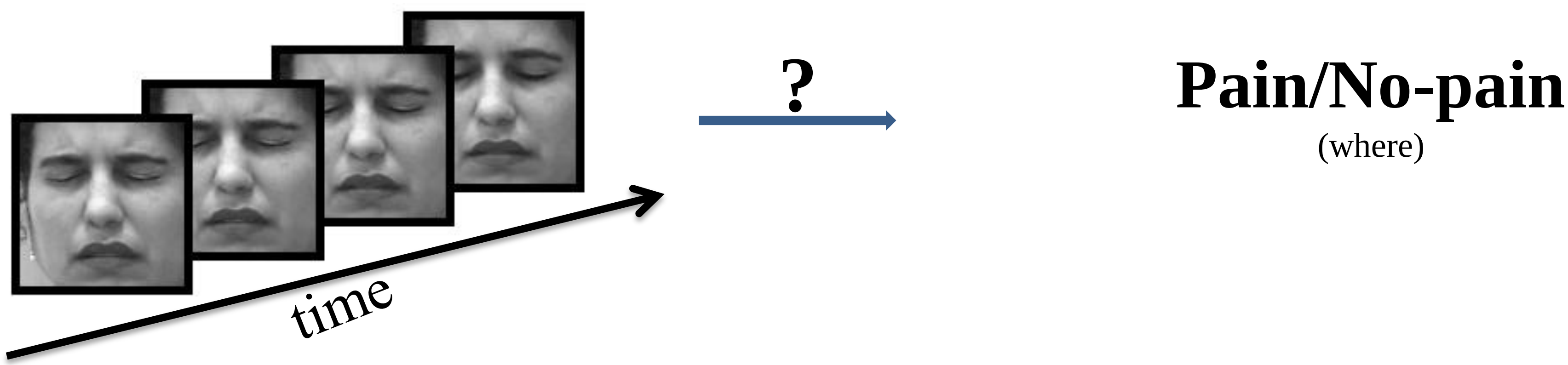
Weakly Supervised Pain Localization using Multiple Instance Learning



Karan Sikka[1], Abhinav Dhall[2], and Marian S. Bartlett[1]

[1]Machine Perception Lab
University of California, San Diego

[2]Australian National University

FG 2013

Motivation

- Pain monitoring critical for clinical applications.

- Spontaneous expression.
  - Classification difficult compared to posed expressions (CK+ dataset).

- Pain has high variability (expression, perception, location and duration)
  - Efficient prediction algorithms.


Frame 30
MCL 2.7.1

Problem Definition

- Subjects undergoing shoulder pain in videos.
  - UNBC MC-Master Pain Dataset*.
  - Ranging from 60-600 frames.

- Classifying and localizing pain in videos.
  - Sequence level ground-truth labels.

Lucey et. al., PAINFUL DATA: The UNBC-McMaster Shoulder Pain Expression Archive Database, FG'11

Challenges

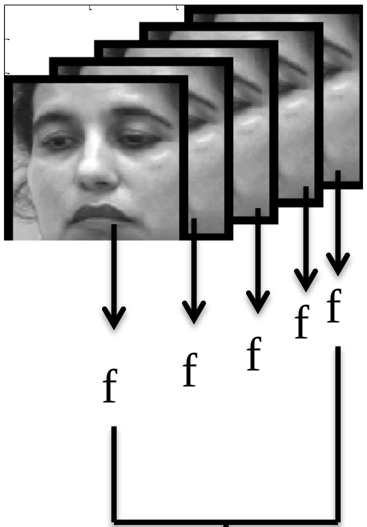
- Ambiguity introduced by **sequence level labels**.
  - Time points and duration of pain unknown apriori.

- Incorporating dynamics/temporal information.

- Temporal segmentation is hard in itself.

Previous Approaches

‘Classical’ Fixed Length Features



Pooling

SVM

f ~ Features for each frame

1. BoW using Local Features.
2. Active Appearance (AAM) based Texture and Shape features.
3. Gabor
5. LBP

Pooling

- Avg or max
- Fixed Length Features

*Laptev et. al., Learning Realistic Human Actions From Movies, CVPR’08

Previous Approaches

‘Classical’ Fixed Length Features

- Most common approach.

- Works well when action spans whole videos
  - Facial expression classification (CK+ dataset).
  - Action classification (KTH dataset).

- Pooling features will not work well for long videos.
  - Kills the signal of interest.
  - Localized instead of global approaches required.

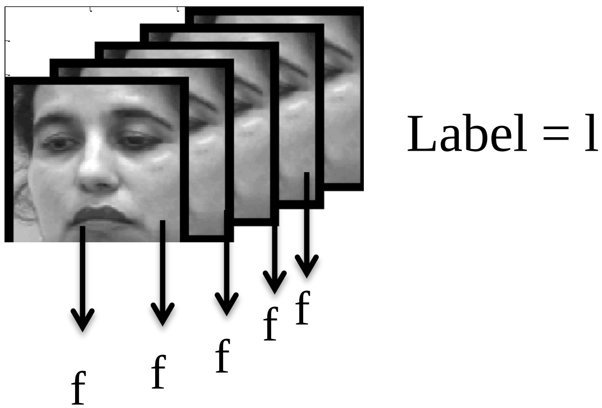
Previous Approaches
'Classical' Fixed Length Features

- Most intuitive approach.

- Works well when action spans whole videos
  - Facial expression classification (CK+ dataset).
  - Action classification (KTH and hollywood dataset).

- Pooling features doesn't work well in all cases.
  - Kills the signal of interest.
  - Localized instead of global approaches required.

Previous Approaches
Frame Level Features

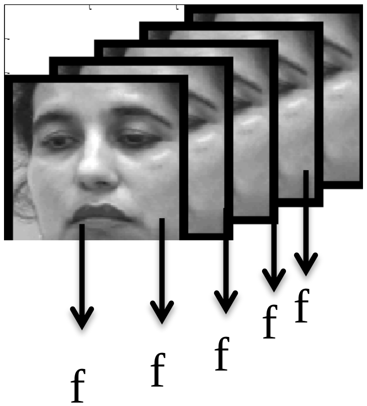


- AAM ➡ Clustering ➡ SVM
- Assign labels of sequence to each frames.
- Test
  - $Score(video) = Avg(Output(frames))$.

Ashraf *et.al*, ICMI'08.

Previous Approaches

Frame Level Features



Label = 1

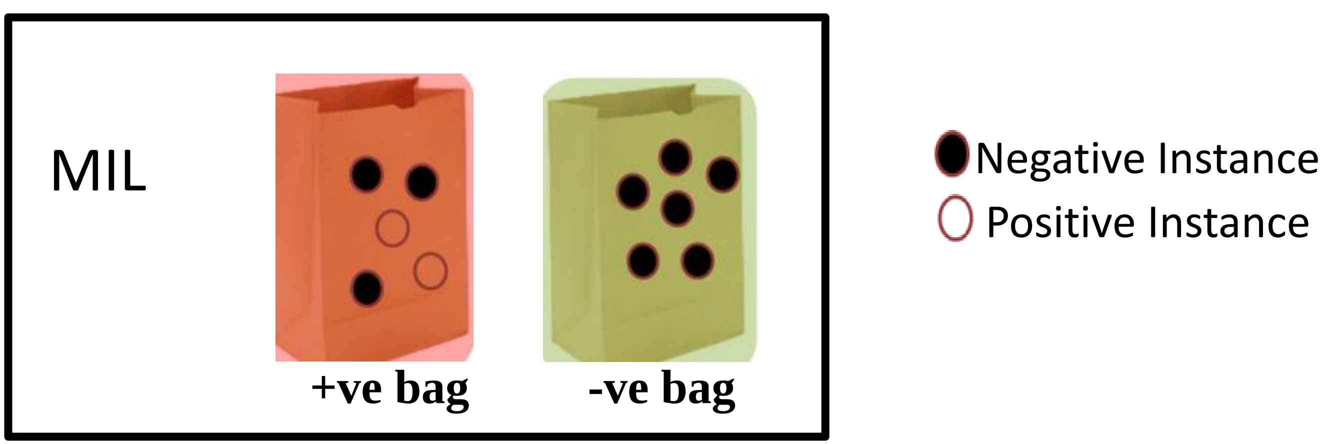
- AAM ➡ DCT ➡ SVM
- Assign labels of sequence to each frames.
- Test
  - $Score(video) = Avg(Output(frames))$.

