

Abstract

The population of the World Wide Web with media of all types such as texts, images, videos and audio files in recent years raised the attractiveness of multimedia retrieval. With our work on the influence of dependencies between modalities and features we investigate why these approaches still do not perform convincingly better than plain text search approaches when applied to large, noisy collections like web content, even though these approaches have more information at their hands.

This article suggests that, due to the size and noise, the modality's dependencies necessary for efficient information fusion becomes small and hard to exploit. Preliminary experiments with two multi modal collections underpin this statement.