TIME-SERIES ACTIVE SEARCH FOR QUICK RETRIEVAL OF AUDIO AND VIDEO

Kunio Kashino, Gavin Smith and Hiroshi Murase*

NTT Basic Research Laboratories
3-1 Morinosato-Wakamiya, Atsugi-shi, Kanagawa, 243-0198 Japan.
kunio@ca-sun1.brl.ntt.co.jp, murase@eye.brl.ntt.co.jp
* Currently with Cambridge University, gas1003@eng.cam.ac.uk

ABSTRACT

This paper proposes a search method that can quickly detect and locate known sound (video) in a long audio (video) stream. The method is based on active search [1]. Active search reduces the number of candidate matches between reference and input signals by approximately 10 to 100 times compared to exhaustive search, while guaranteeing the same retrieval accuracy. We proposed a quick search method in [2], and here we focus on improvement of the accuracy. Thus the feature used has been extended to the audio power spectrum and temporal division of the histogram windows has been introduced to incorporate time information. Tests carried out under practical circumstances clearly show the accuracy improvement. The proposed method is still so fast that it can correctly retrieve a 15-s commercial in a 6-h recording of TV broadcasting within 2 s, once the features are calculated.

1. INTRODUCTION

This paper discusses a method to search quickly through a long audio or video stream (termed an *input signal*) to detect and locate a known reference audio or video signal (termed a *reference signal*). One application in mind is searching and retrieval of music from unlabeled audio archives, videos or the Internet. Another is monitoring occurrences of a TV commercial or the theme music of a TV program.

Even if a reference signal is known, a huge amount of computation is required for the feature matching when a long input signal is assumed. Adopting heuristic time-skipping in the matching process may partially reduce the computational load, but may also result in deterioration of the recall rate [1] (increase of misses).

We proposed a quick audio retrieval method using the active search algorithm [2]. However, the method was not necessarily accurate enough under practical circumstances (e.g. search for real TV recordings). Therefore we focus on improving retrieval accuracy maintaining the quickness that characterizes active search. To this end, the feature is extended to the audio power spectrum. In addition, the temporally divided histogram windows are introduced to incorporate time information. The framework also integrates video retrieval using color features.

Section 2 overviews the search algorithm. Section 3 evaluates the speed and accuracy of the algorithm using recordings of real TV broadcasting. Concluding remarks are given in Section 4.

[1] The recall rate is defined as the number of correctly retrieved objects divided by the number of objects that should be retrieved. The precision rate (appearing in Section 3) is defined as the number of correctly retrieved objects divided by the number of all retrieved objects.

2. SEARCH ALGORITHM

2.1. Overview

Figure 1 outlines the proposed algorithm. Firstly, the feature vectors are calculated from both the reference signal and input signal. The windows are then applied to both the reference and input feature vectors. The feature vectors over the windows create the histogram. The window length may be the same as the reference signal duration. For the incorporation of time-sequence information, however, the windows can be temporally divided into N_{div} subwindows as shown in the figure. Thirdly, similarity between the reference histogram and input histogram is calculated. When the similarity exceeds a threshold value chosen in advance, the reference signal is detected and located. In the last step, the window on the input signal is shifted forward in time and the search proceeds.

2.2. Feature Extraction

The features are the audio power spectrum and colors.

Audio feature vector $f(k)$ is written as

$$f(k) = (\ f_1(k),\ f_2(k),\ \cdots,\ f_N(k)\ )\ , \tag{1}$$

where k is the sampled time. An element of $f(k)$ is the normalized short-time power spectrum, which is given as

$$f_j(k) = \alpha(k)\ Y_j(k), \tag{2}$$

$$Y_j(k) = \sum_{t=k-M+1}^{k} y_j^2(t), \tag{3}$$

$$k = lM \quad (l = 1, 2, \cdots), \tag{4}$$