Bimodal Codebooks Based Adult Video Detection

Yizhi Liu, Junlin Ouyang, Jianxun Liu

Hunan University of Science and Technology,

Xiangtan, China

Contents

- Introduction
- Related work
- The framework of our approach
- Experiments
- Conclusions

1 Introduction

- More attention is necessary to be paid on adult video detection
- Existing works are mostly based on visual features of keyframes
- However, it is difficult for keyframe-based methods to accurately detect adult videos owing to a large amount of low-resolution videos

1.1 Our motivation

- Multi-modality based approach is much more effective.
  - In adult videos, audio signals, such as periodic moaning and screaming, are conspicuous.
- Therefore, we are motivated by combining audio information with visual keyframes to describe the co-occurrence semantics.

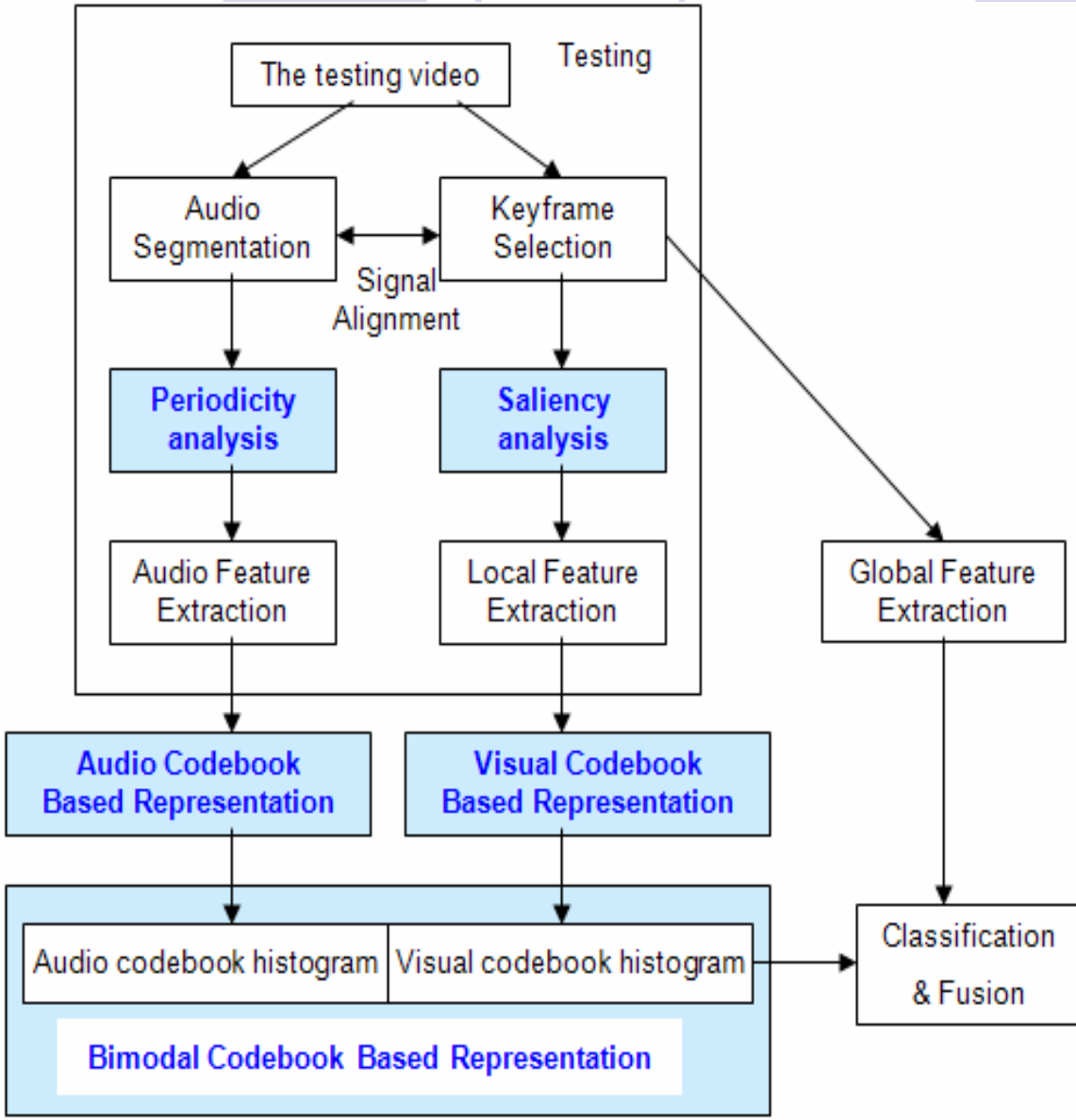
2 Related work

- Content-based adult video detection is traditionally based on visual features of keyframes
  - Single frame: Forsyth et al. [5]; Zeng et al. [6]; Rowley et al. [7]; Tang et al. [8]
  - Multi-frames: Lee et al. [1]; Kim et al. [2] Traditionally, skin-color regions are always extracted as ROI
  - Visual codebook: Deselaers et al. [14]; Wang et al. [15]
  - Multi-modality: Rea et al. [3], Endeshaw et al. [25], Jansohn et al. [26]

2.1 Main challenges in existing methods

- ROI detection
  - It is difficult to differentiate between human skins and other objects with the skin-colors
- audio periodicity analysis
  - Low-level audio features are too similar to be distinguished accurately
- multimodal information fusion
  - Few works have focused on representing the cooccurrence semantics of multimodal signals
- ……

