Very quick audio searching

~ introducing global pruning
to the time-series active search ~

Akisato Kimura, Kunio Kashino,
Takayuki Kurozumi and Hiroshi Murase

NTT Communication Science Labs.

Backgrounds

a 24-hour media information

a month or a year media information

Time-series active search
[Kashino at el,1999]

a quick search method for media information

an even faster method for media information

a target of this work
Objectives

- detect and locate a reference signal
- the signal segments to be detected preserve the same spectrum pattern as the reference signal, except for minor distortions or noises

Time-series active search (TAS)

[Kashino et al. 1999]
**Time-series active search (TAS)**  
[Kashino et al. 1999]

- Takes less than 1 second to detect a 15 second reference signal in a 24-hour stored signal.
- Guarantees search accuracy with respect to the similarity equivalent to L$_1$-distance.

\[ S_1(x, y) = \frac{1}{D} \sum_{i=1}^{L} \min(x_i, y_i) \]  
**histogram intersection**

\[ d_1(x, y) = \sum_{i=1}^{L} |x_i - y_i| \]  
**L$_1$-distance**

\[ x = \{x_i\}_{i=1}^{L}, \ y = \{y_i\}_{i=1}^{L} \]  
(a reference histogram and stored histograms)

takes less than 1 second to detect a 15 second reference signal in a 24-hour stored signal.

guarantees search accuracy with respect to the similarity equivalent to L$_1$-distance.

---

**Approach**

**<TAS method>**
- local pruning (=skipping)

**<proposing method>**
- global pruning + local pruning

- Guaranteeing search accuracy with respect to a certain similarity standard.
Global pruning

TAS (local pruning)

new method (global & local pruning)

matching unnecessarily

Procedure of TAS

<preprocessing stage>
1. extract features from the stored signal

<search stage>
1. extract features from the reference signal and create histograms
2. match histograms
Procedure of proposed method

**<preprocessing stage>**
1. extract features from the stored signal and create histograms
2. divide the histogram space
3. classify the histograms

**<search stage>**
1. extract features from the reference signal and create a histogram
2. perform global pruning
3. match histograms

Extract features from the stored signal and create histograms

- extract features by using a band-pass filter bank
- create histograms by classifying the feature vectors over the window
Divide the histogram space

- each division is used as a unit to determine whether matching is needed
- use a codebook learning algorithm for vector quantization (VQ) (ex. LBG algorithm)
- distance measure: $L_2$-distance (Euclid distance)

$$d_2(x, y) = \sqrt{\sum_{i=1}^{L} (x_i - y_i)^2}$$

- histogram clusters: the divisions of the histogram space

Classify histograms

- classify histograms based on $L_2$-distance and using a certain VQ algorithm
Extract features from the reference signal and create a histogram

- extract features using a band-pass filter bank
- create histograms by classifying the feature vectors

Perform global pruning

- the divided histogram space
- reference histogram
Perform global pruning
<br><br>select clusters that may contain a histogram whose distance from the reference histogram is smaller than the selection threshold

Experiments

- stored signal: a recording of real 24-hour TV broadcasting
- reference signal: 12 sections of 15 seconds randomly chosen from the stored signal
- number of histogram bins: 512
- number of histogram clusters: 512
Review: procedure of proposed method

<preprocessing stage>
1. extract features from stored signal and create histograms
2. divide the histogram space
3. classify the histograms

<search stage>
1. extract features from reference signal and create a histogram
2. perform global pruning
3. match histograms

Results: search time

- search time includes only the time for the histogram matching step

<table>
<thead>
<tr>
<th>Method</th>
<th>Search Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAS</td>
<td>390 msec</td>
</tr>
<tr>
<td>Proposed method</td>
<td>30 msec</td>
</tr>
</tbody>
</table>

(13 times faster than TAS)

(when the search threshold is 0.85)
Results: search time

- detect the segments 13 times faster than TAS (search threshold = 0.85)

Results: number of matches

- reduces the number of matches to 1/10 (search threshold = 0.85)
Conclusions

- We have proposed a global pruning method that enables a very quick search of audio signals.
- It reduces the number of matches to 1/10 (search threshold = 0.85).
- Extension to video search is planned.

Results:

<table>
<thead>
<tr>
<th></th>
<th>preprocessing stage</th>
<th>search stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAS</td>
<td>36 min</td>
<td>410 msec</td>
</tr>
<tr>
<td>proposed method</td>
<td>67.2 min</td>
<td>70 msec</td>
</tr>
</tbody>
</table>

(when the search threshold is 0.85)