Tutorial

ECCV 2020

Recent Advances and Challenges in Facial Micro-Expression Analysis

ME Spotting

John See Multimedia University, Malaysia

Moi-Hoon Yap Manchester Metropolitan University, UK

Su-Jing Wang Chinese Academy of Sciences, China

Jingting Li Chinese Academy of Sciences, China

Sze-Teng Liong Feng Chia University, Taiwan

Outline

ME spotting pipeline

Pre-processing steps

- Facial Landmark Detection & Tracking
- Face Registration
- Optional steps (Masking, Region division)

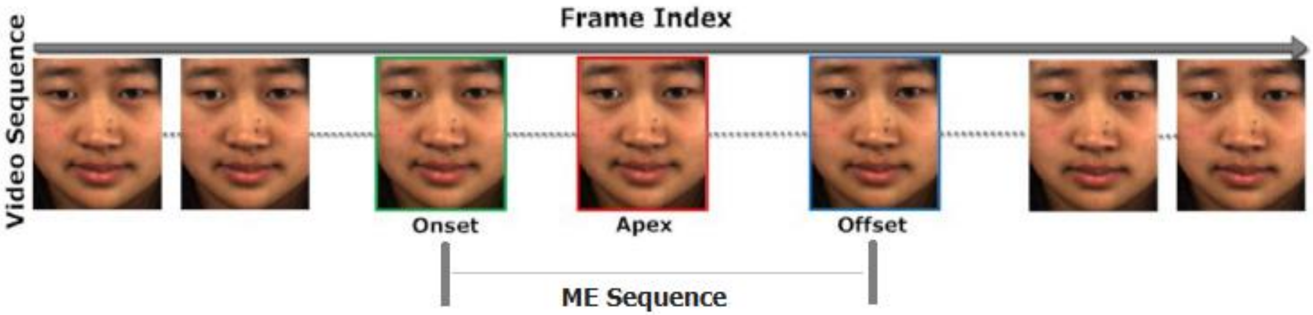
Approaches

- Early attempts with posed data
- ME sequence spotting
  - Feature difference (FD) analysis
  - **Highlighted Work:** Recent Approaches on Spotting ME Sequences (Jingting Li)
- ME apex spotting
- **Highlighted Work:** Automatic Apex Frame Spotting in Micro-Expression Database (Sze-Teng Liong)

ME spotting

- **Spotting**
  - Automatic detection of the temporal interval of a micro facial movement in a sequence of video frames

- **Two Current Flavours:**
  - Spotting ME sequence or window of occurrence
  - Spotting ME apex frame