Lucey *et.al*, ICASP'08.

Previous Approaches
Limitations

1. Assigning sequence label to each frame.
   - Label Ambiguity.
   - ML methods like SVM not robust to outliers.

- Solution: Multiple Instance Learning (MIL).
  - Efficiently handle weakly labeled data.



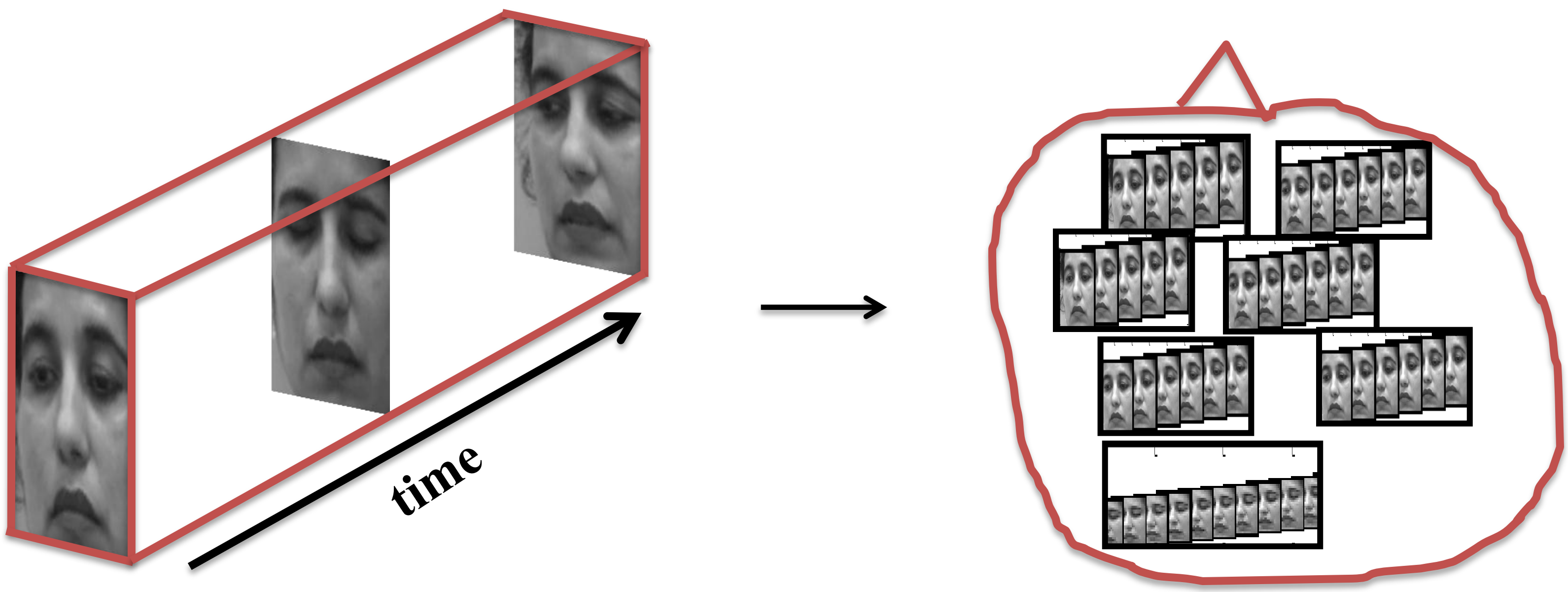
Previous Approaches

Limitations

2. Treated videos as individual frames.
   - Lack of temporal information.
   - Vital for pain classification.

- Solution: Represent sequences as sets of frames: "Multiple segments"

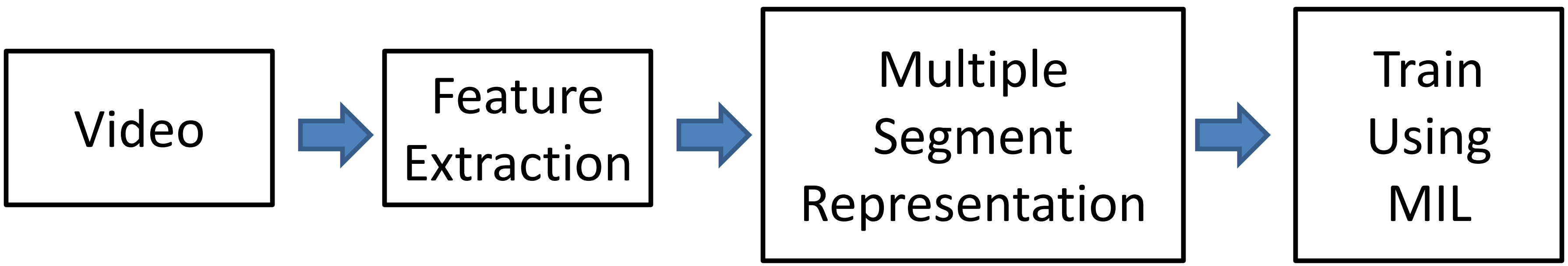
Multiple Segment Representation



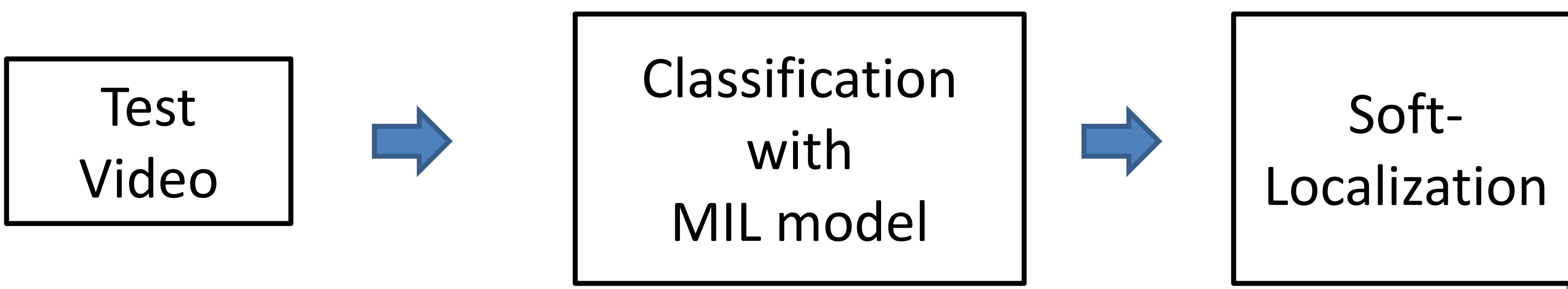
- Extracting at multiple scales and can overlap (no-restriction).
- Allow multiple hypothesis.

