Motion Modeis: Part I

A.L. Yuille and D. Kersten

Abstract

I. Motion

This section describes two types of models. The first performs simple motion measurement. The second integrates motion spatially to obtain a more global percept. .

Motion perception is locally ambiguous as illustrated in figure (1). Short-range motion suffers from the aperture problem while long-range motion has the correspondence problem (long range motion will be covered in the next lecture). Prior models of plausible motion are required to resolve these ambiguities. We briefly describe the history of the slow-and-smooth prior to illustrate how justification has become more quantitative over time.

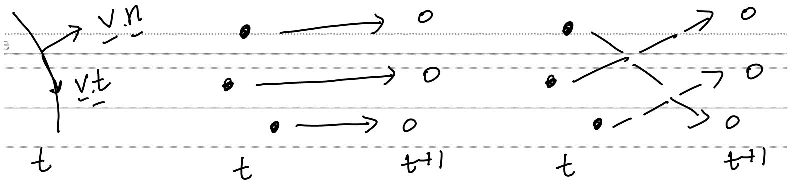

Fig. 1. The aperture problem arises for short-range motion of contours (left) where only the component $\vec{v} \cdot \vec{n}$ of the velocity normal to the contour can be observed and the tangential component $\vec{v} \cdot \vec{t}$ is unknown. The correspondence problem arises for long-range motion because there are many ways to match the dots at time t with those at $t+1$. The human vision system typically prefers slow-and-smooth matching (center panel) although other matches are possible (right panel).

The input to motion measurement is a set of images. Some of the earliest work was done by Reichardt *et al.* who developed a correlation model based on studies of the fly and beetle visual systems [**?**]. Later, mechanisms of directional selectivity were studied in vertebrate retina of the rabbit [**?**]. Some phenomena of human motion perception can be explained by modified versions of these early correlation or correlation-like models [**?**], [**?**], [**?**], [**?**],[**?**]. We describe the basis of these types of models in section (I-A).

Early computational studies (Ullman [**?**]) showed that several perceptual phenomena of long-range motion could be described by a ‘minimal mapping’ theory that, in Bayesian terms, assumed a slowness prior. Subsequent work showed that smoothness priors accounted for findings on short-range motion (Hildreth [**?**]), including the surprising fact that an ellipse rotating in the image plane is perceived to move non-rigidly. Yuille and Grzywacz [**?**] showed that a slow-and-smooth prior could account for a large range of motion perceptual phenomena – including motion capture and motion cooperation – both for short- and long-range motion. Weiss and his collaborators showed that slow and slow-and-smooth priors (Weiss and Adelson [**?**], Weiss et al. [**?**]) could explain other short-range motion phenomena, such as how percepts can change dramatically as we alter the balance between the likelihood and prior terms (i.e. for some stimuli the prior dominates the likelihood and vice versa). These theories are discussed in section (I-C).

A. *Motion Measurement: Spatio-Temporal Filters*

Spatio-temporal filters are biologically plausible ways to measure motion which agree with properties of cells in the visual cortex. The standard model suggests two classes of cells where the first are spatio-temporal filters which are sensitive to the directions of motion while the second combine outputs of these filters to estimate the motion itself [**?**] (Grzywacz and Yuille, Simoncelli, Schrater et al. refs!!).

Measuring the motion velocity assumes that locally the intensity can be modeled as a linear translating pattern:

$$I(\vec{x}, t) = F(\vec{x} - \vec{v}t). \tag{1}$$

Differentiating with respect to $\vec{x}$ and t (using $\vec{\nabla} I = \vec{\nabla} F$ and $\frac{\partial I}{\partial t} = -\vec{v} \cdot \vec{\nabla} F$), gives the *optical flow equation*:

$$\vec{v} \cdot \vec{\nabla} I + \frac{\partial I}{\partial t} = 0. \tag{2}$$

A similar argument applies if we filter the image by any spatial-temporal filter $G^{\mu}(\vec{x}, t)$ to obtain:

$$G^{\mu} * I(\vec{x}, t) = \int G(\vec{x} - \vec{y}, t - s) I(\vec{y}, s) ds d\vec{y}. \tag{3}$$