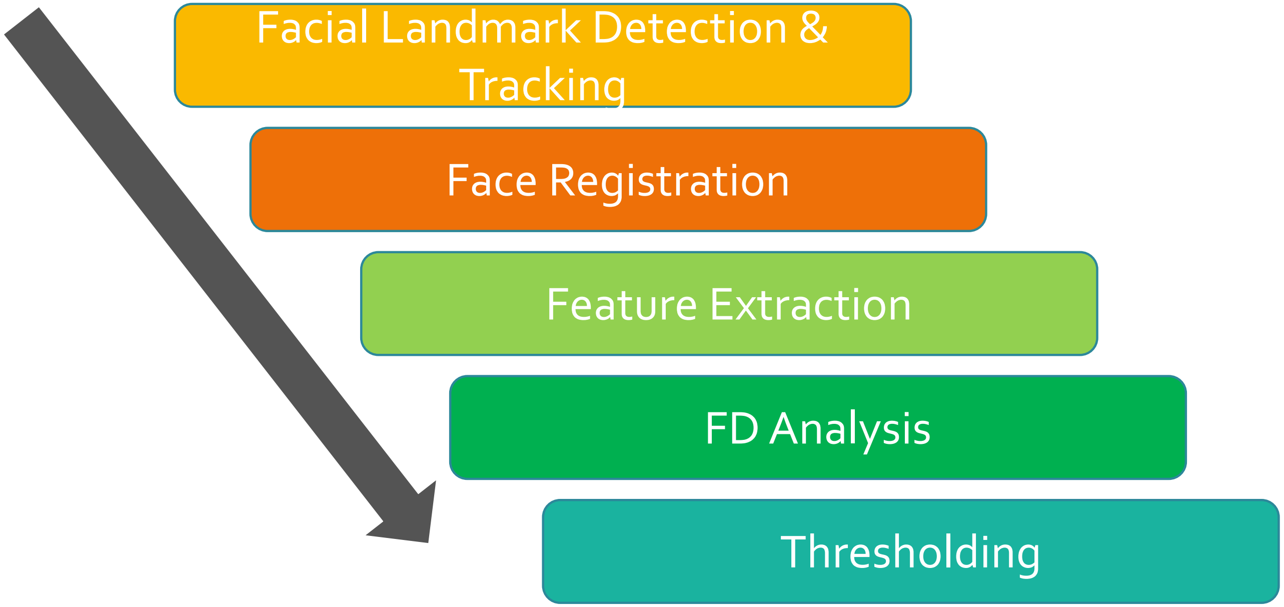


Early Attempts on Posed ME Data

- **Polikovsky et al. (2013)**
  - 3D gradient histograms as descriptor to distinguish onset, offset, apex, neutral
  - Drawbacks:
    - Used posed data which is not challenging and unnatural
    - Treat spotting as a classification problem!
- **Shreve et al. (2013)**
  - Optical strain method to spot macro- and micro-expressions
  - Reported good results (77/96 spotted) on their small, unpublished posed dataset

ME Sequence Spotting Pipeline

Typically[†], a ME sequence spotting process will follow these steps:



Facial Landmark Detection & Tracking

Face Registration

Feature Extraction

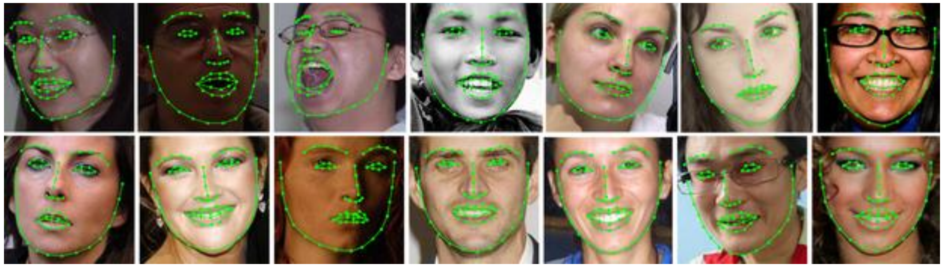
FD Analysis

Thresholding

[†] A majority of spotting works use this pipeline. There are a few works that treat spotting as a classification problem; hence they may have a pipeline that is more similar to one of recognition.

Facial Landmark Detection & Tracking

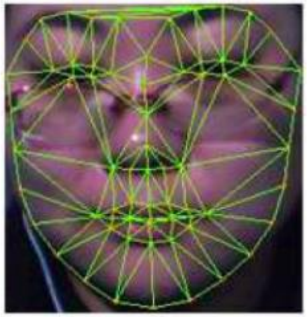
- **Landmark Detection**
  - Some early works manually annotate the first frame with facial landmarks, and proceed to track (Polikovsky et al., 2013)
  - To automate this process, later works apply automatic facial landmark detection:
    - Active Shape Model (ASM)
    - Discriminative Response Map Fitting (DRMF)
    - Constraint Local Model (CLM)



iBUG group, Imperial College London

- **Tracking**
  - Kanade-Lucas-Tomasi (KLT) algorithm
  - Auxiliary Particle Filtering (APF) algorithm

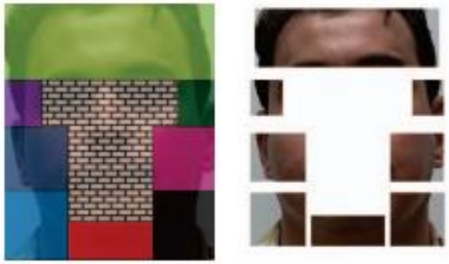
Face Registration

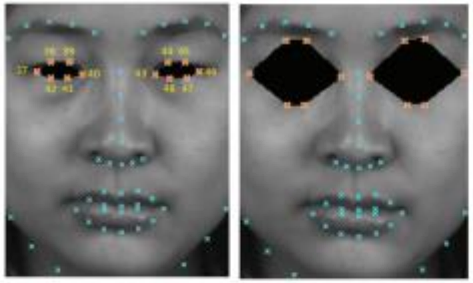


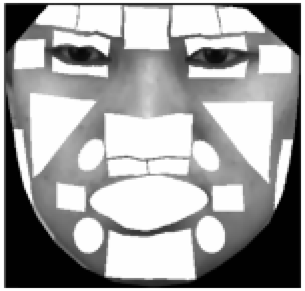
- **Image Registration**
  - A process of aligning two images (the reference and sensed images) in a geometrical manner.
  - In ME → Useful to remove large head translations and rotations that might affect spotting or recognition task.
- 2 major categories of approaches used by ME works:
  - **Area-based (a.k.a. template matching or correlation-like)** – windows of predefined size / entire image utilized to estimate correspondence between images. → **2D-DFT** used by Davison et al. (2016)
  - **Feature-based** – features from local structures (points, lines, regions) are used to find pairwise correspondence between images
    - **Simple affine transform** used by Shreve et al. (2011), Moilanen et al. (2014)
    - **Local Weighted Mean (LWM)** used by Li et al. (2017), Xu et al. (2017) seeks to find 2D transformation matrix using 68 landmark points of the face