3 The framework of our approach

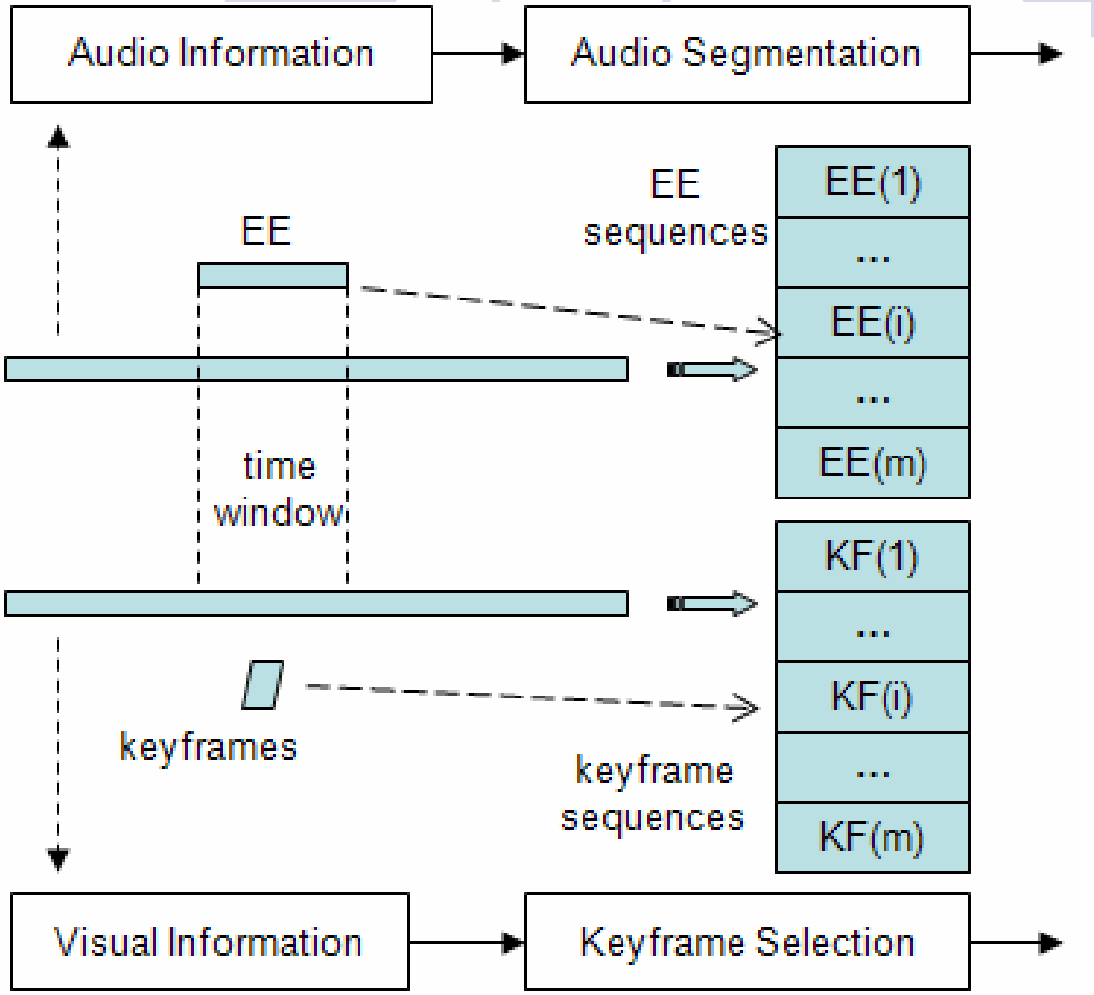


- Our framework includes five modules:
  - signals segmentation and alignment
  - audio periodicity analysis
  - ROI detection
  - bimodal codebook based presentation
  - result decision

3.1 The overview of our framework

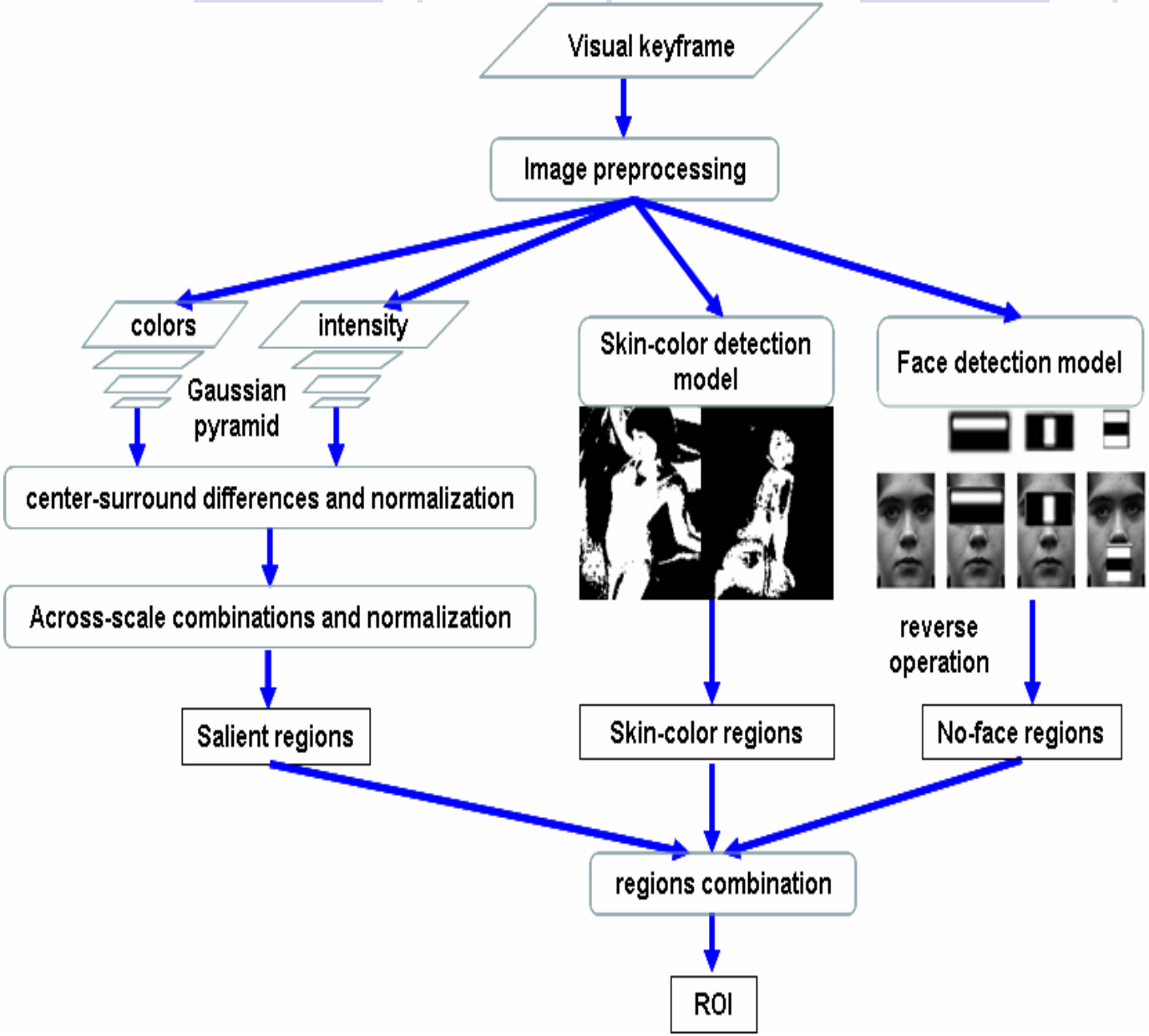
- In our framework, audio frames are segmented into units of energy envelope based on audio periodicity analysis, and visual keyframes are detected based on saliency analysis.
- The lengths of energy envelope (EE) not the same but variable. Subsequently, audio signals are described by the EE sequences with audio codebook based presentation.
- Visual codebook is constructed based on ROI detection, which combines saliency analysis and skin-color detection.
- And visual features are represented by the middle level semantics.
- Results show that our framework achieves more excellent performance than some state-of-the-art methods.

