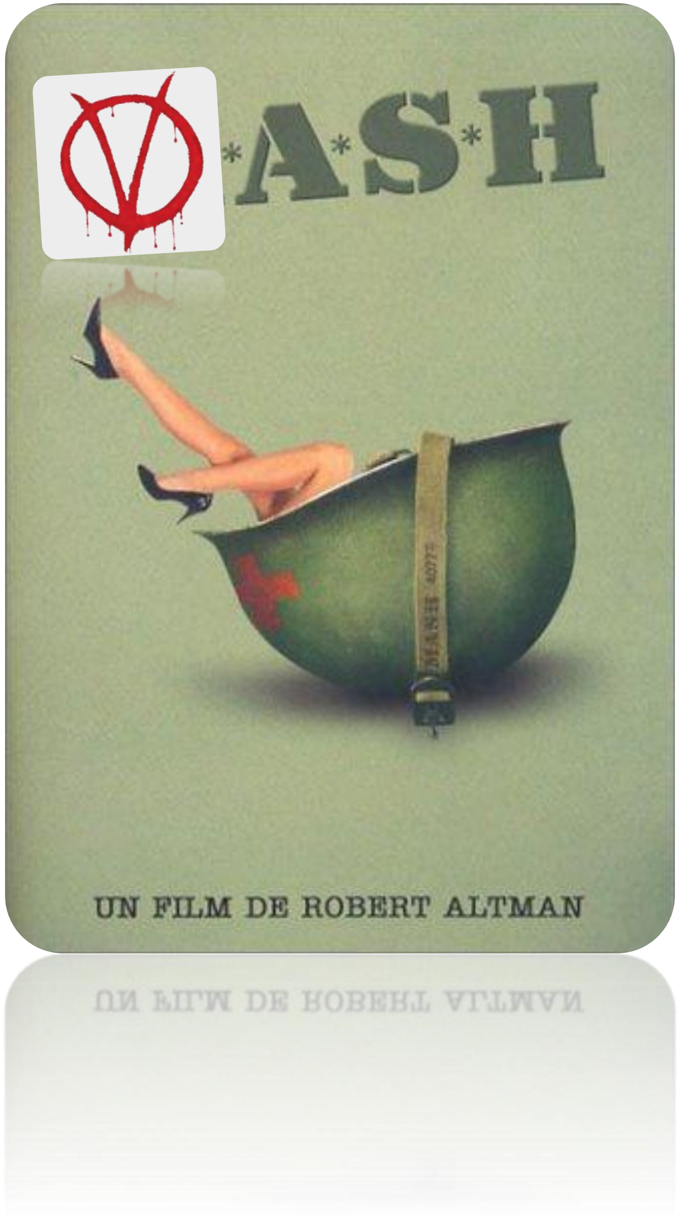
MMAI Term Project
VASH

郝瑞尼 蔡格昇 蔡宗諭

The Vision

“Perceptual Movie Identification”

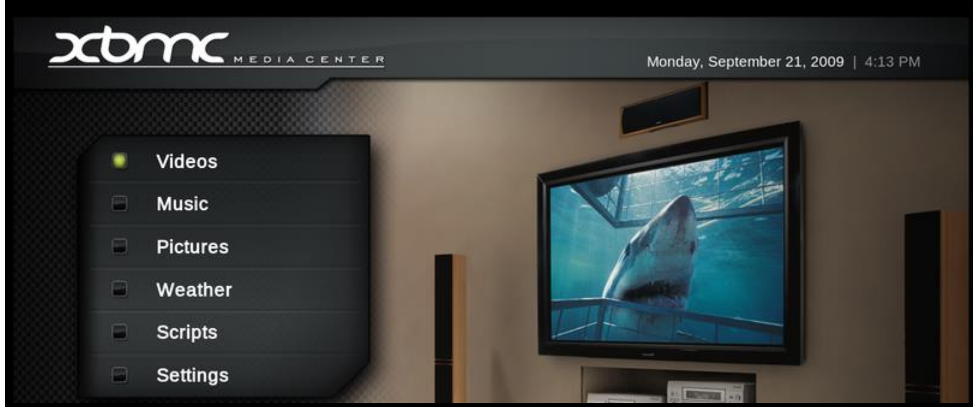
- Look at Movie content
- Identify Movie (name, year, id)
- 100% accuracy and fully automatic

