Using Generative probabilistic models for multimedia retrieval

Thijs Westerveld
CWI/INS1
PO Box 94079, 1090 GB Amsterdam, The Netherlands
`thijs@cwi.nl`

This thesis discusses information retrieval from multimedia archives, focusing on documents containing visual material. We investigate search and retrieval in collections of images and video, where video is defined as a sequence of still images. No assumptions are made with respect to the content of the documents; we concentrate on retrieval from generic, heterogeneous multimedia collections. In this research area a user's query typically consists of one or more example images and the implicit request is: "Find images similar to this one." In addition the query may contain a textual description of the information need. The research presented here addresses three issues within this area.

First, we show how generative probabilistic models can be applied to multimedia retrieval. For each document in the collection a probabilistic model is built. For each of these models we then compute the probability that the query is generated from the model and the documents corresponding to the models with the highest probability are shown to the user. The assumption is that these are the most relevant documents, i.e., those with characteristics corresponding to the query characteristics. Visual information is modelled using Gaussian mixture models and information derived from language (e.g., the speech of the video soundtrack) is modelled using statistical language models.

The second issue addressed is the parallel between the use of generative probabilistic models for multimedia retrieval and comparable models for text. This thesis describes how the techniques developed for language relate to the multimedia techniques presented here and how these parallels can be leveraged.

Third, this thesis studies evaluation. We tested different model variants using a number of collections including the test collections of TRECVID, the international workshop series for benchmarking video retrieval. On average, language-based approaches outperform approaches based on visual information. However, for some queries visual information is important. A combination of both modalities gives the best results when searching a heterogeneous multimedia collection.

Dissertation available from `http://purl.org/utwente/41716` or `http://homepages.cwi.nl/~thijs/pub/phd.html`.