Multiple Segments based Multiple Instance Learning (MS-MIL)

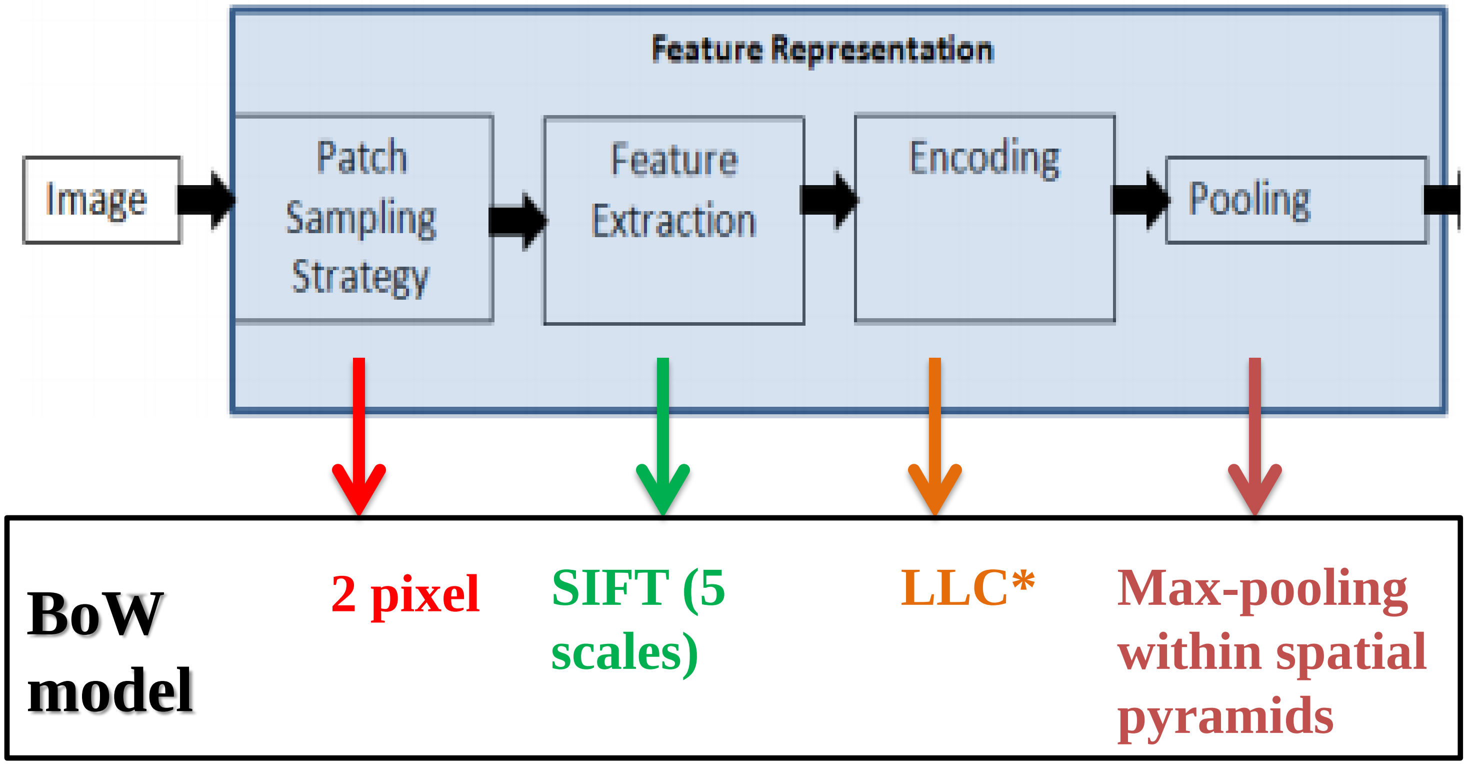
Train



Test



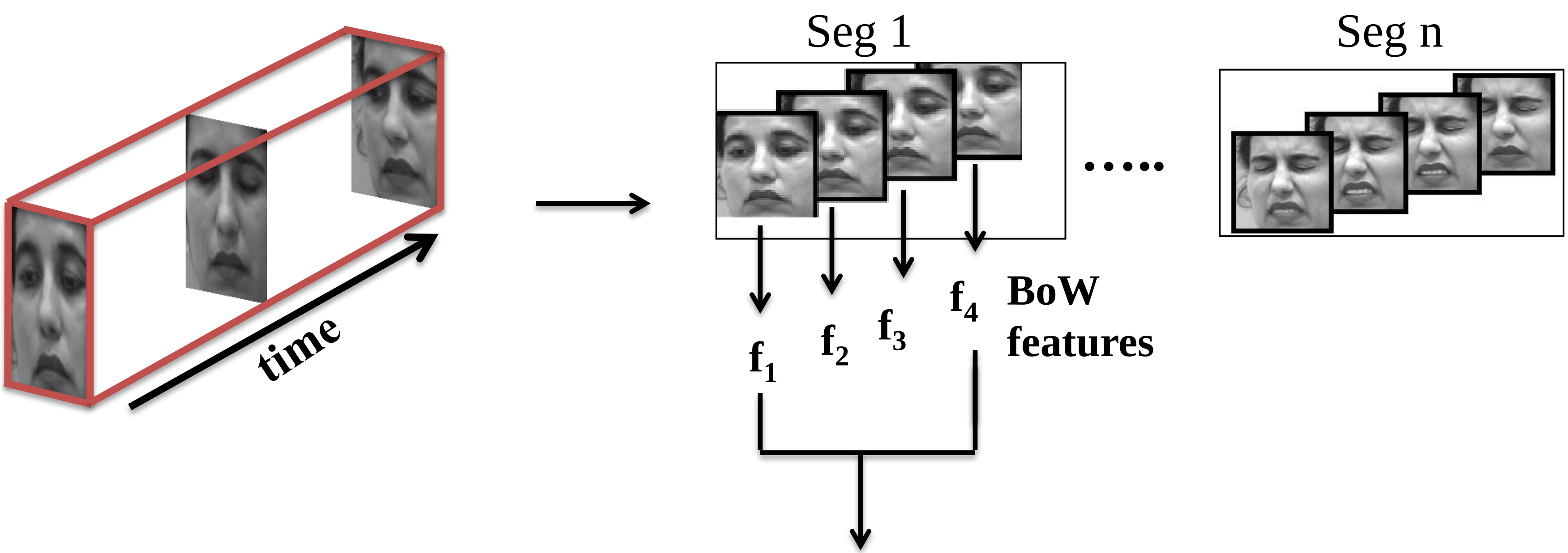
Feature Extraction - Frames



*LLC- Locality constrained Linear Encoding

Sikka *et.al*, ECCV'12.

Multiple Segment Representation



Pooling

f_1=[.................................]

f_2=[.................................]

f_3=[.................................]

f_4=[.................................]

max

F=[.................................]

Max pooling summarizes sparse signals better than average pooling.