Hence each filter gives a constraint on the velocity,

$$\vec{v} \cdot \vec{\nabla} G^{\mu} * I + \frac{\partial G^{\mu} * I}{\partial t} = 0. \tag{4}$$

By applying Fourier analysis to equation (3), it can be shown that filters tuned to frequencies $\vec{\omega}, \omega_t$ will tend to have the biggest response if $\vec{v} \cdot \vec{\omega} + \omega_t = 0$. No realistic filter can be tuned to a single frequency (the filter would be $\exp\{i(\vec{\omega} \cdot \vec{x} + \omega_t t)\}$ which has infinite spatial and temporal extent). But it can be shown (Yuille and Grzywacz!!) that if we use spatio-temporal Gabor functions weakly tuned to $\vec{\omega}, \omega_t$ (Gabor filters satisfy optimal localization in both space-time and frequency) then the biggest responses to the stimuli occur when $\vec{v} \cdot \vec{\omega} + \omega_t = 0$ (it is unlikely that the real receptive fields are spatio-temporal Gabors, but they are probably reasonable approximations).

This gives a mathematical justification for the two stage model. The first set of cells are tuned roughly to spatial-temporal filters and estimate properties of the motion. The next stage uses the activities of the most responsive cells to estimate the velocity. See figure (2).

Fig. 2. Spatio-temporal Gabors

These models give reasonable fits to the activities of cells in the early visual cortex.

B. Smootheness Assumption on Contours: Hildreth's Theory

Consider a contour $\vec{r}(s)$ where s is the arc-length parameter (i.e. $|(d\vec{r})/ds| = 1$). The normal to the contour is $\vec{n}(s)$. Suppose the real velocity of the contour is $\vec{v}(s)$. The normal component of the velocity is $u(s)$. The observation model is:

$$\int ds \{\vec{n}(s) \cdot \vec{n}(s) - u(s)\}^2. \tag{5}$$

This does not provide enough information to estimate $\vec{v}(s)$ uniquely. Hildreth proposed imposing a smoothness requirement of the velocity. This is done by adding a term $\lambda \int ds \{\frac{\partial \vec{v}}{\partial s} \cdot \frac{\partial \vec{v}}{\partial s}\}$ to the observation model. This gives:

$$E[\vec{v}] = \int ds \{\vec{n}(s) \cdot \vec{n}(s) - u(s)\}^2 + \lambda \int ds \{\frac{\partial \vec{v}}{\partial s} \cdot \frac{\partial \vec{v}}{\partial s}\}. \tag{6}$$

The minimum of $E(\vec{v})$ can be found by solving the Euler-Lagrange equations:

$$\lambda \frac{\partial^2 \vec{v}}{\partial s^2} = \{\vec{n}(s) \cdot \vec{v}(s) - u(s)\} \vec{n}(s). \tag{7}$$

The solution to this equation gives roughly the correct solution for the rotating ellipse – i.e. it predicts non-rigid rotation. It is also easy to analyze the solutions and show that this theory only gives the true/veridical result if the velocity $\vec{(}s)$ is constant (i.e. independent of s). To see this, the correct solution occurs only if $\vec{n}(s) \cdot \vec{v}(s) - u(s) = 0$. This implies that $\frac{\partial^2 \vec{v}}{\partial s^2} = 0$. This gives a solution $\frac{\partial \vec{v}}{\partial s} = \vec{c}$, which is constant. But the constraint that the contour is closed means that $\vec{c} = 0$, hence $\vec{v}$ is constant.

C. Simple Slow-and-Smooth: Example of Multi-Dimensional Gaussian .

There is plenty of psychophysical evidence that local motion estimates are pooled to give a global perception. The evidence suggests that this assumes that the motion is locally slow and smooth. This section describes a simple version of this model for short-range motion based on [**?**]. This can be formulated within the probabilistic framework and to simplify the mathematics we restrict the probability distributions to be Gaussians (note this simplification leaves out some of the higher-order smoothness terms which are needed for some phenomena).

The model is formulated as estimating the two dimensional velocities $(U, V) = \{(U_i, V_i) : \ i \in \Lambda\}$ defined over an image lattice Λ. Smoothness is defined over a local neighborhood $Nbh(i)$ defined on the lattice, this is nearest neighbor in this example – see figure (3).