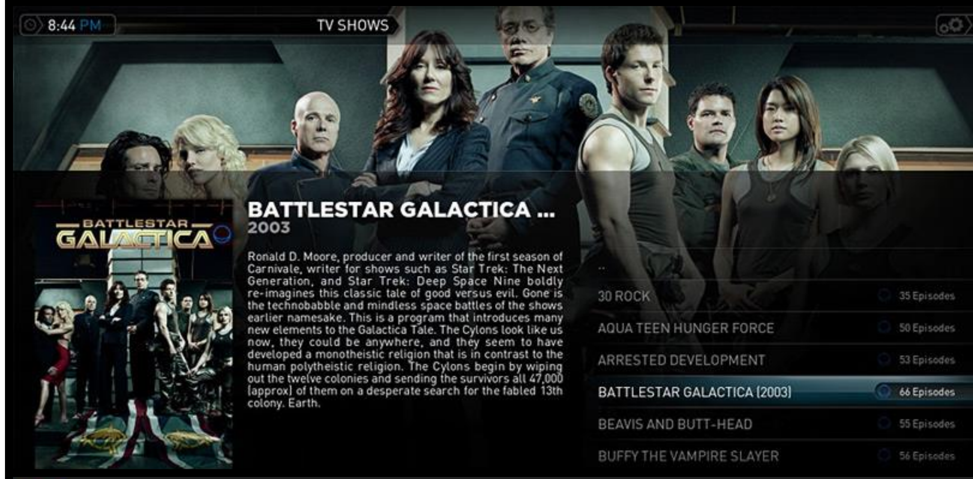


The Why?

- Virtually NO metadata / tagging
- HTPC software used by millions
- Manually organizing media files
- Wasting millions of hours