MS-MIL

Train

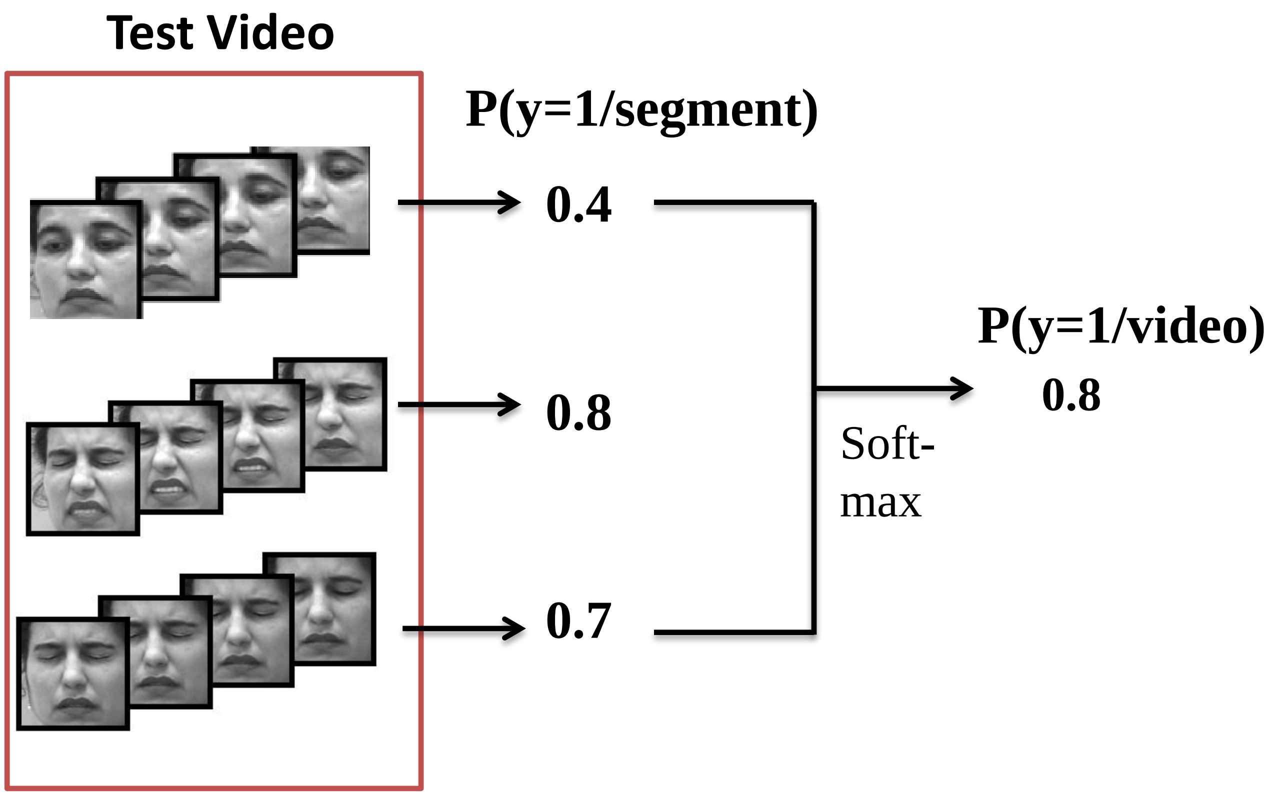
Training Set

Train using
MilBoost
(Viola'06)

- Training videos as Bags

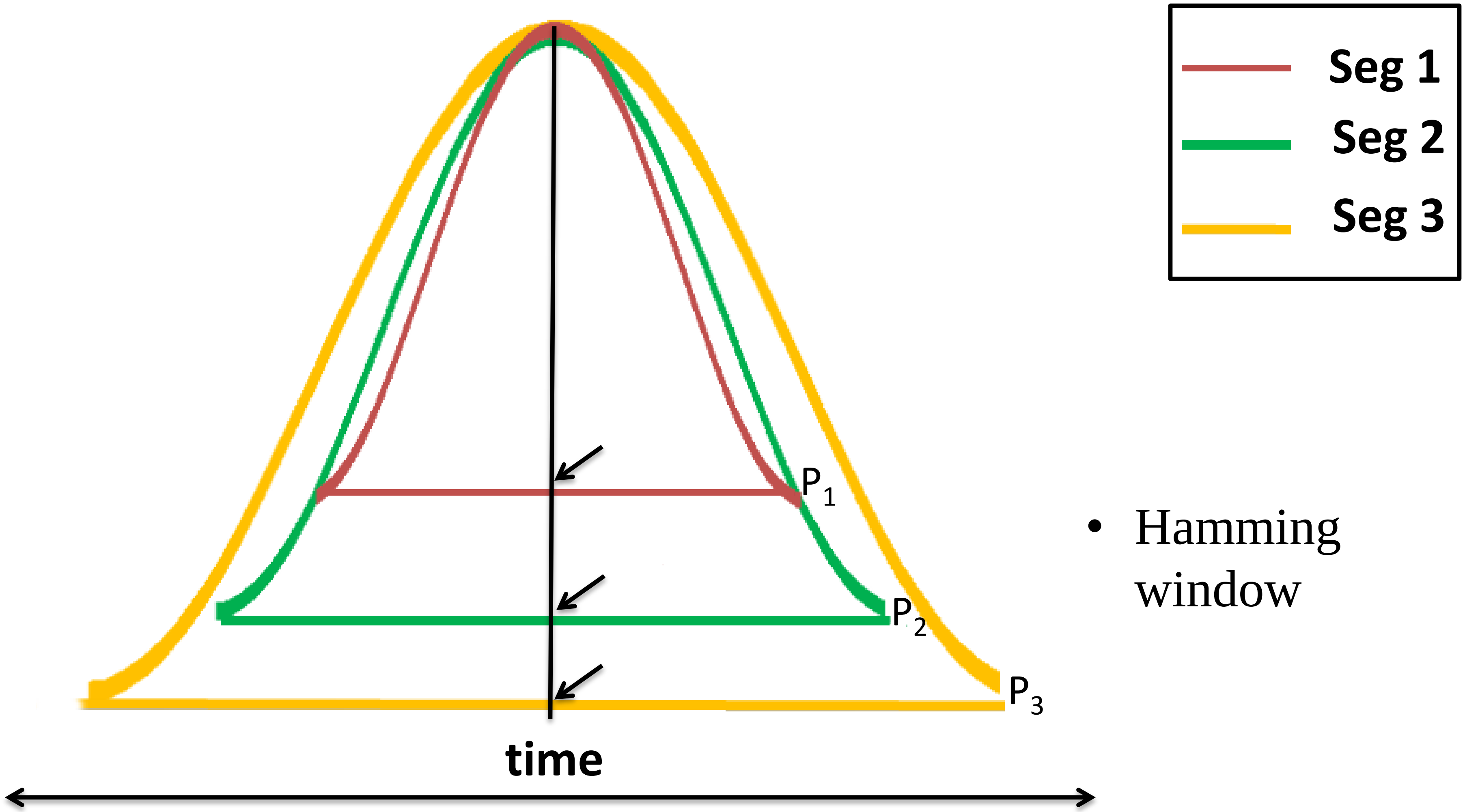
MS-MIL

Test



- MIL has a joint optimization framework.

