Project 1: Face Recognition

A. Task. Given a set of face images, determine (recognize) which one corresponds to a particular target face.

B. Importance. Face recognition is an important application of *image/object recognition.* It requires both low and high-level visual information processing. Target applications include security and banking identification systems.

C. Problem Description. Face images correspond to the projection into the image plane of 3-D (human) heads. Canonical views of the head correspond to snapshots spaced at, say, 30°; if we use a spherical coordinate system centered at the "center" of the head, and fix the polar angle θ at 90°, then the snapshots are described by varying the azymuth ψ at regular (30°) angles, for a fixed radius R. The most common face image is that corresponding to the frontal view $\psi = 0^\circ$.

The standard approach to face recognition is *model-based.* Typically: (i) we identify face features (landmarks), such as the eyes, the nose, and the mouth; (ii) associate the features to (geometrical) models; (iii) determine a face model by "linking" the feature models; (iv) match the face model to the faces in the set.

D. Past Work. Among others, the work of Kanade [1], Turk-Pentland [2], and Yuille [3] developed important concepts for the face recognition problem. Kanade developed a top-down system using a statistical approach to classify and match facial parameters. Turk and Pentland proposed the idea of "eigenfaces". They use an information theoretical approach to face recognition by decomposing a face image into a set of 'eigenface" images; the relative strength of this combination of eigenfaces determines particular (recognized) faces. Yuille proposed the use of deformable templates for face recognition. He models face landmarks in terms of parametrized geometrical models.

References

[1] Kanade, T.
Computer recognition of human faces. Birkhauser Verlag, Basel and Stuttgart, 1977.

[2] Turk, M. and Pentland, A.
Eigenfaces for recognition. Journal of Cognitive Neuroscience, **3**, 1, 71-86, 1991.

[3] Yuille, A.
Deformable templates for face recognition. Journal of Cognitive Neuroscience, **3**, 1, 59-70, 1991.

Project 2: Facial Expression Processing

A. Task. Given a series of face snapshots, analyze, interpret, and synthesize facial expressions. This includes "emotion processing" and lip reading.

B. Importance. It is part of a broad area of research which studies visual behavior as reproduced by artificial systems. Applications include video teleconferencing and systems which combine auditory and visual information.

C. Problem Description. For an arbitrary face image (frontal view) sequence, we want to infer semantic or physical attributes. The first stage corresponds to *face analysis.* This describes the process of decomposing facial characteristics, such as face landmarks and their mutual relations, into models. The second stage corresponds to *face interpretation.* This is guided by the application. For example, if we want to process "emotional states" of a person, we have to compare the model(s) obtained at the analysis stage, to a models describing facial emotion, such as smile (happy), cry (sad), or grin (disgust). The third stage is to synthesize the various emotional states in terms of computer graphics models.

D. Past Work. Among the many works in this area, two are worth mentioning: the Essa [1] and the Waters [2] works. Essa developed a framework for facial emotion processing which couples the estimation of the optical flow with geometrical and physical face models. Taken together, these models and their temporal variance describe "action units" which are associated to particular "emotional states". Waters developed in his pioneer work a framework which couples detailed modeling of facial muscles with computer graphics models.

References

[1] Essa, I. A.
Analysis, interpretation, and synthesis of facial expressions. Technical Report #303, Massachusetts Institute of Technology, Media Laboratory, 1994.

[2] Waters, K. and Terzopoulos, D.
The computer synthesis of expressive faces. Phil. Trans. Royal Society of London B, 1992.

Project 3: Image and Video Watermarking

A. Task. Inserting/removing digital watermarks in images or video sequences.

B. Importance. Is an interdisciplinary research topic. It combines signal and image (video) processing techniques, with coding and information theories. Areas of application include security systems on the the internet and copyright of digital products.

C. Problem Description. Digital watermarks have the goal of "marking" an image or a video sequence with sign which identifies its bearer with the author or owner of the image or video sequence. A watermark should be perceptually invisible and difficult to be removed by parties other than the author/owner of the image.

