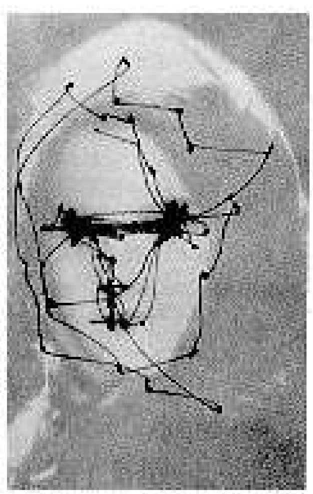
Examination #2

Handed Out: 11am, December 2, 2004
Due: 5pm, December 3, 2004

This take-home exam is to be done individually, though you are free to use any resources such as books, papers, and web pages, as source material for helping you develop your own answers. If you use someone else's work, be sure to cite it. Your answers should be either printed neatly or type written. Please write clearly and concisely; longwinded answers that go off point and discuss extraneous issues will have marks taken off. Use figures and mathematics to aid in giving concise answers. Overall, these questions are intended to contain some open-ended aspects that will be best answered by giving a clear description of the key theoretical and algorithmic issues that are involved and then giving a brief description of a computational approach to a solution. Turn in your answers to Dyer's office (6379 CS) or mailbox (on the 5^{th} floor of CS). There is no length restriction, but I would expect that about 5-7 typed pages would be plenty.

1. Explain why a portrait's eyes appear to "follow you around the room." Give your answer in terms of a homography relationship between the viewer and the picture.

2. The human visual system has photoreceptors called rods and cones that vary in spatial density with their distance from the direction of gaze. In particular, the *fovea* is the central region where the spatial density is highest, but it only covers a visual angle of approximately 2 degrees, which is about the width of your thumb when viewed at arm's length (see right image below)! (Furthermore, the central foveal region only contains cones, which are the color vision receptors.) To avoid seeing the world with "tunnel vision" the human eye performs about 3 saccades per second, which are small eye movements, so that the fovea can view different parts of the scene with high spatial acuity (i.e., detail). Below is an illustration of the sequence of saccades (center) of a person as they looked at a picture of a face (left).


As a result of saccadic eye movements and other visual processing, our perception is of a seemingly high-resolution view of the world. Yet how people integrate these multiple, overlapping, spatially-varying resolution images is not understood. (For more information on this see, for example, http://eyelab.msu.edu/VisualCognition and http://www.psychology.uiowa.edu/faculty/hollingworth/) This problem is also of concern with digital cameras because, while they have increasingly higher spatial resolution (i.e., number of pixels), it is unlikely to be feasible to build cameras that have both very large fields of view and very high pixel density.

For this problem assume we have two digital cameras, one with a small focal length (and hence a wide field of view (WFOV) and a coarse pixel sampling of the visual field) and one with a large focal length (and hence a narrow field of view (NFOV) and a fine sampling of the visual field). Call these two cameras the WFOV and NFOV cameras, respectively, and assume they are rigidly attached together. Using this camera rig mounted on a tripod, we can rotate it and take images from each of the two cameras at any orientation.

a) Given a set of *n* pairs of grayscale images, describe a framework for combining all of these images into a single, variable-resolution image representation. Do *not* assume that you know the relative orientation between images. In other words, describe the theoretical and algorithmic aspects of constructing a mosaic image from these *n* pairs of images.

b) Describe an "active vision" approach to using this rig for actively controlling the rig's orientation in order to recognize a moving 2D object. Assume the rig can capture 3 image pairs per second. The rig must be controlled so that the moving object is "followed" by the camera, while at the same time using the NFOV camera to obtain enough detail to recognize the object. Each NFOV image can only view a small part of the object to be recognized. Given a library of 2D objects to be recognized, one or more NFOV images of different regions on an object may need to be analyzed before recognition can be achieved.