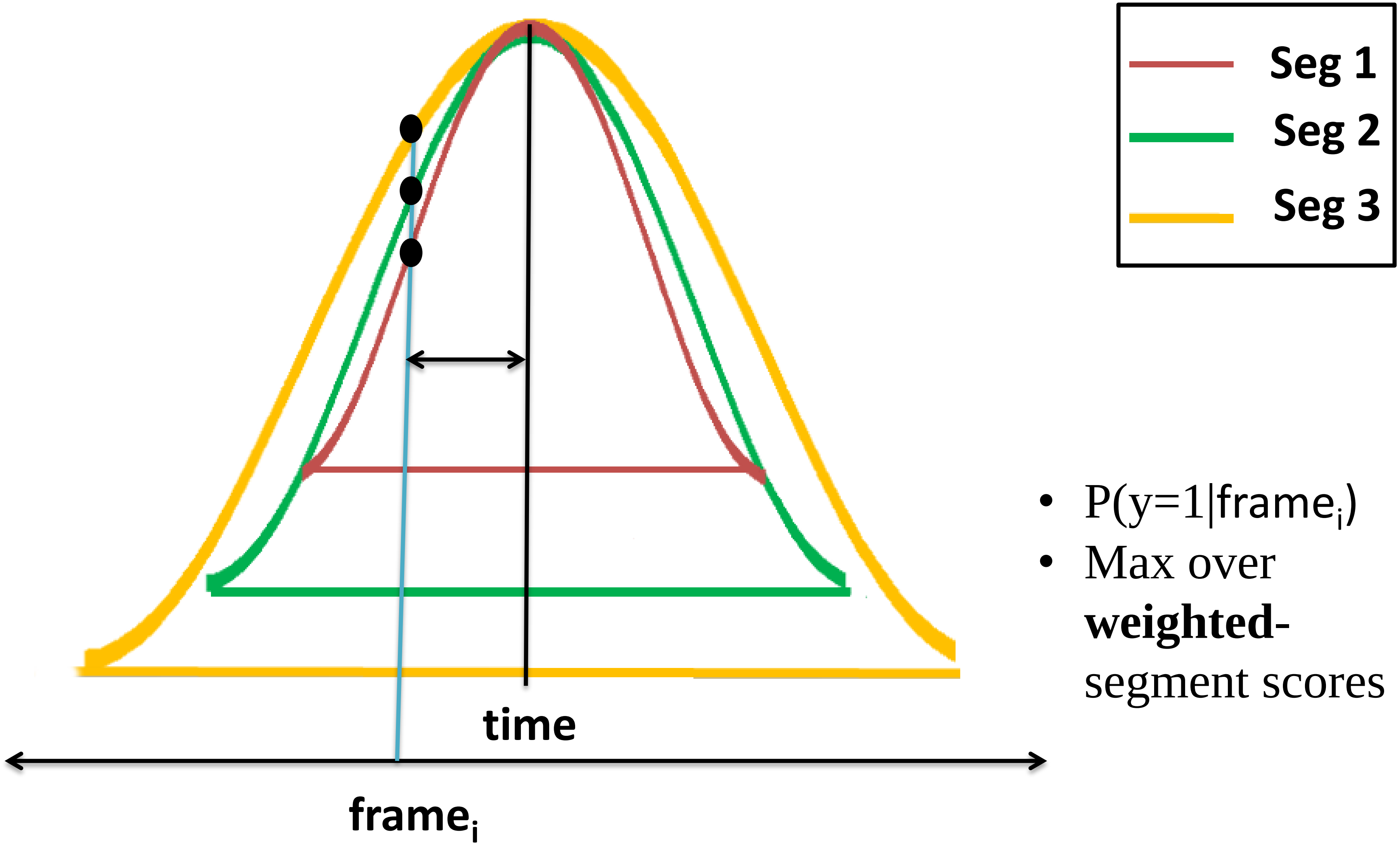
MS-MIL
Localization



- Hamming window

MS-MIL

Localization



- P(y=1|frame$_i$)
- Max over **weighted-** segment scores

Experiments

- Leave one subject out protocal.
- 147 videos from 23 subjects.
- Observer Pain Intensity as ground-truth labels.
  - Binarized.
- Faces aligned using provided AAM features.
- Total classification rate at Equal Error Rate.

Classification Performance

Method	Accuracy	#Subjects	#Samples
MS-MIL	83.7	23	147
Lucey et.al	80.99	20	142
Ashraf et.al (as shown in Lucey at.al)	68.31	20	142
ML-SVM$_{max}$	70.75	23	147
ML-SVM$_{avg}$	76.19	23	147

- Shows gains over previous methods.

Classification Performance

Method	Accuracy	#Subjects	#Samples
MS-MIL	83.7	23	147
Lucey et.al	80.99	20	142
Ashraf et.al (as shown in Lucey at.al)	68.31	20	142
MS-SVM_{avg}	70.75	23	147
MS-SVM_{max}	76.19	23	147

- MS-SVM
  - Each segment assigned the label of the video
  - SVM + score combining rule (max and avg).

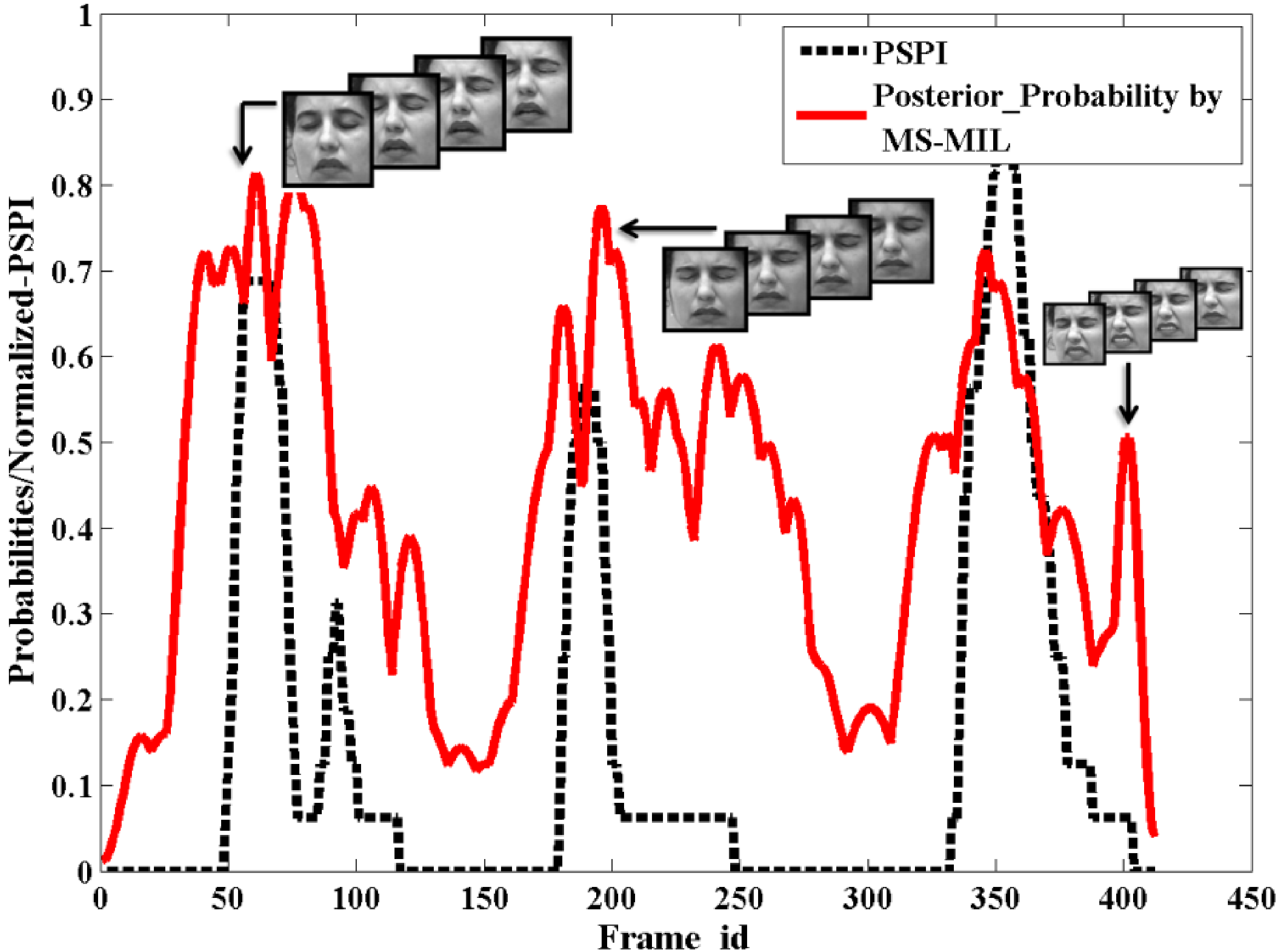
Classification Performance

Method	Accuracy	#Subjects	#Samples
MS-MIL	83.7	23	147
Lucey et.al	80.99	20	142
Ashraf et.al (as shown in Lucey at.al)	68.31	20	142
MS-SVM_{avg}	70.75	23	147
MS-SVM_{max}	76.19	23	147

- MS-MIL performs better than it's traditional ML counterparts.