“ We can do better! “

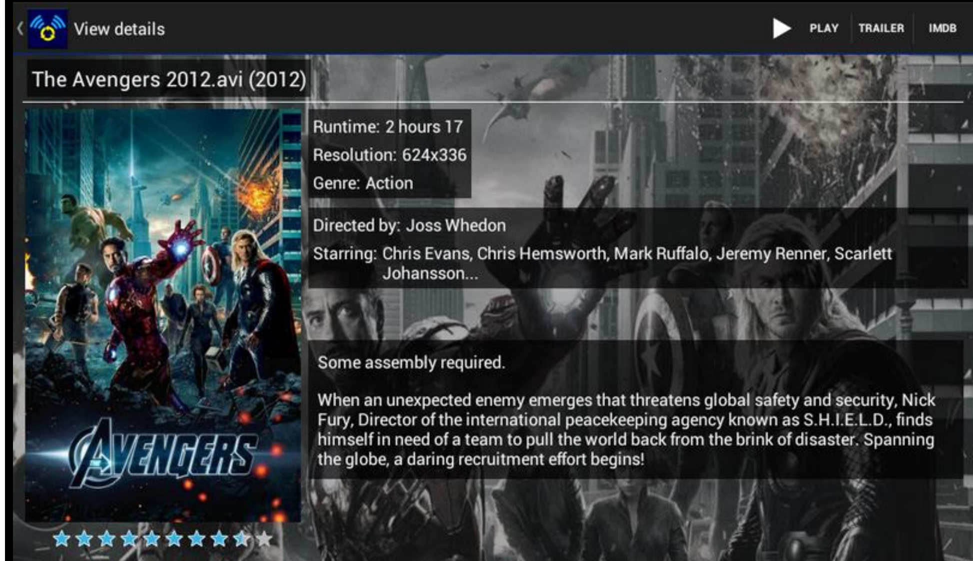




BATTLESTAR GALACTICA ...
2003

Ronald D. Moore, producer and writer of the first season of Carnivale, writer for shows such as Star Trek: The Next Generation, and Star Trek: Deep Space Nine boldly re-imagines this classic tale of good versus evil. Gone is the technobabble and mindless space battles of the shows earlier namesake. This is a program that introduces many new elements to the Galactica Tale. The Cylons look like us now, they could be anywhere, and they seem to have developed a monotheistic religion that is in contrast to the human polytheistic religion. The Cylons begin by wiping out the twelve colonies and sending the survivors all 47,000 (approx) of them on a desperate search for the fabled 13th colony. Earth.

..	
30 ROCK	35 Episodes
AQUA TEEN HUNGER FORCE	50 Episodes
ARRESTED DEVELOPMENT	53 Episodes
BATTLESTAR GALACTICA (2003)	66 Episodes
BEAVIS AND BUTT-HEAD	55 Episodes
BUFFY THE VAMPIRE SLAYER	56 Episodes



The Avengers 2012.avi (2012)

Runtime: 2 hours 17
Resolution: 624x336
Genre: Action

Directed by: Joss Whedon
Starring: Chris Evans, Chris Hemsworth, Mark Ruffalo, Jeremy Renner, Scarlett Johansson...

Some assembly required.

When an unexpected enemy emerges that threatens global safety and security, Nick Fury, Director of the international peacekeeping agency known as S.H.I.E.L.D., finds himself in need of a team to pull the world back from the brink of disaster. Spanning the globe, a daring recruitment effort begins!

I spent 2 days organizing all my movies, series and documentaries. They were already extremely organized, but not good enough for XBMC. So I created an NFO file for every episode. **Was that enough? No.**