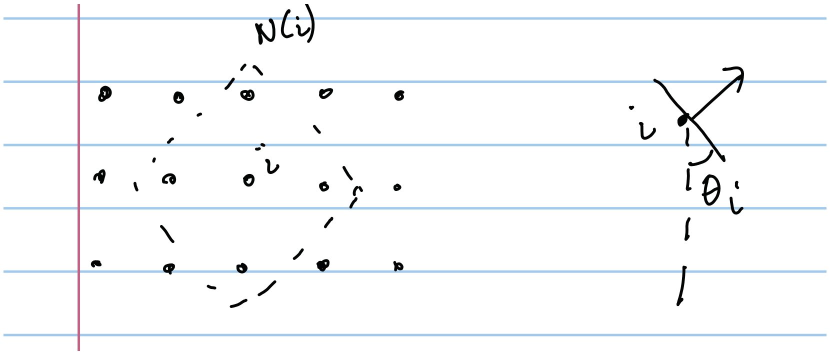

Fig. 3. The slow-and-smooth model is formulated on the image lattice (left) with neighborhood structure given by nearest neighbors in the horizontal and vertical directions. The model assumes that we can only directly observe the velocity component normal to the aperture (right).

The likelihood functions and the slow-and-smoothness prior are defined by Gibbs distributions:

$$P(D|U,V) = \frac{1}{Z} \exp\{-E[D; U, V]\},$$
$$P(U,V) = \frac{1}{Z} \exp\{-E(U,V)\}. \tag{8}$$

With

$$E[D; U, V] = \sum_{i \in \Lambda} \gamma_i (U_i sin\theta_i + V_i \cos\theta_i)^2$$
$$E(U,V) = \alpha \sum_{i \in \Lambda} \{U_i^2 + V_i^2\} + \beta \sum_{i \in \Lambda} \sum_{j \in Nbh(i)} \{(U_i - U_j)^2 + (V_i - V_j)^2\}. \tag{9}$$

The data term assumes that we can only observe one component of the velocity specified by a known angle θ_i. The parameter $\gamma_i = 0$ if there are no observations at lattice site i, and otherwise $\gamma_i = 1/(2\sigma_i^2)$ where σ_i^2 is the variance of the data at i. The prior terms imposes both slowness and smoothness terms – weighted by α and β respectively.

The posterior distribution $P(U,V|D) \propto P(D|U,V)P(U,V)$ is a Gaussian. This is because both $P(D|U,V)$ and $P(U,V)$ are Gaussians (and the conjugate of a Gaussian is also a Gaussian).

We estimate the most probable motion $(\hat{U}, \hat{V})$ from $P(U, V|D)$. For Gaussian distributions, the MAP estimate and the mean estimate are identical. Both reduce to minimizing the energy function $E(U, V) + E(D; U, V)$ which is quadratic in (U, V) and so, after the differentiation, we obtain the following equations:

$$\begin{aligned}
&\alpha \hat{U}_i + \beta \sum_{j \in Nbh(i)} (\hat{U}_i - \hat{U}_j) - \gamma_i \{D_i - \sin\theta_i \hat{U}_i - \cos\theta_i(\hat{V}_i)\} \sin\theta_i, \ \forall i \in \Lambda \\
&\alpha \hat{V}_i + \beta \sum_{j \in Nbh(i)} (\hat{V}_i - \hat{V}_j) - \gamma_i \{D_i - \sin\theta_i \hat{U}_i - \cos\theta_i(\hat{V}_i)\} \cos\theta_i, \ \forall i \in \Lambda.
\end{aligned} \tag{10}$$

These linear equations can be solved by standard packages. But we now give some intuition for them by considering several special cases, shown in figure (4).

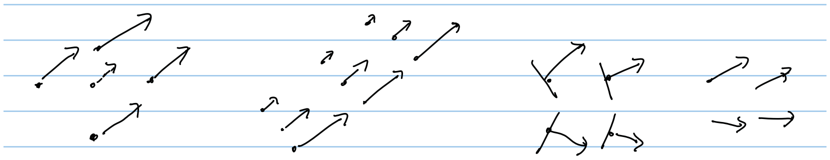

Fig. 4. At lattice nodes with no observations the estimated velocity will be less than the average of its neighbors – see central pixel (far left) and the top left pixels (left). Otherwise the estimated velocity will tend to smooth out the observations (right) yielding a smoother percept (far right).