3.2 Periodicity analysis and signals alignment



- In the time window of every EE, visual keyframe is chosen from many frames
- We present EE segmentation algorithm based on audio periodicity analysis

3.3 Saliency analysis and ROI detection



- We propose a hybrid approach of ROI detection combined three kinds of models:
  - saliency analysis model
  - skin-color model
  - face detection model

A hybrid approach of ROI detection

- The saliency-based model and the contrast-based model are two typical kinds of visual attention models. The former model is time-consuming because of massive computations and the latter one has limited capability to highlight human-beings in the images.
- Therefore, we propose a hybrid algorithm to fuse the two preceding models
- Furthermore, we adopt the skin-color model proposed by Garcia et al. and the face detection model proposed by Viola
- Finally, we take the intersecting part of salient regions, skin-color regions, and no-face regions as ROI

3.4 Bimodal codebooks based presentation

- The procedure of constructing bimodal codebook histogram
  - At first, we select some adult videos and extract audio and visual low-level features
  - After audio periodicity analysis and ROI detection, the audio codebook and the visual codebook are respectively created by K-means clustering algorithm
  - Next, low-level audio and visual features of the testing video are respectively converted into mid-level semantic histograms via the audio or visual codebook
  - The histograms are concatenated to represent the coocurrence semantics of bimodal (audio and visual) signals
  - Finally, we fuse the classification results of bimodal codebooks based presentation with that of visual global features.

4 Experiments

- We collect videos from the Internet and respectively set up a training dataset and a testing dataset.
  - There are forty eight adult videos and three hundred benign ones in the training dataset.
  - And the testing dataset includes fifty adult videos and one hundred and fifty benign ones.
  - We evaluate our approach in the visual studio 2003 environment with the machine of 1.86 GHz Duo CPU and 2GB memory.
- We evaluate our method with receiver operating characteristic (ROC) curves.
  - A ROC space is defined by false positive rate (FPR) and true positive rate (TPR) as x and y axes respectively.
  - On the basis of our previous works [8, 16, 17], we adopt color moments as the global features, SURF as the local features, and SVM classifiers.

