Chapter 12
Constructing and Utilizing Video Ontology for Accurate and Fast Retrieval

Kimiaki Shirahama
Kobe University, Japan

Kuniaki Uehara
Kobe University, Japan

ABSTRACT
This paper examines video retrieval based on Query-By-Example (QBE) approach, where shots relevant to a query are retrieved from large-scale video data based on their similarity to example shots. This involves two crucial problems: The first is that similarity in features does not necessarily imply similarity in semantic content. The second problem is an expensive computational cost to compute the similarity of a huge number of shots to example shots. The authors have developed a method that can filter a large number of shots irrelevant to a query, based on a video ontology that is knowledge base about concepts displayed in a shot. The method utilizes various concept relationships (e.g., generalization/specialization, sibling, part-of, and co-occurrence) defined in the video ontology. In addition, although the video ontology assumes that shots are accurately annotated with concepts, accurate annotation is difficult due to the diversity of forms and appearances of the concepts. Dempster-Shafer theory is used to account the uncertainty in determining the relevance of a shot based on inaccurate annotation of this shot. Experimental results on TRECVID 2009 video data validate the effectiveness of the method.

DOI: 10.4018/978-1-4666-2940-0.ch012

There is significant demand for a video retrieval method that is capable of retrieving shots of interest from a large number of videos efficiently. Based on how to represent a query, existing methods can be classified as Query-By-Keyword (QBK) or Query-By-Example (QBE). In the QBK approach, the user provides keywords that represent the query. The system then retrieves shots that are annotated with the same or related keywords. Under the QBE approach, the user provides ex-
ample shots to represent the query. The system then retrieves shots that are similar to example shots in terms of features like color, edge, motion, etc. QBE has two advantages over QBK. The first advantage is that through features in example shots, the query is represented with no subjectivity. In QBK, representing semantic content with keywords is affected by user subjectivity and lexical ambiguity, as different users may perceive the same content differently (Rui, Huang, & Chang, 1999). The second advantage is that QBE requires no shot annotation as features can be automatically extracted from shots. In contrast, QBK requires shot annotation, but it is impractical to annotate a huge number of shots with keywords for all possible queries. Therefore, the video retrieval method we developed is based on the QBE approach.

However, applying QBE to large-scale video data involves two crucial problems: The first one is that similarity in features does not necessarily imply similarity in semantic content. In other words, QBE retrieves several irrelevant shots that contain similar features to example shots, but show different semantic content. One main reason is overfitting, resulting from the insufficiency of example shots, compared to the high-dimensionality of features. A user can only provide a small number of example shots for a query. However, each shot is represented using high-dimensional features, for example, a bag-of-visual-words representation having over 1,000 dimensions. Due to the small number of example shots available, obtaining reliable statistical information about feature dimensions is difficult. Consequently, retrieved shots are similar to example shots in terms of feature dimensions that are very specific to the example shots, but are not useful for characterizing the query. For example, in Figure 1, when Ex. 1, 2, and 3 are provided as example shots for the query “tall buildings are shown”, Shot 1, 2, and 3 are incorrectly retrieved. This is because Shot 1, 2, and 3 are similar to Ex. 1, 2, and 3 in terms of feature dimensions that characterize few edges in sky regions. The second problem is an expensive computational cost to calculate the similarity of a huge number of shots to example shots. The computational cost of QBE linearly increases depending on the number of shots.

To overcome the above problems, as preprocessing of QBE, we develop a method that filters a large number of irrelevant shots to a query based on concept detection results. Figure 1 shows detection results for three concepts: Building, Cityspace, and Person. Each shot is represented as a vector of detection scores, each of which represents the presence of a concept in the shot. A higher detection score for a concept indicates that this concept is more likely to appear in a shot. In Figure 1, Building and Cityspace are likely to appear in Ex. 1, 2, and 3, although unlikely to appear in Shot 1, 2, and 3. Several researchers have used concept detection results in video retrieval recently (Ngo et al., 2009; Snoek et al., 2009). Unlike a small number of example shots in QBE, the detector for each concept is built using a large number of shots (more than 10,000 shots), that are annotated to represent the

Figure 1. Example of an overfit retrieval result

<table>
<thead>
<tr>
<th>Query: Tall buildings are shown</th>
<th>Retrieved by overfitting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ex. 1</td>
</tr>
<tr>
<td>Building:</td>
<td>0.9</td>
</tr>
<tr>
<td>Cityspace:</td>
<td>0.7</td>
</tr>
<tr>
<td>Person:</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Related Content

Hybrid Query Refinement: A Strategy for a Distance Based Index Structure to Refine Multimedia Queries
Kasturi Chatterjee and Shu-Ching Chen (2011). *International Journal of Multimedia Data Engineering and Management* (pp. 52-71).
[www.igi-global.com/article/hybrid-query-refinement/58051?camid=4v1a](www.igi-global.com/article/hybrid-query-refinement/58051?camid=4v1a)

Network Deployment for Social Benefits in Developing Countries
[www.igi-global.com/chapter/network-deployment-social-benefits-developing/17515?camid=4v1a](www.igi-global.com/chapter/network-deployment-social-benefits-developing/17515?camid=4v1a)

Perceptual-Based Visualization of Auditory Information Using Visual Texture
[www.igi-global.com/chapter/perceptual-based-visualization-auditory-information/8426?camid=4v1a](www.igi-global.com/chapter/perceptual-based-visualization-auditory-information/8426?camid=4v1a)

IP Mobility Support in Hybrid Wired-Mobile Ad Hoc Networks
[www.igi-global.com/chapter/mobility-support-hybrid-wired-mobile/50329?camid=4v1a](www.igi-global.com/chapter/mobility-support-hybrid-wired-mobile/50329?camid=4v1a)