同济大学计算机科学与技术系

多媒体与智能计算实验室

MIC-TJU at MediaEval Violent Scenes Detection (VSD) 2014

Bowen Zhang, Yun Yi and Hanli Wang

Tongji University

Content

Content

Introduction

- Violent Scene Detection (VSD) contains two sub-task: Main Task and Generalization Task.
- Train set: 24 movies from Hollywood
- Test set:
  - Main Task: 5 movies from Hollywood
  - Generalization Task: 86 web videos

Introduction

- **Challenges:**
  - Difficulties in feature detection since no shot boundary given
  - Camera jitters make it hard to track trajectories

Shot from Main Task

Shot from Generalization Task

Introduction

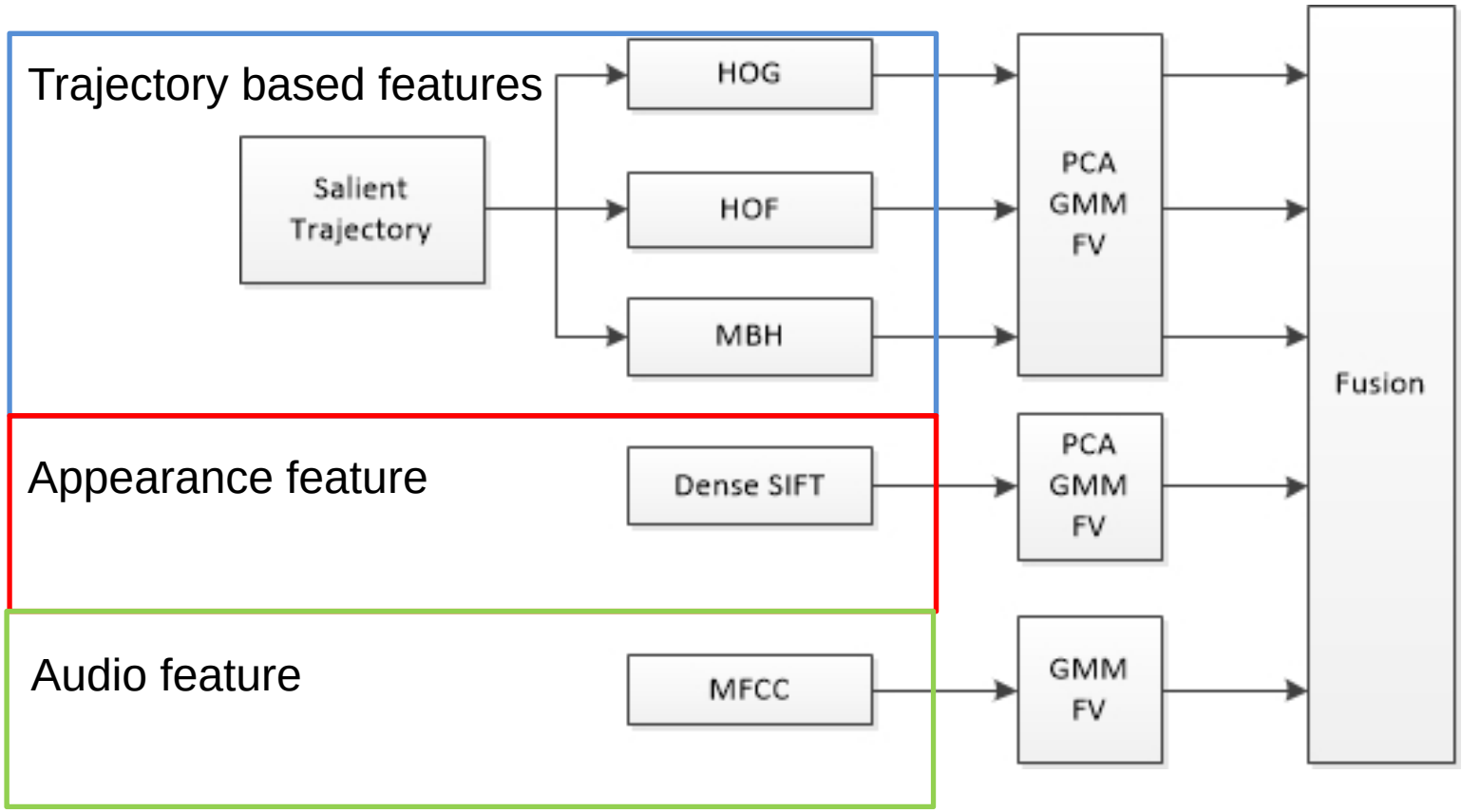
- **Motivation**
  - Use shot boundary detection algorithm to detect shot boundary.
  - Use camera motion elimination technique to remove camera jitters.