4.1 Evaluation of the proposed ROI detection method

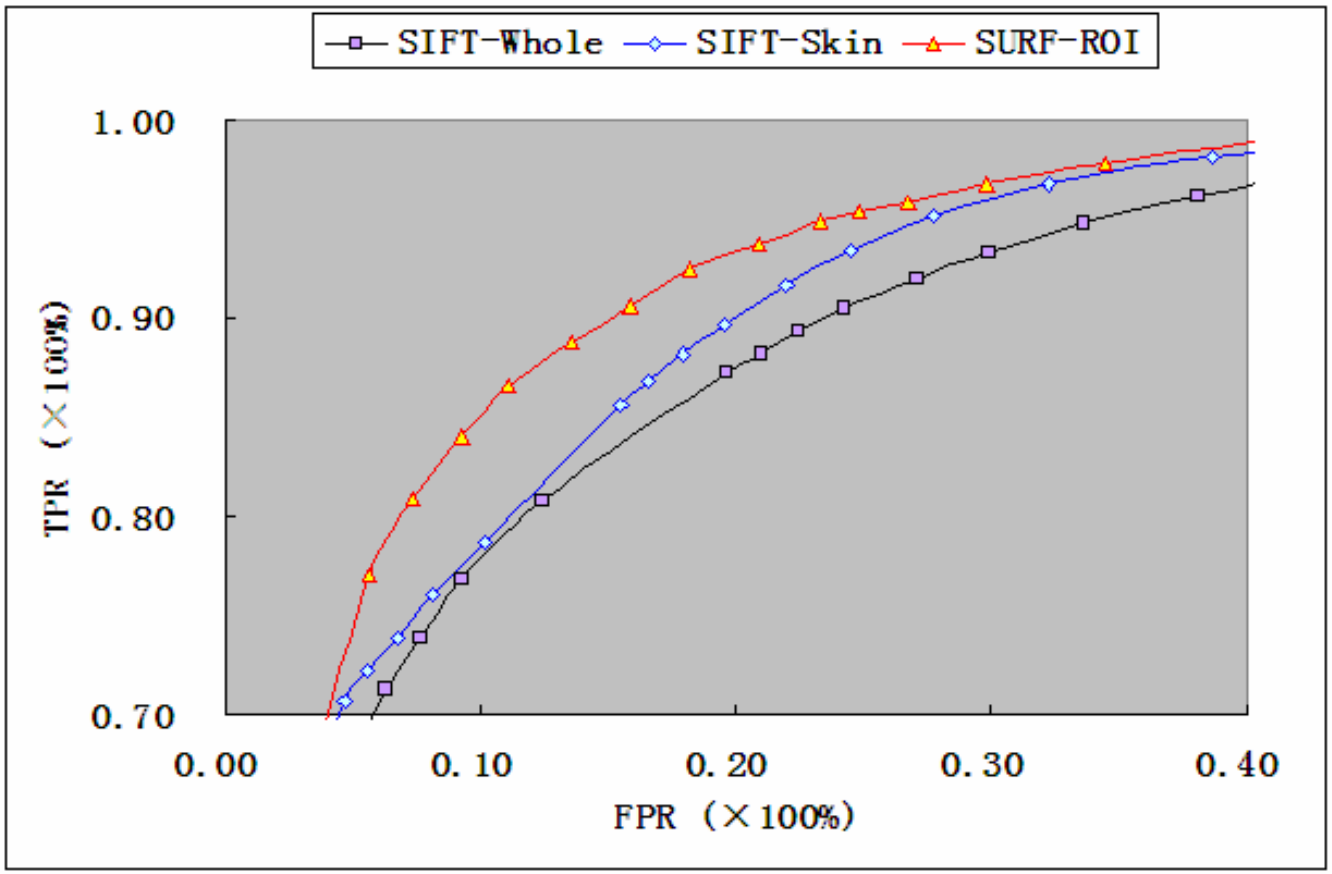
- According to the evaluation results, we can conclude that the proposed ROI detection method achieves good performance and the precision reaches 91.33% in average.
- It is able to detect ROI in adult images more precisely than using Itti's model [10], Ma's model [11], and Garcia's model [12] alone.