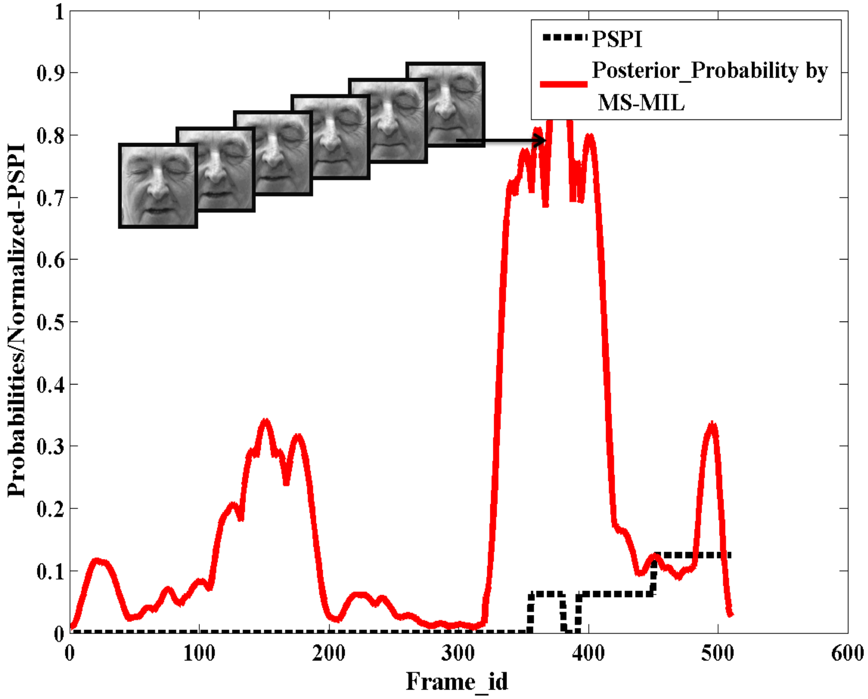
Localization Performance

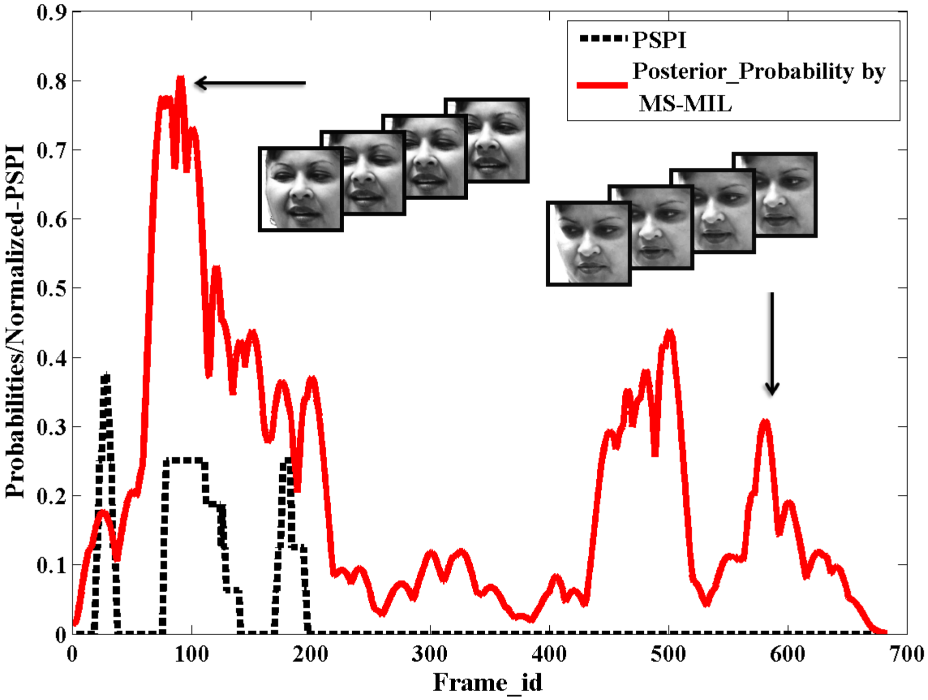
- Compared per-frame probabilities predicted by MS-MIL with human expert pain labels.
- PSPI computes pain intensity based on FACS.
  - PSPI sums intensities of 4 Action Units.
  - Prkachin & Solomon’08.
- Normalized PSPI to 0-1.

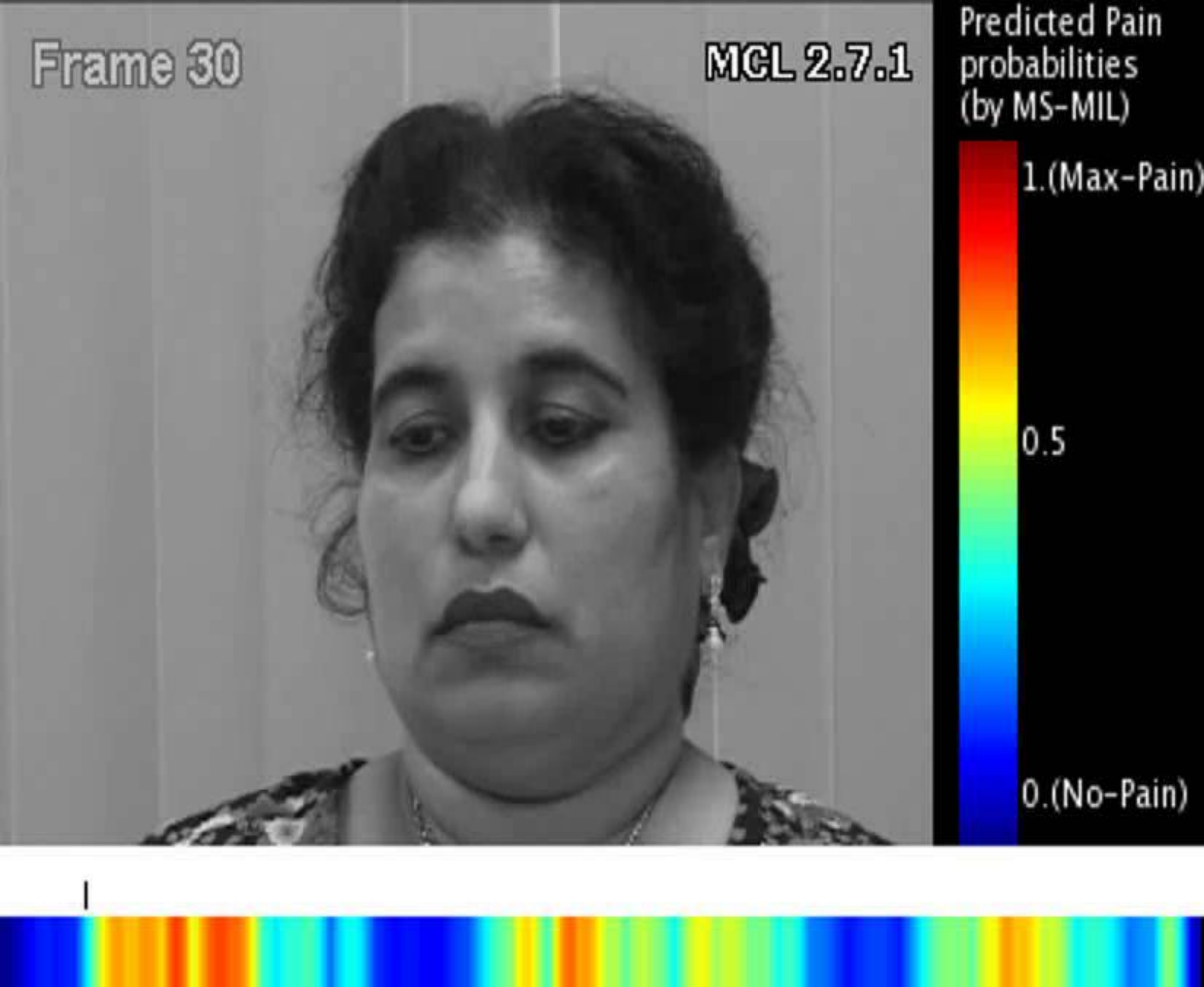


PSPI
Posterior_Probability by MS-MIL
Probabilities/Normalized-PSPI
Frame_id
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
50
100
150
200
250
300
350
400
450