douglasrac

Junior Member

Posts: 9
Joined: Sep 2013
Reputation: 0

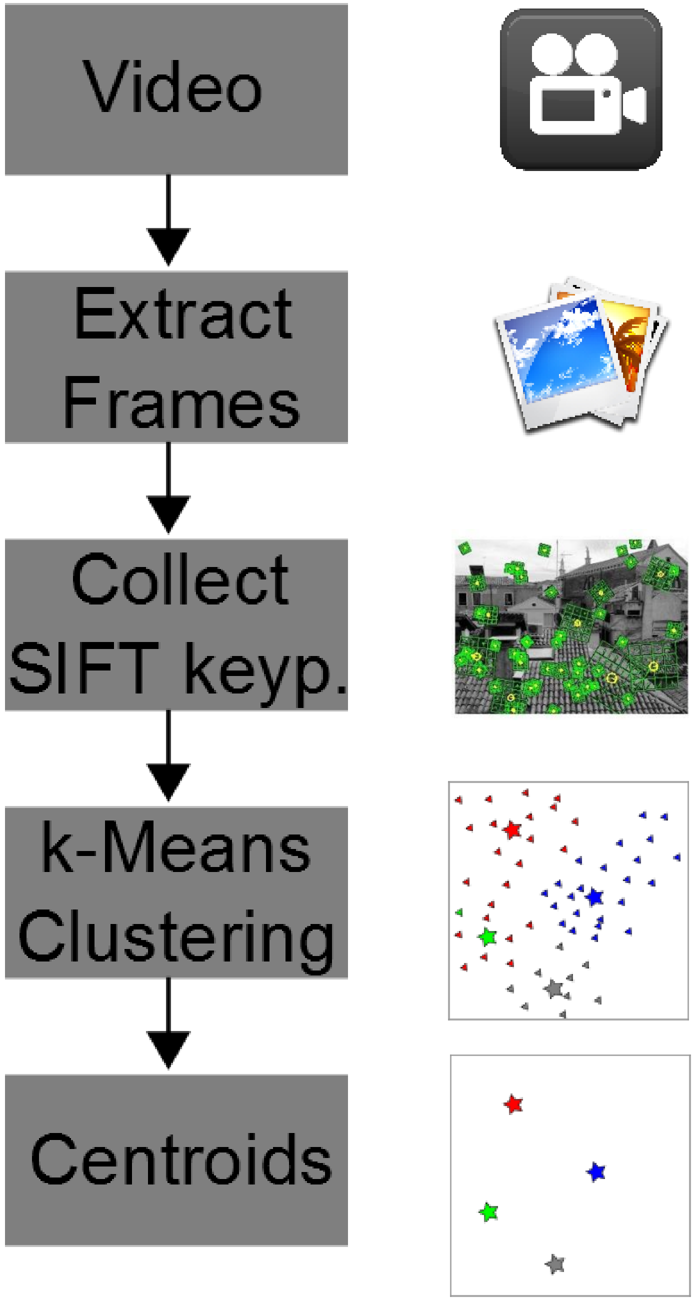
The How

- Sample key frames from video files
- Extract SIFT keypoints for video (sub-)sequences
- Compute signature using Histogram of Visual Words
- Match signature against our reference database
- Return video identity

“ Lifeng Shang et al, 2010, Real-time large scale near-duplicate web video retrieval ”

Generation of Visual Word Dictionary

- Select a number of frames from videos
- Compute and aggregate
- SIFT keypoints
- Run k-means clustering
- The resulting k-centroids are the Visual Words



Video

Extract Frames

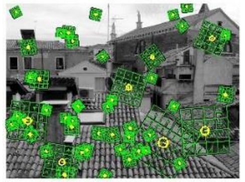
Collect SIFT keyp.