	GOOD	ACCEPT	FAILED
Group 1	41.71%	49.43%	8.86%
Group 2	41.43%	51.14%	7.43%
Group 3	46.29%	44.00%	9.71%
Average	***91.33%***		8.67%

Examples of the proposed ROI detection method

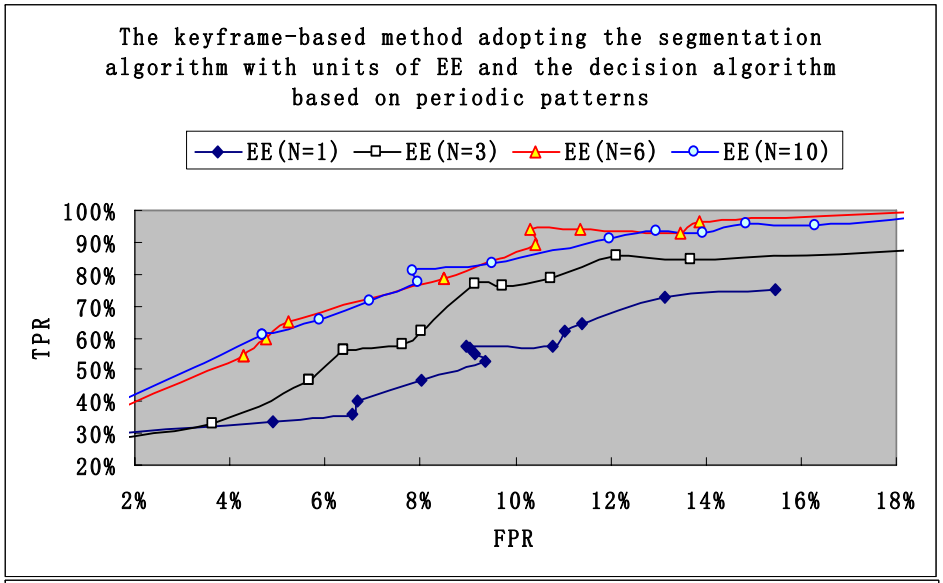
4.2 Evaluation of a codebook based method

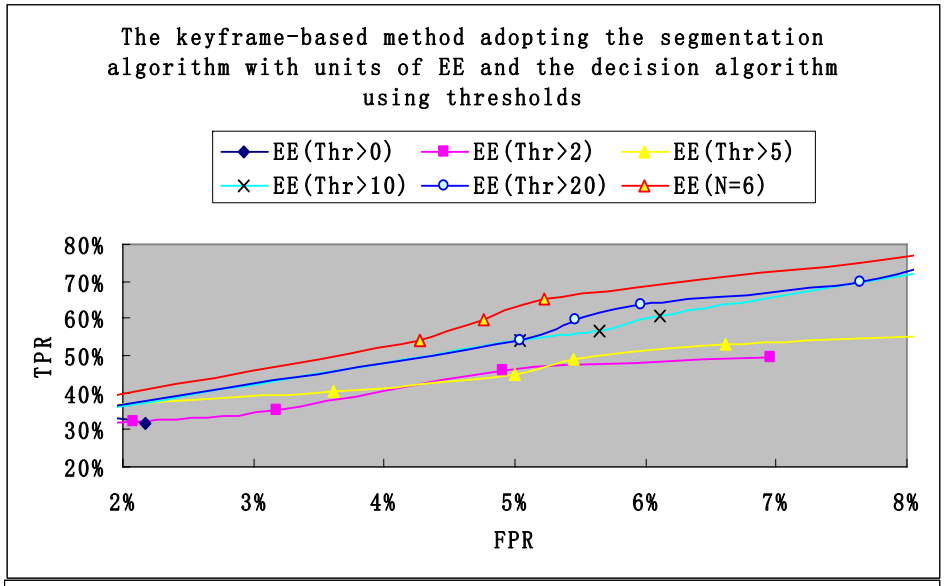
- Results show that the proposed ROI-based codebook algorithm is able to remarkably improve many other visual codebook based algorithms.

