Course notes, CS664, 9/23/04

- Administrivia: short in-class quiz at the end of next lecture (Tuesday). PS1 will be out soon (implement energy minimization algorithms for stereo).

- Addendum: additional important camera internal parameter (besides focal length and sampling grid): principal point, aka center of the image, which is the intersection of the optical axis and the imaging plane.

TRACKING

- Contour-based tracking with deformable models (snakes): consider a curve on the image that is attracted to, e.g., edges. (More precisely, changes in intensity)

- Imagine a spring on a hillside, rolling downhill to a stop. The hillside is a potential function defined over the image. Typically based on the gradient (attract to large changes in intensity, i.e. edges).

- Again, we have

$$E_{data} + \lambda \cdot E_{smooth}$$

  We can represent the curve as a spline (set of control points). Called "snakes" or "active contour models", and a big hit of vision.

- As we will see there are fast methods to minimize this energy function. But it's not clear that you want the global minimum, especially when using them for tracking.

- Main advantage of "physics-based vision" is that the algorithms are somewhat intuitive (this is often true in energy minimization, as long as there is a good optimization algorithm... otherwise, you can't distinguish optimization algorithm failure from energy function failure). Parameters can control, for example, spring stiffness.

- Problems with contour tracking: (1) serious occlusions; consider ambiguity when passing behind an object (2) splitting/merging is not handled by snakes (3) problems with high texture, as usual

- Snakes can be generalized to 3D (for instance, in medical data),

often called "balloon models".

RECOGNITION (CLASSIC)

- Recognition problem definition: position and pose from model library (usually)

- Simple example, not usually considered vision: OCR

- Face recognition vs detection.

- Models are just images; this is sometimes called an *appearance* model, as opposed to most models in recognition which are geometric.

- Face recognition is actually a very solvable problem under the right assumptions (mugshot database, frontal views, similar lighting).

- You can understand most of its limitations by realizing that the most effective methods use the L2 distance. There are more complex models, that try to find eyes, nose, lips — these tend not to work, however.

- 2D vs 3D (images,models)

- Model representation (geometric): usually polygonal, somewhat based on computer graphics and on CAD. Curved object representations exist, but are intractable (though they lead to beautiful complex mathematics).

- Basic idea is that a model in a given pose and position will result in a set of edges.

- Choices of features

- Models are usually rigid. Not much work on model acquisition.

- Transformation groups range from simple to complex. Simple is fast but doesn't model various real image effects.

- Articulated objects