Content

System Description

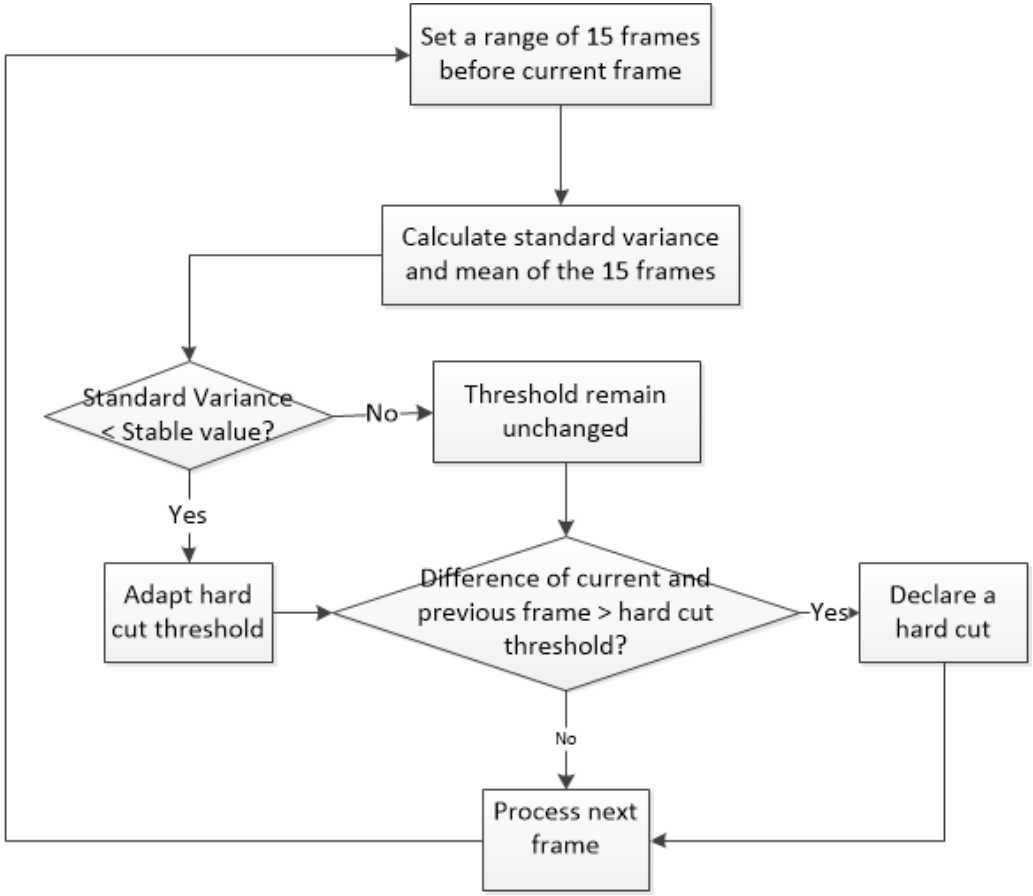
- **Our approach:**
  - Shot boundary detection
  - Trajectory based video features
  - Appearance video feature
  - Audio feature

