Introduction

- Huge number of multimedia contents, like videos;
- YouTube receives around 300 hours of video per minute;
- Video retrieval is a challenging task;
- A single video may contain multiple categories (genres).

Literature

- There are two approaches to represent video content:
  1) Spatio-temporal methods [1];
  2) Spatial-only methods [2];
- A popular approach for spatial methods is visual dictionaries.

Experimental Protocol

VIDEO GENRE RETRIEVAL
- Genre Tagging Task at MediaEval 2012 [3]:
  - 14,838 videos (3,288 hours) collected from Blip.tv;
  - 5,288 videos (3,943,375 frames) for training (36%);
  - 9,550 videos (7,273,996 frames) for testing (74%);
  - 26 different genres assigned by Blip.tv;
  - 7 content descriptors of separated frames [4];
  - Balanced training with 800 frames of each genre;
  - 200 queries (5% of the dataset size) replicated 5 times;
  - Retrieval effectiveness P10 and MAP.

VIDEO EVENT RETRIEVAL
- Event Video (EVVE) dataset [5]:
  - 2,995 videos (166 hours) collected from YouTube;
  - 13 events (categories);
  - 620 (20%) of query videos;
  - 2,375 (80%) of reference videos;
  - Retrieval effectiveness MAP.

Bag of Genres

- New: visual words are labeled regions determined by a classification model;
- Contribution: each dimension of the feature space spanned by the model is associated to a semantic concept.

Experimental Results – Genre Retrieval

- Approach performs similar to state-of-the-art methods on MediaEval dataset and it is the state-of-the-art on EVVE dataset;
- In the proposed visual dictionary, visual words are obtained by a supervised classifier; the method is compact: each dimension corresponds to a semantic concept;
- Future work includes the evaluation of other methods for feature extraction and classification strategies.

Experimental Results – Event Retrieval

- Performance of different methods for event retrieval on EVVE dataset. BoG+BIC outperformed all the baselines by a large margin.

Bibliography