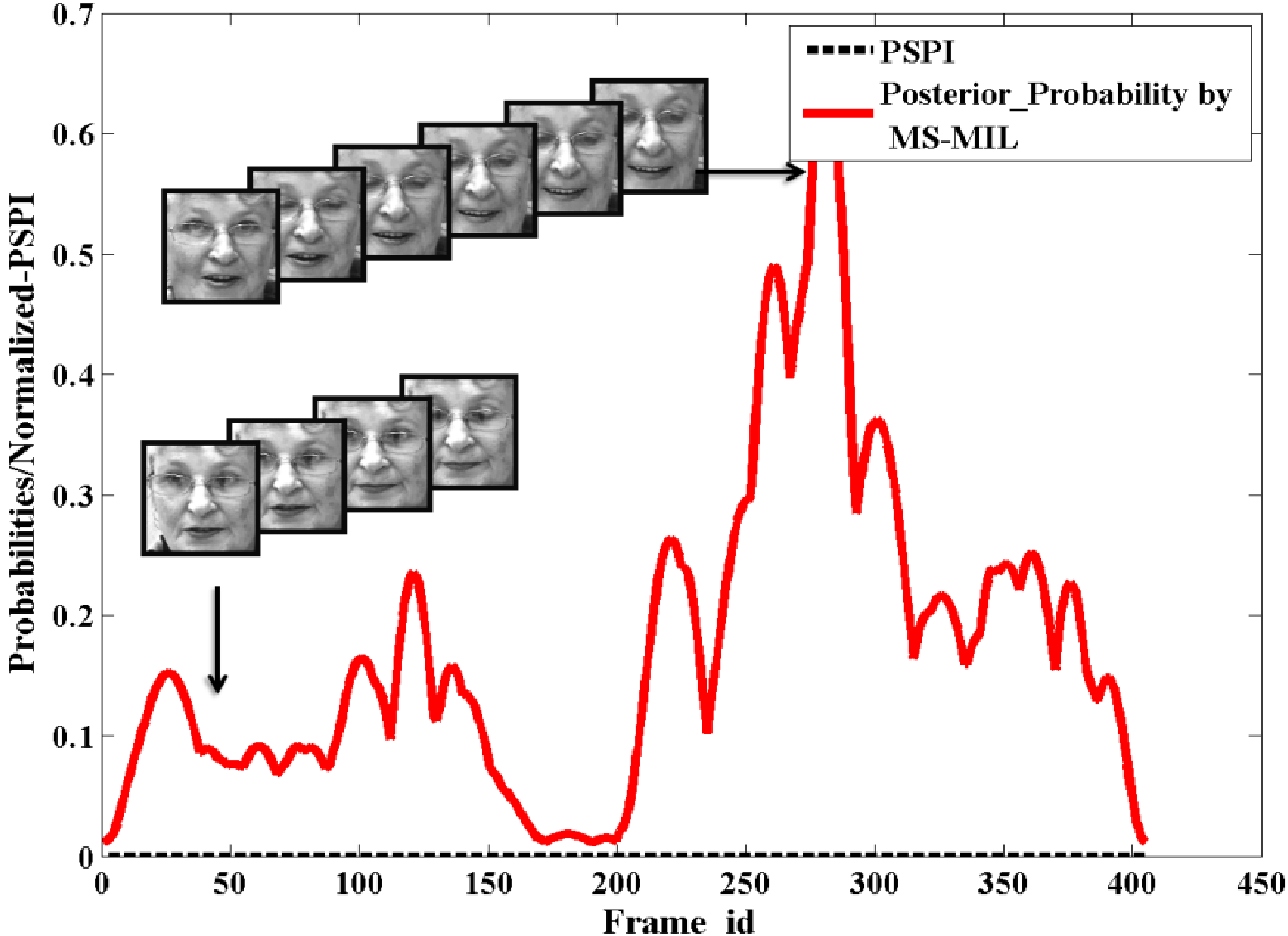
Localization Performance



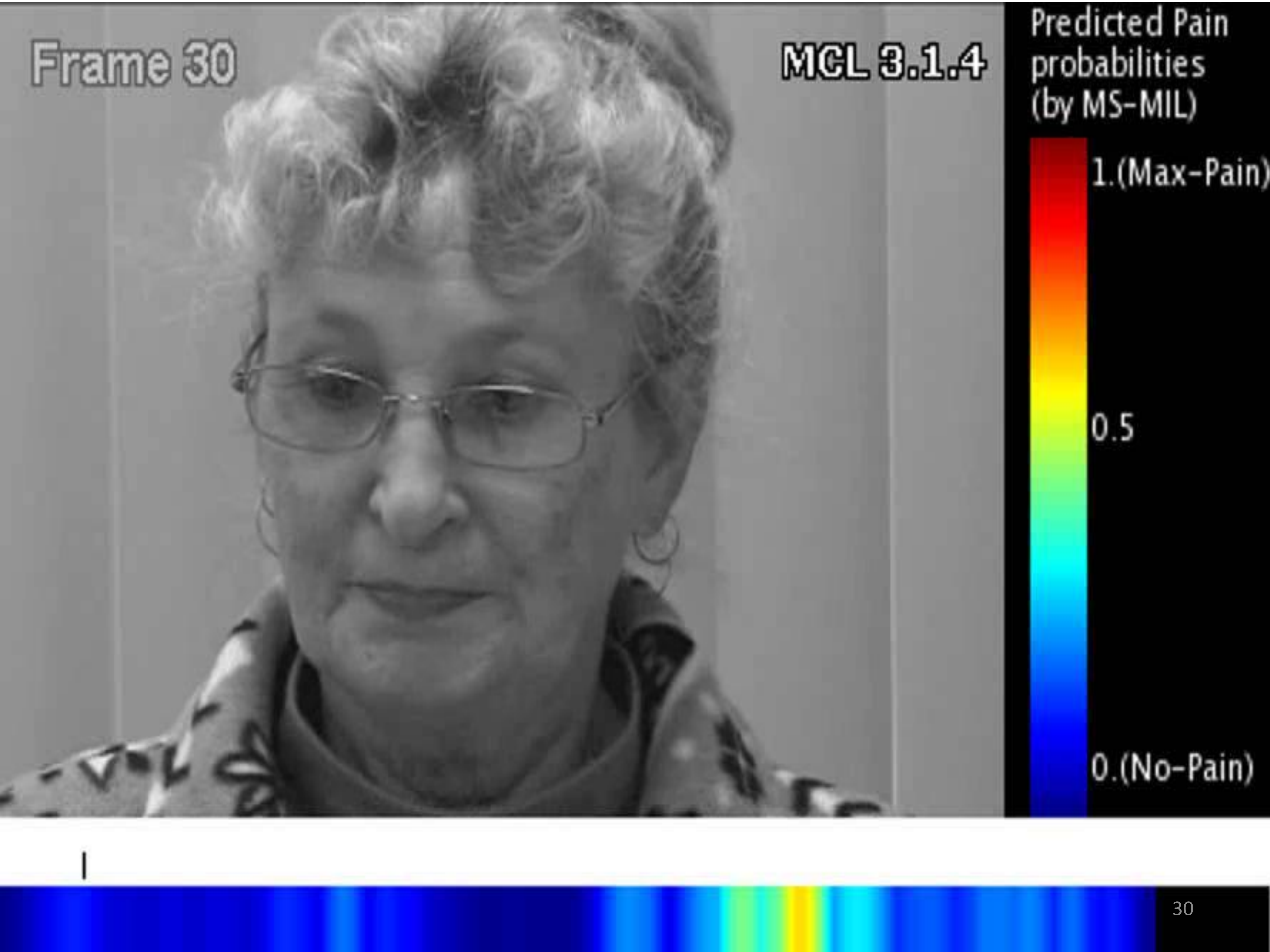




Frame 30
MCL 2.7.1
Predicted Pain probabilities (by MS-MIL)
1.(Max-Pain)
0.5
0.(No-Pain)



PSPI
Posterior_Probability by MS-MIL
Probabilities/Normalized-PSPI
Frame_id
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
50
100
150
200
250
300
350
400
450


Frame 30
MCL 3.1.4
Predicted Pain probabilities (by MS-MIL)
1.(Max-Pain)
0.5
0.(No-Pain)

Conclusion

- Proposed Novel approach to problem of classifying and localizing pain.

