The LIG Multi-Criteria System for Video Retrieval

Stéphane Ayache
Laboratoire d’InfoRmatique en Images et Systèmes d’information
5, Avenue Pierre Mendès-France
69676 Bron Cedex, France
Stephane.Ayache@imag.fr

Georges Quénot
Laboratoire d’Informatique de Grenoble
CNRS
BP 53
38041 Grenoble Cedex 9, France
Georges.Quenot@imag.fr

ABSTRACT
The LIG search system uses a user-controlled combination of five criteria: keywords, similarity to example images, semantic categories, similarity to already identified positive images, and temporal closeness to already identified positive images.

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

Keywords: Video retrieval, multi-criteria search.

The LIG search system [1] uses a user-controlled combination of five criteria: keywords, similarity to example images, semantic categories, similarity to already identified positive images, and temporal closeness to already identified positive images (Figure 1). The system outputs an ordered list of relevant shots for each topic after interaction with the user (initial query and multiple relevance feedback).

Figure 1: The LIG video retrieval system

Criterion 1: Keyword based search. The keyword based search is done using a vector space model. The words present in the ASR-MT transcription are used as vector space dimensions. Stemming ans stopword list are used. Relevance is first assigned to speech segments and then projected onto overlapping shots.

Criterion 2: Similarity to image examples. Visual similarity between key frames and image examples is looked for using color and texture characteristics (4x4x4 color histograms and 8x5 Gabor transforms). Distances are computed, normalized and the turned into a relevance value for each feature. A 65% color and 35% texture linear combination is used.

Criterion 3: Feature based search. The goal of this part is to help focusing on specific categories of the video shots, according to a non-crisp labeling of their keyframes. All keyframes are automatically labeled according to the 36 categories used in the TRECVID high level feature detection task.

Criterion 4: Visual similarity to already identified positive images. Visual similarity to already retrieved images can be used for the search. These images have to be marked as positive examples for similarity based search by the user (relevance feedback). The search is performed in the same way as for the original image examples. Key frames are ranked according to their visual closeness to the positive examples. The images selected for similarity-based search need not to be actually positives for the current search.

Criterion 5: Temporal closeness to already identified positive images. Temporal closeness (within the video stream) to already retrieved images can be used for the search. These images have to be marked as positive examples for temporal closeness based search by the user (relevance feedback). Key frames are ranked according to their temporal closeness to the positive examples. The images selected for temporal closeness-based search need not to be actually positives for the current search.

Combination of search criteria. The user can dynamically define his search strategy according to the topic and/or the looking of the retrieved images. Each search criterion can be independently configured and given a global weight for the search. Relevance are computed independently for each criterion and for each key frame (or subshot). The per-criterion relevances are then linearly combined according to the criteria weights to produce the final key frame relevance. A relevance is computed for each shot as the maximum of the relevances associated to each key frame (or subshot).

1. REFERENCES