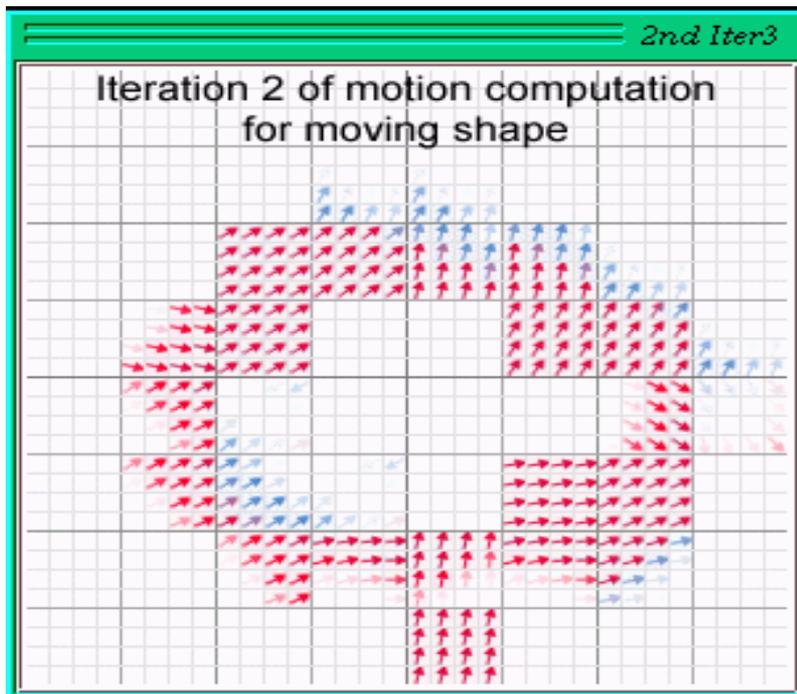


Learning low-level vision



We have developed a machine learning based method which applies to various problems in low-level vision. We seek the scene interpretation that best explains image data. For example, we may want to estimate the projected velocities (scene) which best explain two consecutive image frames (image).

We use computer graphics to generate synthetic data, and model the statistical relationship between images and scenes in the synthetic world. We then use that model to estimate scene corresponding to a given image.

This yields an efficient method to form low-level scene interpretations, which should apply to a variety of low-level vision problems. We have demonstrated the technique for motion analysis and estimating high resolution images from low-resolution ones.

Background and objectives: We want to get a computer to solve vision problems which are trivial for people: interpret a line drawing; estimate the 3-d shape of an object depicted in a photograph; estimate depth from a stereo pair of image; estimate motion from an image sequence. Of course, algorithms exist for many of these problems, but many are brittle. We seek to exploit the memory capacity of modern computers for solving these problems. We developed a common machine learning framework which applies to all these problems. We hope that learning-based, memory-intensive approach will be more reliable than other algorithms. There may be many applications of this technology. This research might ultimately lead to a vision chip, which could input image data and output a mid-level scene representation, such as 3-d shape, or reflectances. It might lead to a method to estimate high resolution images from low-resolution ones.

Technical discussion: We ask: can a visual system correctly interpret a visual scene if it models (1) the probability that any local scene patch generated the local image, and (2) the probability that any local scene is the neighbor to any other? The first probabilities allow making scene estimates from local image data, and the second allow these local estimates to propagate.

First, we synthetically generate images and their underlying scene representations, using computer graphics. For example, for the motion estimation problem, our training images were moving, irregularly shaped blobs.

Second, we place the image and scene data in a Markov network. We break the images and scenes into localized patches where image patches connect with underlying scene patches; scene patches also connect with neighboring scene patches. The neighbor relationship can be with regard to position, scale, orientation, etc.

Third, we propagate probabilities in the Markov network, taking advantage of a "factorization approximation", where we ignore the effects of network loops. This method is fast, and in practise for the problems we have studied, proves to be reliable, as well.

TR2000-05

Learning low-level vision

(longer journal version, Intl. Journal of Computer Vision, 40(1), pp. 25-47, 2000) William T. Freeman, Egon C. Pasztor, and Owen T. Carmichael

TR99-12

Learning low-level vision

(conference version, Intl. Conf. on Computer Vision, Corfu, Greece, 1999) William T. Freeman, Egon C. Pasztor

TR99-08

Markov networks for low-level vision

William T. Freeman, Egon C. Pasztor

TR99-05

Learning to estimate scenes from images

William T. Freeman, Egon C. Pasztor,
Neural Information Processing Systems 11, 1998.