- Highlighted limitations of previous approaches and motivations for current algorithm.

- Compared MS-MIL with
  - Previous Approaches
  - Traditional ML counterparts.

- Localization compared with ground-truth index (PSPI).

Questions?

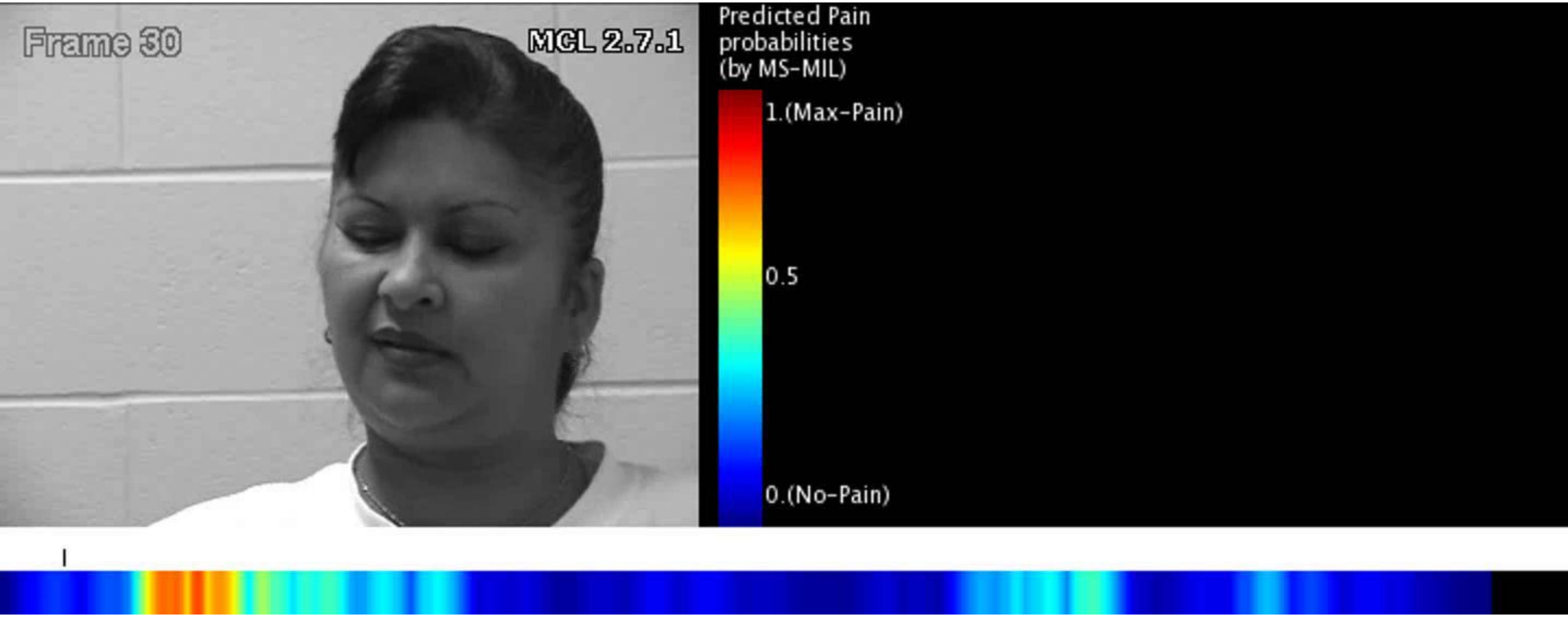
Karan Sikka

Abhinav Dhall

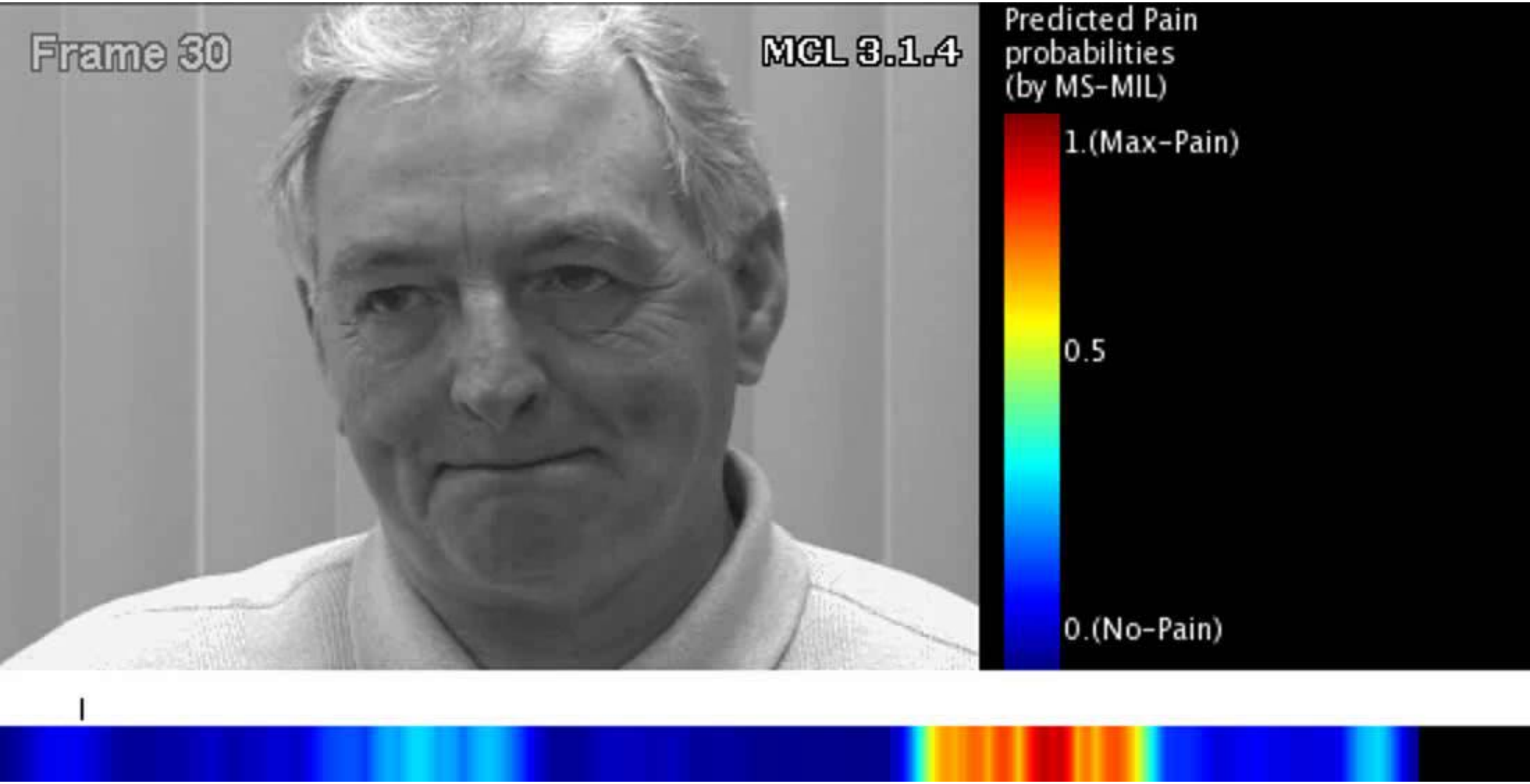
Dr. Marian S. Bartlett

Machine Perception Lab, UCSD

Thanks



Frame 30
MCL 2.7.1
Predicted Pain probabilities (by MS-MIL)
1.(Max-Pain)
0.5
0.(No-Pain)


Frame 30
MCL 3.1.4
Predicted Pain probabilities (by MS-MIL)
1.(Max-Pain)
0.5
0.(No-Pain)