D. Past Work. Work on digital watermarks has received increased attention in the last five years. An interesting work has been recently published by I.J. Cox, J. Kilian, T. Leighton, and T. Shamoon [1]. They developed an universal watermark technique for multimedia applications. They adapt spread spectrum techniques for inserting watermarks into the spectral components of the data.

References

[1] Cox, I., Kilian, J., Leighton, T., and Shamoon, T. *Secure spread spectrum watermarking for multimedia.* Technical Report, NEC Research Institute, #95 – 10, 1995.

Project 4: Segmentation and Tracking in Dynamic Scenes

A. Task. Given a video sequence of a dynamic scene we want to segment the objects (persons, cars, etc.) relative to the static background - the figure/ground segmentation problem. Also, once segmented, we want to track specific objects which are pre-determined by the application being implemented.

B. Importance. Segmentation is among the most important problems in visual information processing. Motion segmentation is specially important for video applications. Automatic object tracking, which is restricted in this project to 2-D region tracking, is fundamental in applications such as target tracking and sports broadcasting.

C. Project Description. This project deals with the particular application of NFL video. The goal is to segment and track a player and/or the ball. This requires a multistage process. First, we have to determine the motion induced in the image by camera motion/zooming - the image background velocity. Second, we subtract from the images the background velocity and segment the objects. Third, we determine the velocity of the segmented objects. Fourth, we track the objects. The challenge here is to come up with an efficient method to perform these tasks close to or in real-time.

D. Past Work. See www.

Project 5: Temporal Segmentation of Video

A. Task. Given a video sequence, determine in which frames scene changes occur.

B. Importance. Temporal segmentation is a major task in video production and content-based handling. It may require low and high-level visual information processing, depending on how much is known about the structure of the "world" and the problem being addressed. Target applications include assisted non-linear video editing systems, and digital video library indexing.

C. Problem Description. Video is a temporal sequence of frames. Prior knowledge about the world and problem domain can be used to determine when interesting events, such as scene changes occur. These selected frames contain a substantial amount of information, and for some applications the extraction of high level content information from those frames is important.

In this project we focus on segmentation of a NFL game into plays. Video is shot according to a specific rule that says that there is always an scoreboard shot in between plays. We want to segment the video sequence into plays by detecting the scoreboard shots to allow the extraction of textual information present in these frames. Segmentation is a fundamental step towards for autonomous game annotation.

Past Work. See www.

Project 6: Content Extraction from Video.

A. Task. Given a video sequence segmented into small clips (see project 5), extract textual content from scene change frames.

B. Importance. OCR (Optical Character Recognition) is a document processing technique that allows textual information captured in image form to be recognized and stored in more compact and efficient formats, i.e. ASCII format. In off-the-shelf OCR systems, as those that come bundled with scanners, distortions introduced by perspective, uncalibrated cameras and uneven illumination are not supported. Non-standard typesets are also a no-no. Applications as automatic video annotation for digital video library indexing require OCR functionality in a much less restricted conditions.

C. Problem Description. In a video sequence, some of the frames contain a substantial amount of information in textual form, and for some applications the extraction of this kind of high level information is important.

In this project we focus on the analysis of scoreboard shots and player jersey shots from several NFL videos segmented in time. The textual information present in these scenes is fundamental for game annotation and posterior game analysis during coaching time.

D. Past Work. See www.

Project 7: Object Recognition via Geometric Invariance

A. Task. Given a sequence of images from a 2-D man made object, determine which object among a set of known targets is present in the scene.

B. Importance. Object recognition is a major research topic. The specific case of a planar 2-D object immerse in a 3-D world is of interest because most man made objects have planar surfaces and the more complex problem of recognizing 3-D objects may be decomposed in smaller problems.

Geometric Invariance is a powerful tool because it provides features that are invariant to camera and object motion. Some applications are automatic target recognition systems, autonomous robot vision for manufacturing, navigation and exploration.

C. Problem Description. Recognition of 2-D objects using a sequence of shots of the object in an arbitrary and unknown position and orientation in 3-D, is not a trivial task especially under noisy conditions. This project entails image database construction, image pre-processing, computation of geometric invariants, and finally GI based-classifier performance analysis.

D. Past Work. See www.