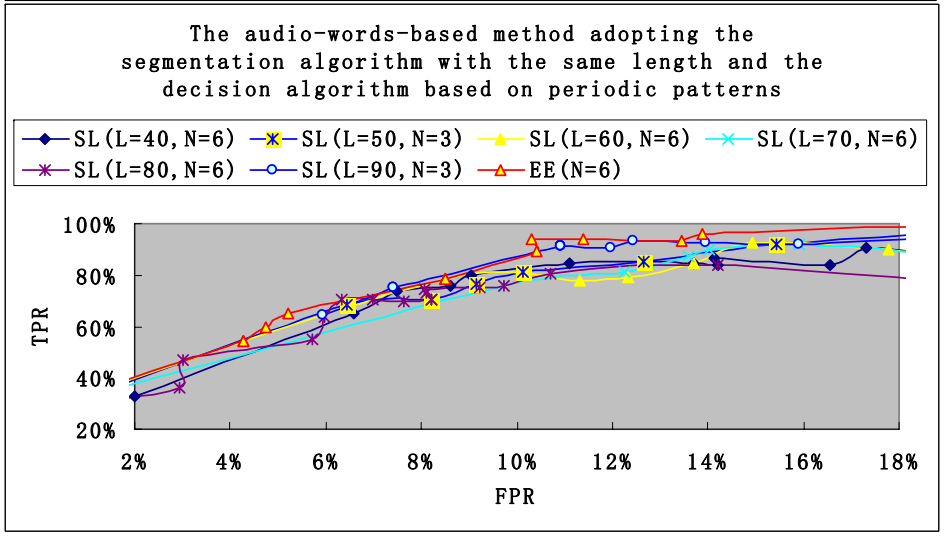


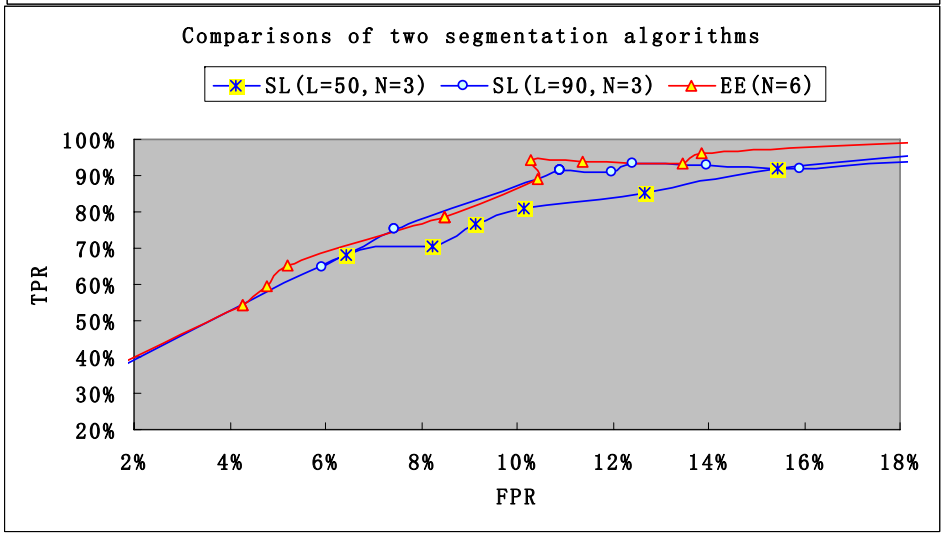
4.3 Evaluation of audio codebook based method

