motion estimation by assuming that each layer undergoes an affine motion. That is, assuming that for each layer, the motion can be parameterized by:

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \begin{pmatrix} a & b \\ d & e \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} c \\ f \end{pmatrix} \tag{5}$$

and then trying to estimate a, b, c, d, e, f. The referrence for parametric motion estimation is Ayer & Sawhney. However an affine motion model is not a good approximation if the scene is not roughly planar. Weiss has studied the case of nonparametric motion estimation.

The model selection problem (figuring out the number of groups necessary for the mixture estimation) is in general very hard to solve. There have been many different approaches to solve the model selection problem. One approach is to find the *smallest* number of layers consistent with the motion.

5 Conclusions

The main message of this lecture was that grouping based on common fate is more difficult and is different than grouping based on similarity. Also grouping based on motion can not be done locally.