System Description



Content

Shot boundary detection



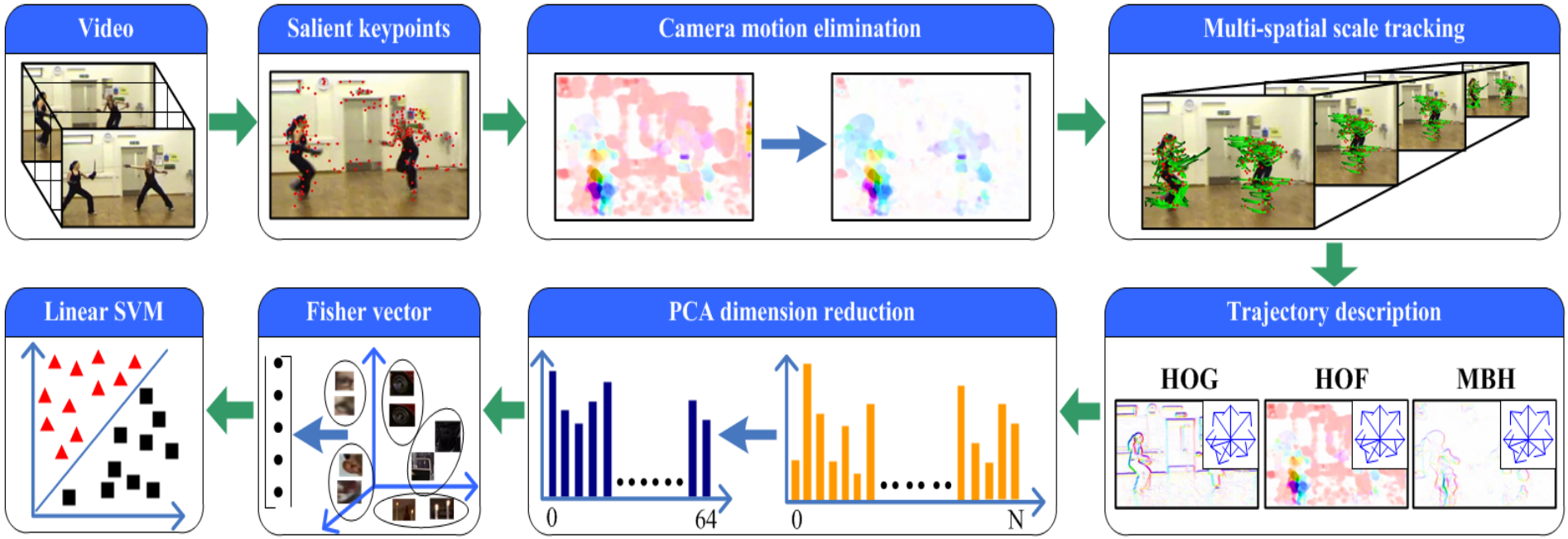
- **Based on difference of histograms**
- **Adaptive threshold**

Content

Video and audio features

- **Video and audio features**
  - **Trajectory based features**
  - **Camera motion elimination**
  - **Appearance feature**
  - **Audio feature**

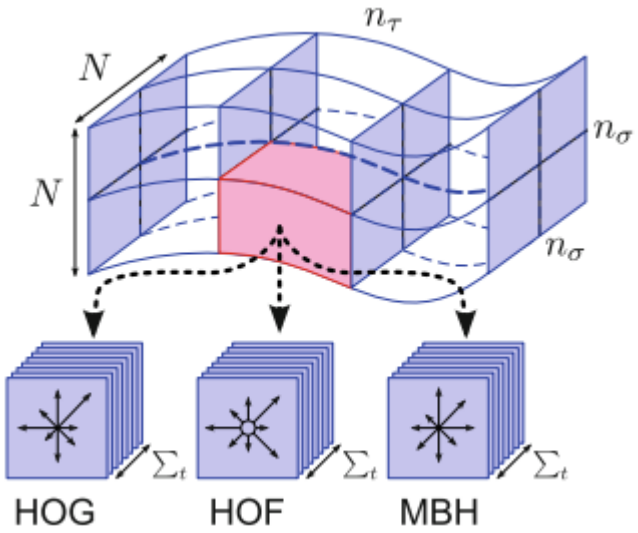
Trajectory based features



- ✓ Saliant Trajectory
- ✓ Camera motion elimitation
- ✓ Classification

Trajectory based features

- Why?
  - Choose good points for tracking.
- Saliant trajectory appraoch:
  - SIFT keypoints detection
  - Dense optical flow
  - Multiple spacial scale tracking
  - Trajectory descriptions
    - HOG (96=2×2×8×3)
    - HOF (108=2×2×9×3)
    - MBH (192=2×2×8×3×2)

