EE E6887 Statistical Pattern Recognition
Course Project

Due Date: December 12th 2005

Topic: Image classification

Data Set:

The data set includes 3504 images from the broadcast news domain. This set has been extracted from the TRECVID 2005 development set. Manually labeled ground truths of two concepts "*building*" and "*explosion_fire*" for each image are included. There are 555 images labeled as *building* and 136 images labeled as *explosion_fire* in this data set. For each image, two image features, grid color moments and Gabor texture, are provided.

Task:

The objective of this project is for you to gain experience in designing statistical classifiers for solving practical problems by applying the knowledge and tools learned in this course. You are asked to develop image concept classifiers, which take as input the image features and predict the concept labels. You are allowed to use any tool learned in this class as well as promising ones found in the related literature. For example, methods like SVM, GMM Bayesian classifier, or boosting have been frequently used in practice. You are encouraged to investigate advanced topics as well. For example, you may explore fusion of classifiers using different features, models, or families of classifiers. In addition, you my study the techniques of feature correlation analysis and feature selection and analyze their effects on classification performance.

In evaluating the results of your classifiers, please compute the precision and recall values of your classifier. If your classifier includes any parameters that can be adjusted, please report the precision-recall curve showing the performances at different parameter values. Remember to partition the data set into K folds and conduct a cross-validation procedure in measuring the performance. For simplicity, keep K small. A two-fold partition will be fine.

Requirements:

- A project report of 10 pages, including text description, analysis, and all the figures. Remember to include explanations about why you choose the specific approaches, your empirical design decisions, and insights you learn in intermediate analysis steps as well as the final results. Please provide sufficient details that are needed by future readers in case they want to repeat your experiments.
- A power-point file summarizing your approaches, experiments, and analysis results

Sample images from the data set:

Building:

Explosion_fire:

Others: