

Image Database Retrieval

9807-NTT03

Progress Report: July 1, 1999 December 31, 1999

Paul Viola

Project Overview

In this research project we plan to study and create systems that can scan images and video to locate items of interest. For example, such a system should be able to scan a travel documentary for images of distinct locations and objects, like “Buddhist temples”, “gothic cathedrals”, or “statues on horseback”. We believe that by leveraging our existing work in this area, we can play a key role in setting the standard for research in visual information retrieval. At the same time, this research provides an excellent opportunity for transition to practical applications. We believe that the complementary skills of MIT and NTT are well suited for pursuing this dual path of developing practical applications of image indexing in conjunction with fundamental progress in associated science and engineering.

Progress Through December 1999

We have made progress on several problems related to the core goals of the Image Database retrieval problems:

- **Detection of Faces in Images:** In the recent past there has been rapid progress on the detection of faces from a frontal viewpoint, using templates or related approaches. There has been much less progress on the detection of faces from a more general viewpoint, in portrait or 3/4 view. We have developed a new approach that attempts to model the appearance of faces as a statistical distribution over features. This approach parallels our recent work on the detection of visual texture. We call this approach **Objects as Textures** .

We believe that this approach is actually quite general and will allow us to detect more complex patterns, such as the appearance of the human body. This problem is very hard because of the variety of poses that the human body can assume. Faces and people are a critical aspect of image databases.

- **New Statistical Model of Image Database Features:** Most classical approaches for image database indexing rely in a very small number of generic features (color, vertical edges, horizontal edges, etc.). Unfortunately, by themselves these features are not very selective (e.g. almost every image contains some vertical edges). In order for queries to be selective in such systems, the user must carefully specify the values for these features (e.g. 9% red pixels and 13% vertical edges). These precise measurements are very dependent on the image background and object scale — which adversely effects performance. We argue that the features used to represent images should be unique and selective. In such a system each feature is present in only a small percentage of images.

We have created a mechanism of computing features of this type called **Complex Features** . A retrieval system based on this insight works much better than previous systems. The algorithms for constructed this feature set, and the query system itself, is computationally efficient.

- **Recognition of Scanned Documents:** The approaches above are optimized for retrieval in photo databases. Of related interest is searching databases of scanned documents. Currently documents can be read using OCR, but there is no attempt to interpret the mathematical expressions in engineering or technical documents.

We have constructed a system that can automatically interpret mathematical expressions in such documents. This provides a new mechanism for searching technical documents. Based on the same ideas we have built an interactive handwritten mathematical expression recognizer. The system provides a friendly and intuitive interface for the entry of mathematical expressions.

New Results since July 1999

- We have developed a new technique for learning image database queries that is based on the boosting algorithm from machine learning. This algorithm allows us to learn queries from a small number of images. In addition the algorithm is specifically tailored to learn a query that is very efficient for retrieval. The key idea is that the query returned uses the smallest number of features that can distinguish the negative from positive examples. In many cases the query will utilize only 10 or 20 of the 45,000 possible features. As a result query retrieval is very fast — we could potentially search a database containing millions of images in less than a second. These results are described in a pre-print paper available from the MIT/NTT web page: <http://www.ai.mit.edu/projects/ntt>.

Research Plan for the Next Six Months

- Incorporate new types of media, such as video and virtual reality.
- Further refine the query learning process..