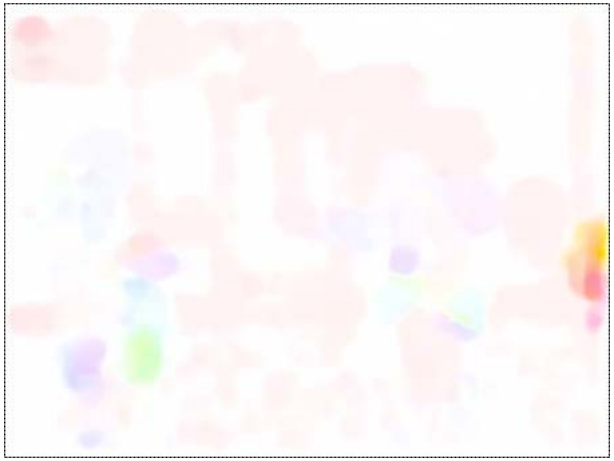


Trajectory based features

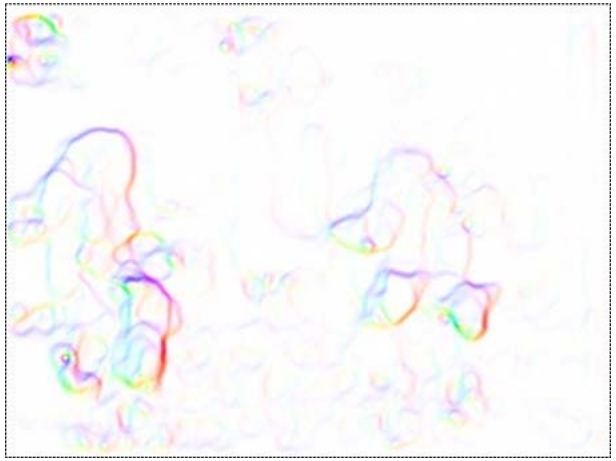
(a) Frame k

(b) Frame k+1

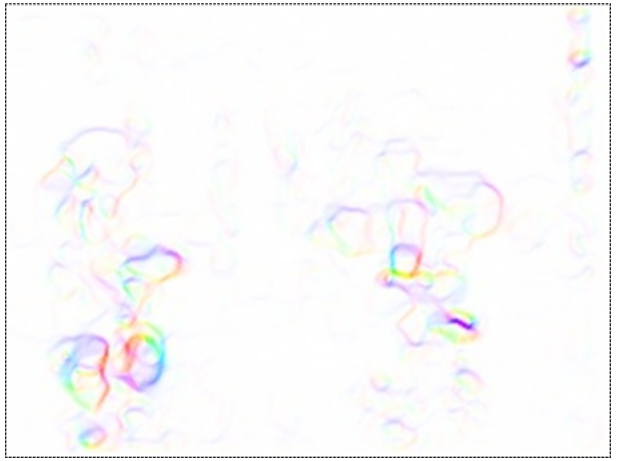
(c) HOG



(d) HOF



(e) MBHx



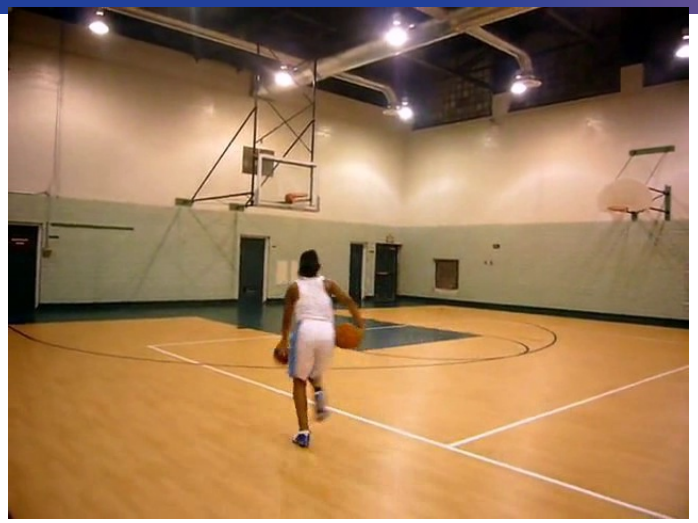
(f) MBHy

In Munsell system, orientation is indicated by color and magnitude by saturation.

