ABSTRACT

This chapter introduces the application of active learning in video annotation. The insufficiency of training data is a major obstacle in learning-based video annotation. Active learning is a promising approach to dealing with this difficulty. It iteratively annotates a selected set of most informative samples, such that the obtained training set is more effective than that gathered randomly. The authors present a brief review of the typical active learning approaches. They categorize the sample selection strategies in these methods into five criteria, that is, risk reduction, uncertainty, positivity, density, and diversity. In particular, they introduce the Support Vector Machine (SVM)-based active learning scheme which has been widely applied. Afterwards, they analyze the deficiency of the existing active learning methods for video annotation, that is, in most of these methods the to-be-annotated concepts are treated equally without preference and only one modality is applied. To address these two issues, the authors introduce a multi-concept multi-modality active learning scheme. This scheme is able to better explore human labeling effort by considering both the learnabilities of different concepts and the potential of different modalities.
**INTRODUCTION**

With rapid advances in storage devices, networks and compression techniques, large-scale video data is becoming available to more and more average users. The management of these data becomes a challenging task. To deal with this issue, it has been a common theme to develop techniques for deriving metadata from videos to describe their content at syntactic and semantic levels. With the help of these metadata, the manipulations of video data can be easily accomplished, such as delivery, summarization, and retrieval.

Video annotation is an elementary step to obtain these metadata. Ideally, video annotation is formulated as a classification task and it can be accomplished by machine learning methods. However, due to the large gap between low-level features and to-be-annotated semantic concepts, typically learning methods need a large labeled training set to guarantee reasonable annotation accuracy. But because of the high labor costs of manual annotation (experiments prove that typically annotating 1 hour of video with 100 concepts can take anywhere between 8 to 15 hours (Lin, Tseng, & Smith, 2003)), this requirement is usually difficult to meet.

Active learning has proved effective in dealing with this issue (Ayache & Quénot, 2007; Chen & Hauptmann, 2005; Cohen, Ghahramani, & Jordan, 1996; Panda, Goh, & Chang, 2006). It works in an iterative way. In each round, the most informative samples are selected and then annotated manually, such that the obtained training set is more effective than that gathered randomly. In this chapter, we discuss the application of active learning in video annotation, which mainly consists of two parts. In the first part, we provide a survey of the existing active learning approaches, especially their applications in image/video annotation and retrieval. We analyze the sample selection strategies in these methods and categorize them into five criteria: risk reduction, uncertainty, positivity, density, and diversity. Afterwards, we detail a widely-applied Support Vector Machine (SVM)-based active learning approach. In the second part, we analyze the deficiency of the existing active learning methods in video annotation, including: (1) the to-be-annotated concepts are treated equally without preference and (2) only a single modality is applied. To address these two issues, we introduce a multi-concept multi-modality active learning scheme, in which multiple concepts and multiple modalities can be simultaneously taken into consideration such that human effort can be more sufficiently exploited.

The organization of the rest of this chapter is as follows. In Section II, we briefly review the related works, including video annotation and active learning. In Section III, we provide a survey of the sample selection strategies for active learning. In Section IV, we introduce the SVM-based active learning scheme. In Section V, we discuss the deficiency of the existing active learning approaches for video annotation, and then we present the multi-concept multi-modality active learning scheme in Section VI. Finally, we conclude the paper in Section VII.

**RELATED WORKS**

**Video Annotation**

Video annotation (also referred to as “video concept detection”, “video semantic annotation” (Naphade & Smith, 2004a), or “high-level feature extraction” (TRECVID, http://www-nlpir.nist.gov/projects/trecvid)) is regarded as a promising approach to facilitating higher level manipulations of video data by bridg-