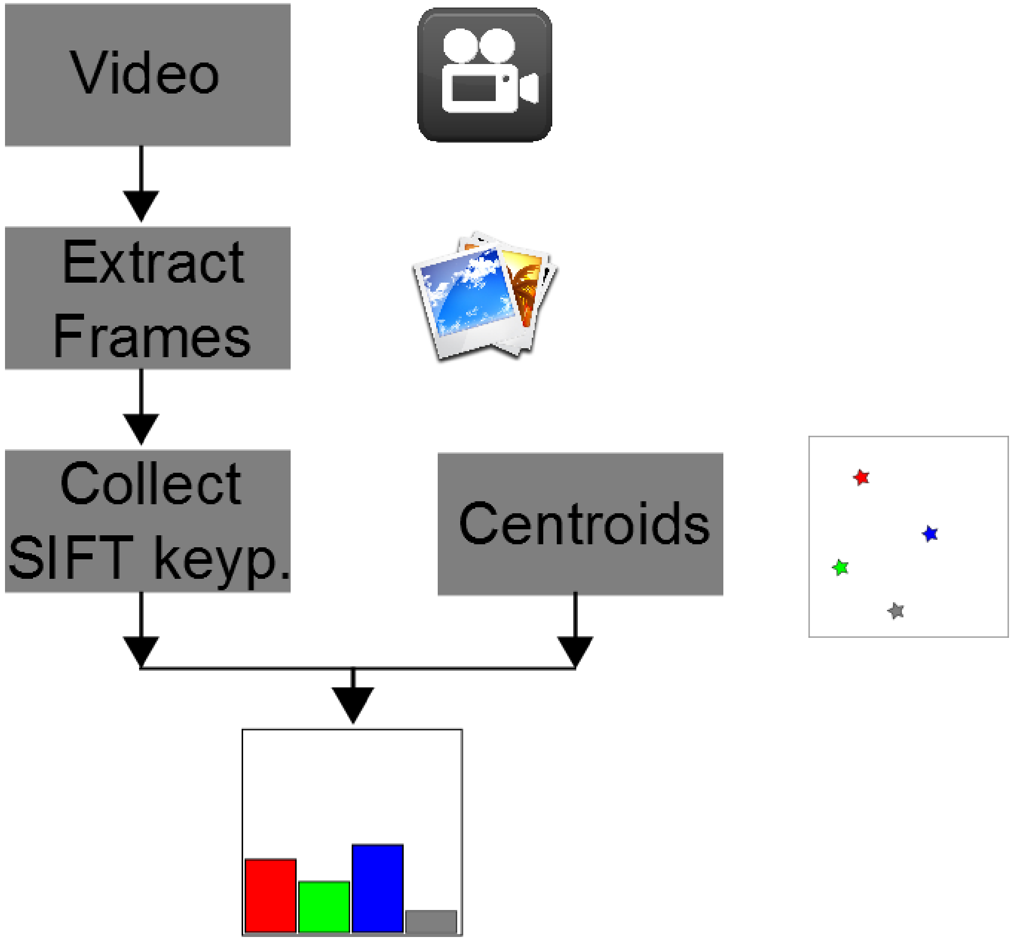
k-Means Clustering

Centroids

Training Signatures

- Which frames to select?
  - ✓ Choose certain frames at certain intervals
- Compute SIFT key points
- Quantize keypoints to Visual Words
  - ✓ Compute the distances to each centroid and pick the closest one
- Signature of a Histogram of Visual Words, e.g. VW 1 occurred 5 times





Video

Extract Frames

Collect SIFT keyp.

Centroids

How does SIFT work?

- Candidate keypoints are found by searching extrema in the Gaussian-scale space
- Further refinement
- Compute orientation and magnitude
- Combine to keypoint descriptor

“ D. G. Lowe, 2004, Distinctive Image Features from Scale-Invariant Keypoints, Int. Journal of Computer Vision “

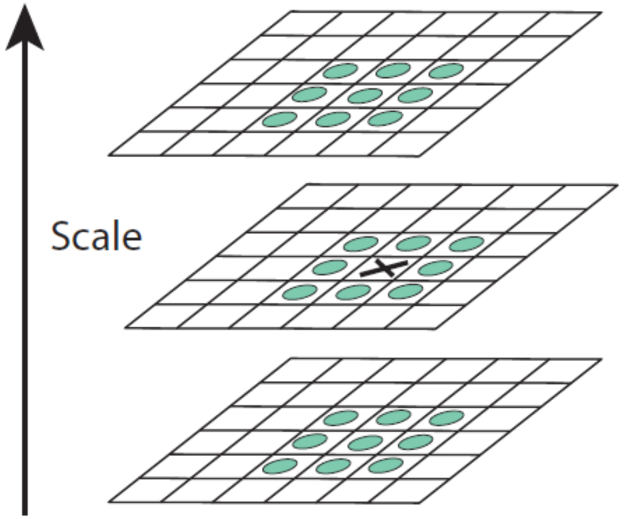
SIFT keypoint discovery

Extrema

- Use a DoG approximation for faster computation

$$\underbrace{G(x,y,k\sigma) - G(x,y,\sigma)}_{\text{DoG}} \approx (k-1)\underbrace{\sigma^2\nabla^2 G}_{\text{LoG}}$$

- Give the current and two neighboring scales, select a point if all 26 neighbors have smaller or greater value