First, suppose we are at a position where there is no observation and so $\gamma_i = 0$. In this case, the estimated velocity at i is a sub-average of the velocities of its neighbors:

$$\hat{U}_i = \frac{\beta \sum_{j \in Nbh(i)} \hat{U}_j}{\alpha + |Nbh|\beta}, \ \hat{V}_i = \frac{\beta \sum_{j \in Nbh(i)} \hat{V}_j}{\alpha + |Nbh|\beta}. \tag{11}$$

If there is no slowness – i.e. $\alpha = 0$ – then the velocity estimate $(\hat{U}_i, \hat{U}_j)$ is an average of the velocity of its neighbors, but if $\alpha > 0$ then the estimates are lower. So slowness means that estimate of motion speed decreases in regions where there are no observations, in agreement with experiments. If there is no smoothness – $\beta = 0$, then the estimate of velocity is zero at node i.

Second, consider a lattice node where there is an observation. This gives estimates which encourage similarity to the motion estimates for the neighbors and also agreement with the observations.

$$\begin{aligned}
\hat{U}_i &= \frac{\beta \sum_{j \in Nbh(i)} \hat{U}_j + \gamma_i D_i \sin\theta_i}{\alpha + \beta|Nbh| + \gamma_i \sin^2\theta_i}, \\
\hat{V}_i &= \frac{\beta \sum_{j \in Nbh(i)} \hat{V}_j + \gamma_i D_i \cos\theta_i}{\alpha + \beta|Nbh| + \gamma_i \cos^2\theta_i}.
\end{aligned} \tag{12}$$

A special case occurs when we set β which removes the smoothness constraint. In this case we obtain:

$$\hat{U}_i = \frac{\gamma_i D_i \sin\theta_i}{\alpha + \gamma_i \sin^2\theta_i}, \ \hat{V}_i = \frac{\gamma_i D_i \cos\theta_i}{\alpha + \gamma_i \cos^2\theta_i}. \tag{13}$$

It can be shown that this encourages the estimated motion to be direction $(\sin\theta_i, \cos\theta_i)$.

Weiss *et al* [**?**] imposed a slowness prior and studied the trade-of's between the measurement and the prior terms, by varying the strength of the parameters γ_i. This showed the effect of luminance contrast on the perception of motion where the speed and direction of moving patterns depends on the contrast. For example, this analysis explains the odd combination of facts that a thin horizontally moving rhombus appears to move diagonally at low contrasts and horizontally at high contrasts, whereas a fat rhombus appears to move horizontally at all contrasts. When contrast is low, retinal image information becomes less reliable, and so the Bayesian ideal observer shifts more weight to the prior probability distribution on motion velocity; this shift in relative weight alters the optimal estimate of speed and direction. This can be explained by the model above, by seeing how the behavior varies if we changes the data parameter γ.

The original use of the slow-and-smooth prior was justified on intuitive grounds and on its ability to model experimental findings. Theoretical arguments [**?**],[**?**] showed that slow-and-smoothness priors result from the geometry of image formation (perspective or orthographic projection) provided we assume that objects in the three-dimensional scene are equally likely to move in all directions and make weak, and reasonable, assumptions about the distribution of their speed. But this did not give quantitative form for the priors. Studies of human performance give ways to estimate motion priors (Stocker and Simoncelli [**?**]) which, for slow-and-smooth motion, show that the priors that humans use are more robust than those originally proposed [**?**],[**?**] More recently, it has been possible to learn motion priors for natural image motions (Black and Roth [**?**],[**?**]). Their findings show biases to slow-and-smooth and are similar, but more robust, to the specific forms used in the models. The differences between these models correspond to different choices of potentials.

D. Slow-And-Smooth: Spatial Fall-off

Psychophysical experiments show that dots can be captured by the motion of neighboring dots. This requires an interaction between dots that falls off with distance. We now show how the alow-and-smooth theory can account for this.

