Active Vision

Active vision was introduced by Y. Aloimonos and co-workers, and by R. Bajcsy. Earlier, non-mathematical, work in the same spirit was done by J.J. Gibson.

Active Vision is based on the variation of controllable parameters of the image formation process.

If the image formation process depends smoothly on these parameters then continuous variation of the parameters result in continuous variation in the image.

Thus, there is often no need to rely on smoothness of scene parameters, such as illumination and object shape, as is the case in the regularization approach.

Complicated non-linear visual processes can sometimes be converted into linear problems, thereby eliminating the multiple solution problem and improving stability.

Application of Active Vision to Shape from Shading

(Aloimonos, Y. and Bandyopadhyay, A., "Active Vision", First International Conference on Computer Vision, London, 1987, pp 35-54)

Assume a Lambertian shading model, although the non-Lambertian case can be handled as well.

Assume perspective projection and assume that the camera can be moved such that its optic axis always remains parallel to the z axis.

We have control over the motion of the camera.

Motion of the camera induces changes in the image intensity measured at each camera pixel.

Scale the distances involved so that the camera focal length = 1.

Denote the location of an image point (on the camera image sensor) as x_i, y_i ($z_i = -1$), and the 3-D coordinates of a point on the object being viewed as (x_o, y_o, z_o).

The origin of the coordinate system is taken to be at the lens centre of the camera in its initial position, with the z axis aligned with the optical axis of the camera.

The location of the camera's lens centre is denoted (x_c, y_c).

The laws of perspective projective gives:

$$x_i - x_c = -\frac{x_o - x_c}{z_o}$$

$$y_i - y_c = -\frac{y_o - y_c}{z_o}$$

As the camera moves (i.e. as (x_c, y_c) varies under our control) the image $I(x, y)$ changes.

There are two ways to consider the image change:

- Eulerian Flow - look at the intensity variation related to a fixed image pixel. As the camera moves this pixel will correspond to a different object point.
- Lagrangian Flow - look at the intensity variation related to a fixed object point. As the camera moves this object point will project to a different image pixel.

Use the Lagrangian approach. Look at isophotes and the epipolar lines in the image.

In a Lambertian surface image intensity of an object point does not depend on the camera position. Thus the isophotes will move with the object points.

Apply some sort of labelling to the different isophotes (α) and epipolar lines (β). For example the intensity value itself could be used for α.

An image point with a given label α, β always corresponds to the same object point.

As the camera moves, we can track a point with a given label α, β to obtain

$$x_i(\alpha, \beta, x_c, y_c), y_i(\alpha, \beta, x_c, y_c)$$

Differentiating the perspective projection equations we get:

$$\frac{\partial x_i}{\partial \alpha} = -\frac{1}{z_o}\frac{\partial x_o}{\partial \alpha} + \frac{(x_o - x_c)}{z_o^2}\frac{\partial z_o}{\partial \alpha}$$

$$\frac{\partial y_i}{\partial \alpha} = -\frac{1}{z_o}\frac{\partial y_o}{\partial \alpha} + \frac{(y_o - y_c)}{z_o^2}\frac{\partial z_o}{\partial \alpha}$$

$$\frac{\partial x_i}{\partial \beta} = -\frac{1}{z_o}\frac{\partial x_o}{\partial \beta} + \frac{(x_o - x_c)}{z_o^2}\frac{\partial z_o}{\partial \beta}$$

$$\frac{\partial y_i}{\partial \beta} = -\frac{1}{z_o}\frac{\partial y_o}{\partial \beta} + \frac{(y_o - y_c)}{z_o^2}\frac{\partial z_o}{\partial \beta}$$

The LHS contains known (measured) values, while the RHS contains the unknowns.

By substituting the perspective equations in the above we can eliminate z_o^2 to give (in vector form):

$$\frac{\partial \vec{x}_i}{\partial \alpha} = -\frac{1}{z_o}\frac{\partial \vec{x}_o}{\partial \alpha} - \frac{(x_i - x_c)}{z_o}\frac{\partial z_o}{\partial \alpha}$$

and a similar equation in β.

As the camera moves the relative α and β values between two object points do not change, but the corresponding $\vec{x}_i$ do change (the distance between the isophotes or epipolar lines for the two points). From this change the object shape can be inferred.

For a short baseline (small camera motion) we can get approximation to derivatives with respect to $\vec{x}_c$:

$$\frac{\partial^2 x_i}{\partial \alpha \partial \vec{x}_c} = -\left(\frac{\partial \vec{x}_i}{\partial \vec{x}_c} - \delta_{ij}\right)\frac{1}{z_o}\frac{\partial z_o}{\partial \alpha}$$

We can then multiply the previous equation as well as the perspective equation and the first derivative equations by z_o and get a $Linear$ system of equations:

$$(\vec{x}_i - \vec{x}_c)z_o + \vec{x}_o - \vec{x}_c = 0$$

$$\frac{\partial \vec{x}_i}{\partial \alpha} z_o + \frac{\partial \vec{x}_o}{\partial \alpha} + (\vec{x}_i - \vec{x}_c)\frac{\partial z_o}{\partial \alpha} = 0$$

$$\frac{\partial^2 x_i}{\partial \alpha \partial \vec{x}_c} z_o + \left(\frac{\partial \vec{x}_i}{\partial \vec{x}_c} - \delta_{ij}\right)\frac{\partial z_o}{\partial \alpha} = 0$$

We can then solve this system of six equations for the six unknowns $\vec{x}_o, z_o, \partial \vec{x}_o / \partial \alpha, \partial z_o / \partial \alpha$

We need not get the measurements all from a single image. We can take many measurements along the trajectory of a given point α, β and average them. This reduces the effect of measurement noise.

Doing the above process for the equations involving both α and β will yield the object point coordinates $\vec{x}_o$ and two tangent vectors on the object ($\vec{t}_\alpha = \partial\vec{x}_o/\partial\alpha$ and $\vec{t}_\beta = \partial\vec{x}_o/\partial\beta$)

From these tangent vectors we can compute the normal to the object surface via:

$$\hat{n} = \frac{1}{\sqrt{\vec{t}_\alpha \cdot \vec{t}_\beta}}(\vec{t}_\alpha \times \vec{t}_\beta)$$