- Parameter regulation, and performance evaluation



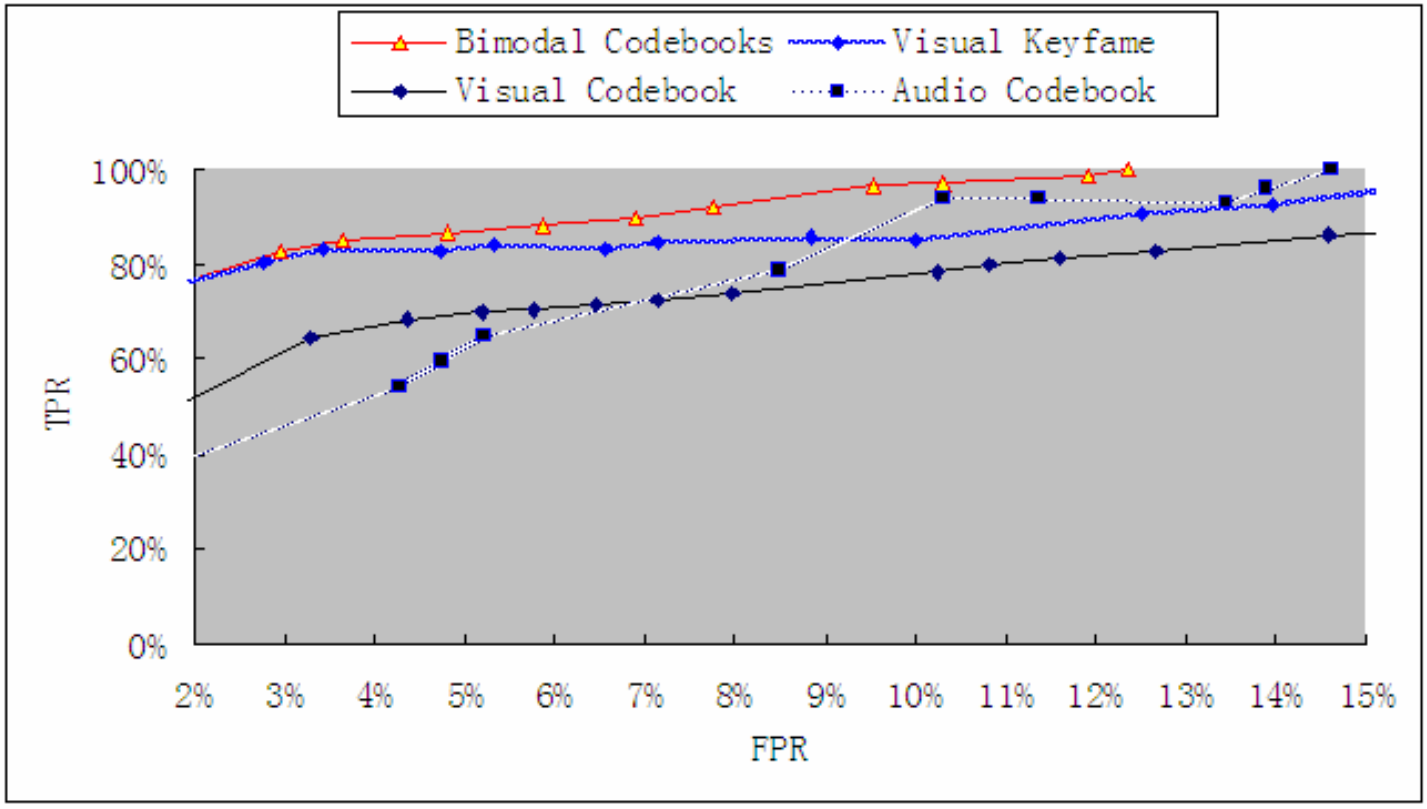






4.4 Evaluation of bimodal codebooks based presentation

- Experimental results show that our approach outperforms the traditional one which is based on visual features, and achieves satisfactory performance. The true positive rate achieves 96.7% while the false positive rate is about 10%.



5 Conclusions

- Multi-modality based adult video detection is an effective approach of filtering pornography. But the performance of existing methods is not good enough owing to lacking accurate multi-modality semantics representation.
- Therefore, we put forward bimodal codebooks based adult video detection which merges periodicity analysis based audio codebook representation and saliency analysis based visual codebook representation.
- Our intention is not only to combine the two modalities of visual images and audio signals, but also to narrow down the semantic gap between low-level features and high-level concepts by constructing bimodal codebooks.
- The results show that our approach outperforms some state-of-the-art methods.

湖南科技大学
Hunan University of Science and Technology

Thank you.