Consider a set of dots moving with velocities $\{\vec{v}_i : i = 1, ..., N\}$ at positions $\{\vec{x}_i : i = 1, ..., N\}$. This gives a data term $\sum_{i=1}^{N} |\vec{v}(\vec{x}_i) - \vec{v}_i|^2$. Then add a slow-and-smooth term. We can express this as $\lambda \int d\vec{v}\vec{L}\vec{v}$, where L is a differential operator.

$$\sum_{i=1}^{N} |\vec{v}(\vec{x}_i) - \vec{v}_i|^2 + \lambda \int d\vec{v}\vec{L}\vec{v}. \tag{14}$$

To simply the analysis we will work in one-dimension. This replaces $\vec{v}$ and $\vec{v}_i$ by v and v_i.

The representer theorem states that the solution can be given by a linear combination of the Green's function $G(x)$ of the operator L – i.e. $LG(x) = \delta(x)$. More precisely:

$$v(x) = \sum_{i=1}^{N} \alpha_i G(\vec{x} - \vec{x}_i), \tag{15}$$

where $\vec{\alpha} = (\alpha_1, ..., \alpha_N)$ obey:

$$(I - \lambda G)\vec{\alpha} = \vec{v}. \tag{16}$$

Here I is the $N \times N$ identity matrix, G is the $N \times N$ matrix with components $G(x_i - x_j)$.

The solution depends on the form of the Green function. We can, for example, select the differential operator L so that the Green function G is a Gaussian. In this case the velocity falls off with distance away from the data points. This enables the slow-and-smooth model to capture other motions that are nearby and not ones which are further away.

It can be shown that the Green's function will fall off with distance provided we include a slowness terms $\int d\vec{x}|\vec{v}(\vec{x})|^2$.

References

[1] E. H. Adelson and J. R. Bergen. 1985.

[2] H. Barlow. 1996.

[3] H. B. Barlow and R. W. Levick. 1965.

[4] M. J. Black and S. Roth. On the spatial statistics of optical flow. *International Journal of Computer Vision (IJCV)*, 74(1):33–50, August 2007.

[5] B. Hassenstein and W. Reichardt. 1956.

[6] S. He and R. H. Masland. 1997.

[7] E. C. Hildreth. The computation of the velocity field. *Proceedings of the Royal Society of London, Series B*, 221:189–220, 1984.

[8] C. Koch, V. Torre, and T. Poggio. 1986.

[9] H. Lu, T. Lin, A. Lee, L. Vese, and A. Yuille. Functional form of motion priors in human motion perception. In *Advances in neural information processing systems*, pages 1495–1503, Cambridge, MA, June 2010. Advances in neural information processing systems.

[10] H. Lu, T. Lin, A. Lee, L. Vese, and A. L. Yuille. Functional form of motion priors in human motion perception. In *Advances in Neural Information Processing Systems*, pages 1495–1503, Cambridge, 2010. MIT Press.

[11] S. Roth and M. J. Black. On the Spatial Statistics of Optical Flow. *International Journal of Computer Vision*, 74(1):33–50, Jan. 2007.

[12] V. Santen, J. P. H., and G. Sperling. 1985.

[13] A. A. Stocker and E. P. Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4):578–585, 2006.

[14] S. Ullman. *The interpretation of Visual Motion*. MIT Press, Cambridge, MA, 1979.

[15] Y. Weiss and E. H. Adelson. Slow and smooth: A bayesian theory for the combination of local motion signals in human vision. Technical Report 1624, Massachusetts Institute of Technology, 1998.

[16] Y. Weiss, E. P. Simoncelli, and E. H. Adelson. Motion illusions as optimal percepts. *Nature Neuroscience*, 5:598–604, 2002.

[17] A. Yuille, P. Y. Burgi, and N. M. Grzywacz. Visual motion estimation and prediction: A probabilistic network model for temporal coherence. *Computer Vision, 1998. Sixth International Conference on*, pages 973–978, 1998.

[18] A. L. Yuille and S. Ullman. Rigidity and smoothness of motion: Justifying the smoothness assumption in motion measurement. In S. Ullman and W. Richards, editors, *Image understanding*. 1989.