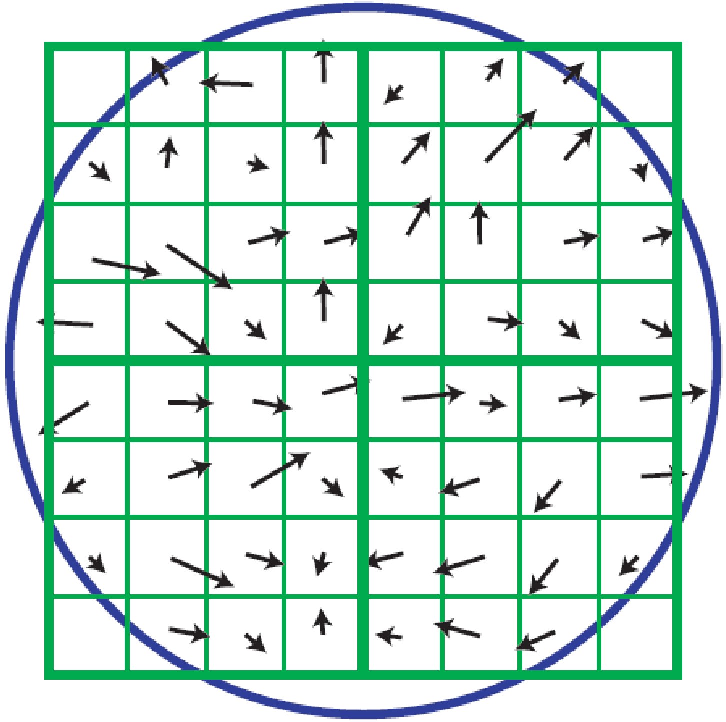


Orientation

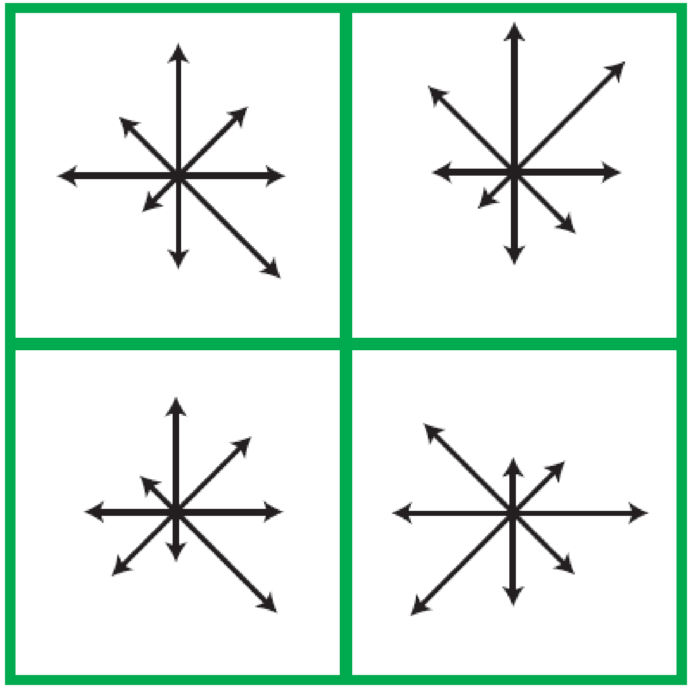
$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right)$$

SIFT keypoint discovery



The orientations in an area around the keypoint



Binning - Creating a histogram from the orientations in the left figure

Note: In his paper, Lowe recommends 4x4x8 histograms instead of 2x2x8 ones.

Evaluation

DataSet 1: True Blood – Season 5 (12 episodes, real-world data)

DataSet 2: CC_WEB_VIDEO: Near-Duplicate Web Video Dataset

Evaluate with k=100 and k=1000 visual words respectively

Evaluate against 9 attack types (logo, subtitles, crop, etc)

Test each video for each attack (Null-Attack is naturally 100%)