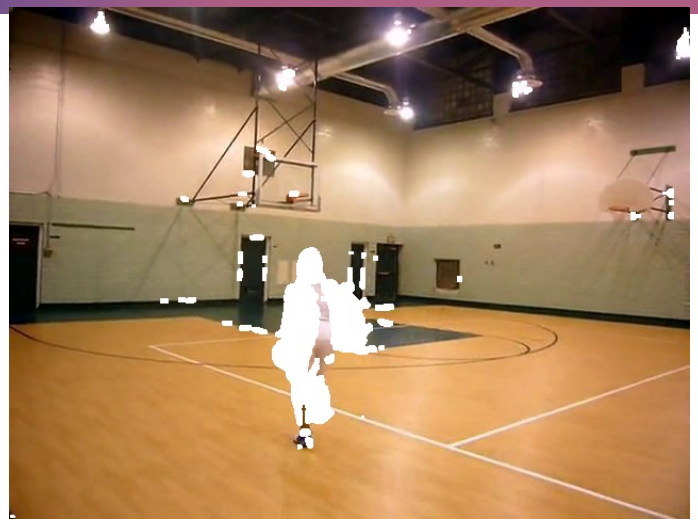
Camera motion elimitation

- **Why?**
  - **Action may be fused by camera motion.**
- **How?**
  - **Background detection (Pixel-Based Adaptive Segmenter method)[2]**
  - **Keypoints match**
  - **Homography estimation (Random Sample Consensus)**
  - **Frame rectification**

Camera motion elimitation



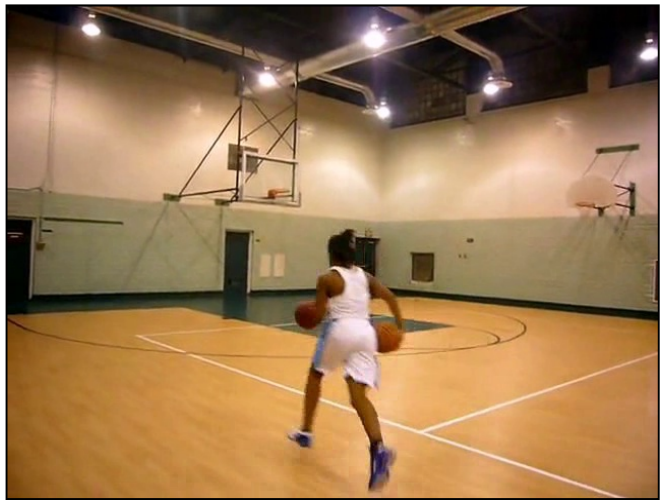
(a) Frame k



(b) Background

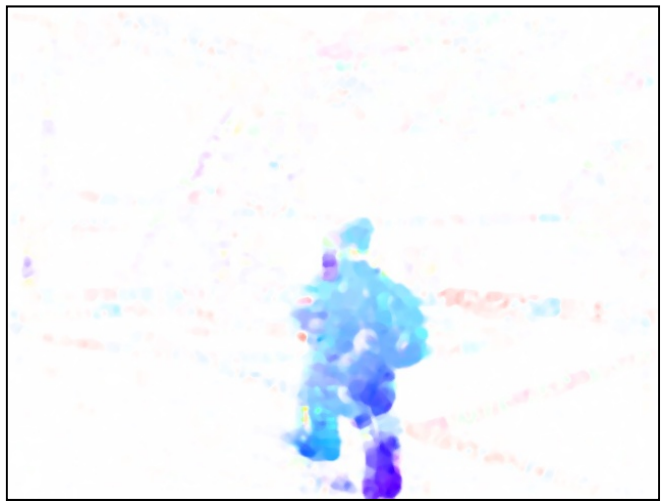
(c) Keypoint match

Camera motion elimitation

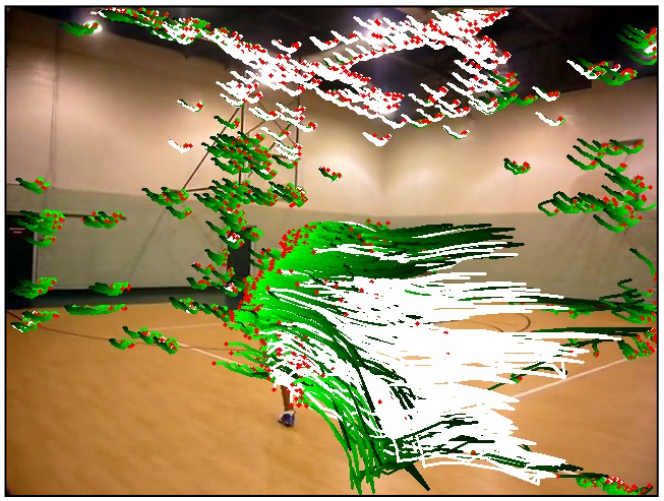


(a) Frame k

(b) Optical flow



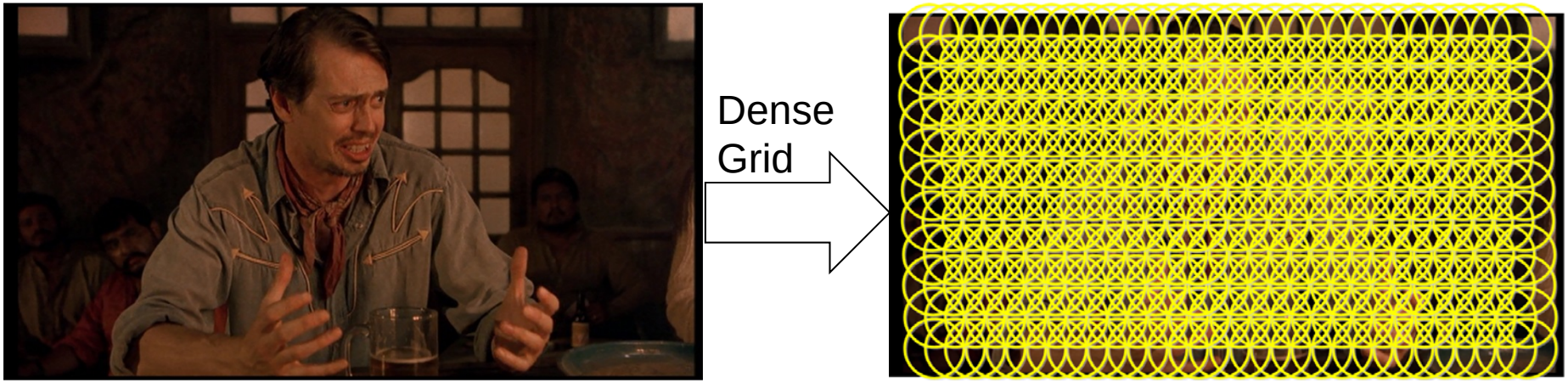
(c) Warped optical flow



(d) Trajectory

Appearance feature

- **Dense SIFT**
  - **Dense Grid: 21 × 21 patches with 4 pixel steps and 5 scales**
  - **SIFT: Calculate SIFT on dense grid.**



Audio feature

- **MFCC**
  - **The time window for each MFCC is 32 ms.**
  - **There is 50% overlap between two adjacent windows.**
  - **We integrate delta and double-delta of 20 dimensions MFCC vector into the original MFCC vector to generate a 60-dimension MFCC vector.**

Content

Conclusion

- **Configuration of submitted runs**

Run	Trajectory based Features	Appearance Feature	Audio Feature	Fusion	Weights
Run 1	HOG, HOF, MBH	-	MFCC	Late Fusion	4:1
Run 2	HOG, HOF, MBH	Dense SIFT	MFCC	Double Fusion	4:1
Run 3	HOG, HOF, MBH	Dense SIFT	MFCC	Double Fusion	1:1
Run 4	HOG, HOF, MBH	Dense SIFT	MFCC	Late Fusion	4:1:1
Run 5	HOG, HOF, MBH	Dense SIFT	MFCC	Late Fusion	4:1:1

Conclusion

- **Configuration of submitted runs**
  - **In the late fusion, an arithmetic sum of scores outputted from SVM for video features (trajectory based features and appearance feature) and audio feature is calculated.**
  - **In double fusion, we firstly early fuse video features and then late fuse video features and audio feature.**
  - **The weight setting segmented by colon in Table stands for the weights applied to different kinds of features during late fusion.**

Conclusion

- **Results**

Run	Main Task	Generalization Task
Run 1	44.17%	56.01%
Run 2	43.07%	56.52%
Run 3	44.60%	55.56%
Run 4	39.23%	56.62%
Run 5	38.50%	56.00%

- **Metrics of result is MAP2014.**
- **Dense SIFT improves score.**
- **Generalization Task outperform Main Task, because shots in Generalization Task do not change as frequent as that in Main Task.**

Thank you!