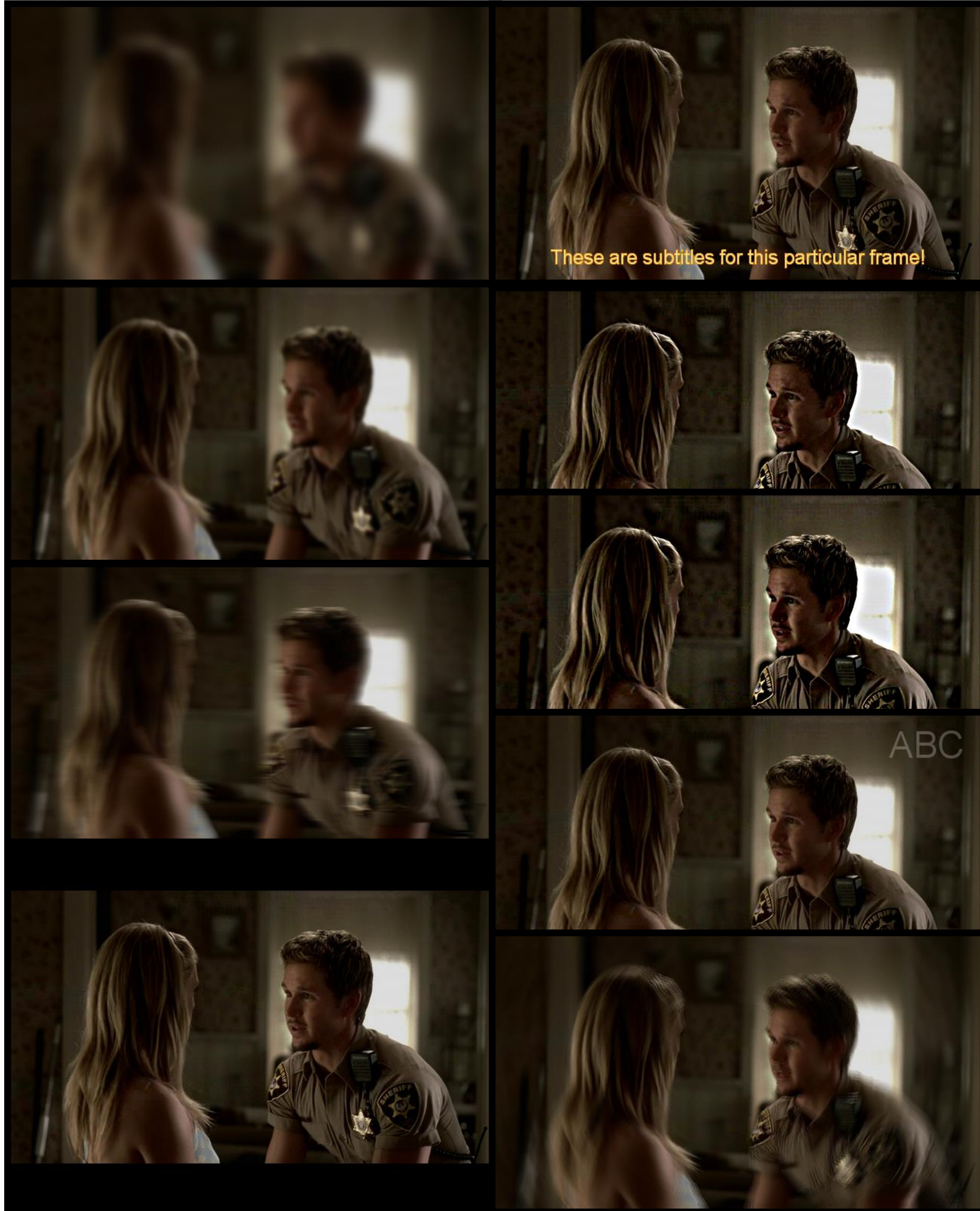
Performance (on Intel Core 2 Duo 2.54 GHz)

• Generate Centroids	43:30 min
• Train Signatures	02:02 min
• Query	9 sec

Test Results

Attack Type	Accuracy k=100	Accuracy k=1000
Heavy Blur	8 %	8 %
Light Blur	38 %	75 %
Motion Blur	8 %	8 %
Radial Blur	8 %	8 %
Crop	100 %	100 %
Logo	100 %	100 %
Heavy Sharpen	33 %	92 %
Light Sharpen	42 %	100 %
Subtitles	96 %	100 %

k … size of visual word vocabulary



Future Works

- More evaluation with real world data required
    - Detection of the same movie with slightly different length
    - Detection of the same movie in different codecs (and key frame intervals)
    - Bigger data sets for testing real world attacks

- Distributed collection of vocabulary and signatures

- Fusion of Features (e.g. SIFT features + Color Histogram)

THE END

謝謝

Questions?

Visit us at http://vash.in