(Optional) Masking

- A **masking** step can be useful to remove noise caused by undesired facial motions
  - **Shreve et al. (2011)** – used a "T-shaped" static mask to remove middle portions of the face and the eyes
  - **Liong et al. (2016)** – used eye regions extracted based on facial landmarks to reduce false spotting of eye blinking motion
  - **Davison et al. (2016)** – used a binary mask to obtain 26 FACS-based facial regions, which were representative of locations of the face containing a single or a group of AUs





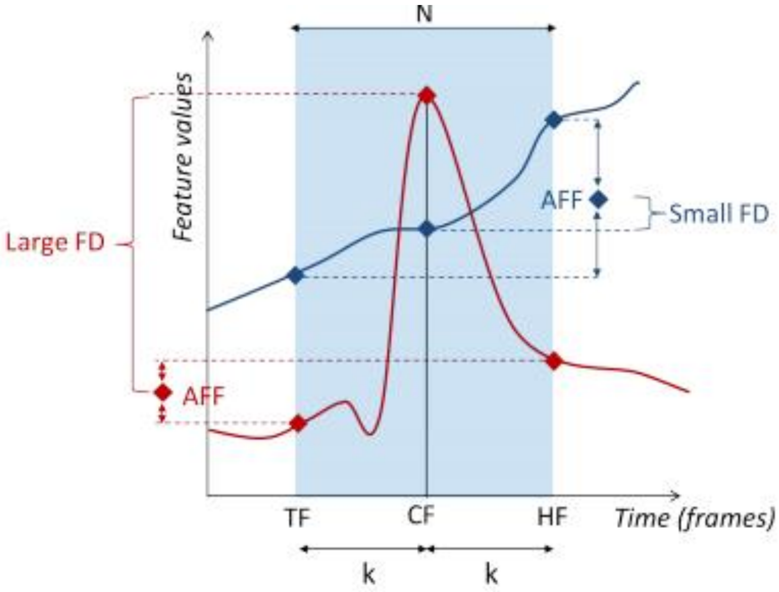


(Optional) Region Division

- Psychological findings (Porter & ten Brinke, 2008): ME analysis should be done on the upper and lower halves separately instead of together
- **Region division** encourages splitting the face into important separate segments to achieve “localized spotting”
  - **Ad-hoc ROI segments:**
    - 3 regions (upper, middle, lower); **(Shreve et al., 2009)**
    - 8 regions (forehead, left & right of eye, left & right of cheek, left & right of mouth, chin) **(Shreve et al., 2011)**
    - 4 quadrants **(Shreve et al., 2014)**
    - FACS action unit (AU) regions **(popular in many works)**
  - **Block/grid segments:**
    - $m \times n$ blocks **(popular in many works)**
  - **Delaunay triangulated segments (Davison et al., 2016)**

Feature Difference (FD) Analysis

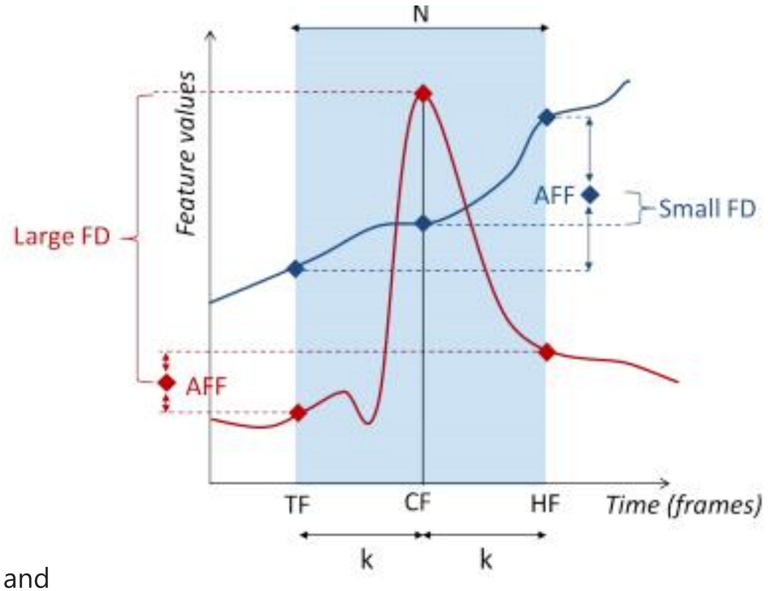
- **FD Analysis:** Compares differences of video frame features within a specified interval
- **Terminologies:**
  - CF: Current frame, N : Micro-interval
  - HF: Head frame (k-th frame after CF)
  - TF: Tail frame (k-th frame before CF)
  - $k = \frac{1}{2}(N - 1)$
  - AFF: Average feature frame, feature vector Representing the average of features TF and HF.



Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., & Pietikäinen, M. (2017). Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. IEEE Trans. Affective Computing, 9(4), 563-577.

Feature Difference (FD) Analysis

- **FD Analysis:**
  - Compare the features of CF against its AFF by computing the FD of the pair of feature histograms using χ^2 distance
  - Do this for all CFs, except for the first and last k frames
- **Intuition**:
  - Large FD ➔ A rapid facial movement, onset-offset occurs within the time window
  - Small FD ➔ A slower, gradual facial movement



Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., & Pietikäinen, M. (2017). Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. IEEE Trans. Affective Computing, 9(4), 563-577.

Feature Difference (FD) Analysis

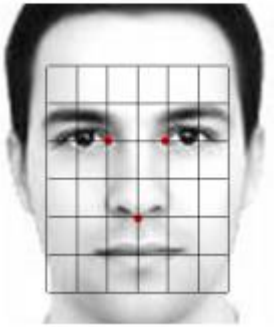
- **Localizing spotting:**
  - For each CF, FD values are computed for each block (e.g. 6x6 grid = 36 blocks)
  - Since the occurrence of an ME will result in larger FD values in some (but not all) blocks, we take the average of M largest FDs as the difference vector

$$F_i = \frac{1}{M} \sum_{\beta=1}^{M} d_{i,j_\beta}$$

  - The contrasted difference vector finds how far it is from the average F_i of TF and HF

$$C_i = F_i - \frac{1}{2}(F_{i+k} + F_{i-k})$$

- **Determine threshold**:
  - Threshold ➔ $T = C_{\text{mean}} + \tau \times (C_{\text{max}} - C_{\text{mean}})$
  - $\tau = [0, 1]$ is a parameter for obtaining different thresholds



Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., & Pietikäinen, M. (2017). Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. IEEE Trans. Affective Computing, 9(4), 563-577.

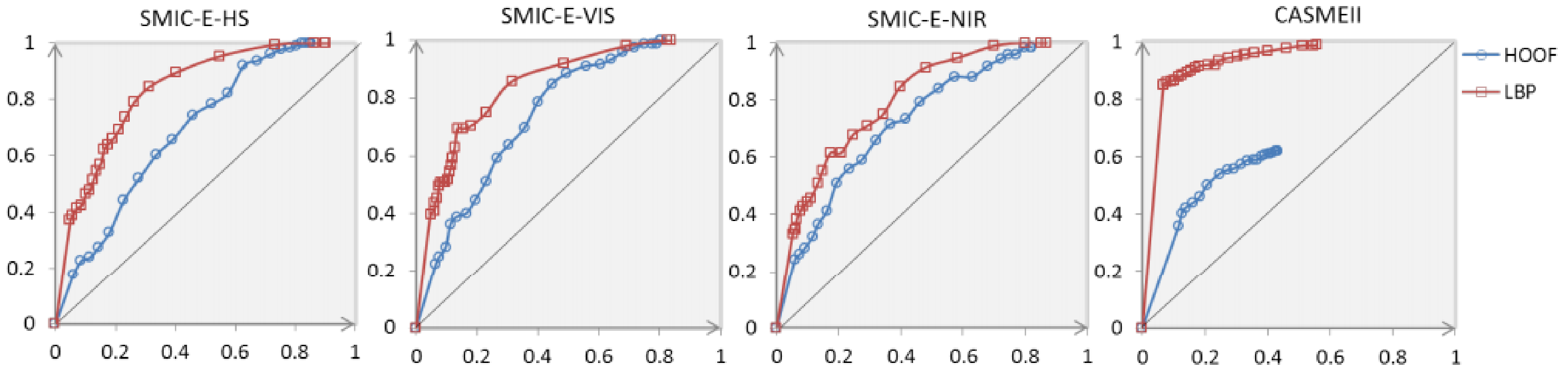
Feature Difference (FD) Analysis

- **Peak detection:**
  - Minimum peak distance for peak detection is set to $k/2$
  - Spotted peaks are compared with ground truth labels
    - If one spotted peak is located within the frame range of $\left[onset - \frac{N-1}{4}, offset + \frac{N-1}{4}\right]$, the frames in the spotted sequence are counted as true positives, otherwise the N frames will be counted as false positive frames.

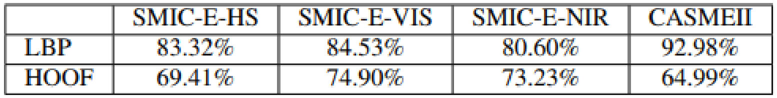
Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., & Pietikäinen, M. (2017). Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. IEEE Trans. Affective Computing, 9(4), 563-577.

Results from Li et al. (2017)

- **Spotting on CASME II and the SMIC datasets**

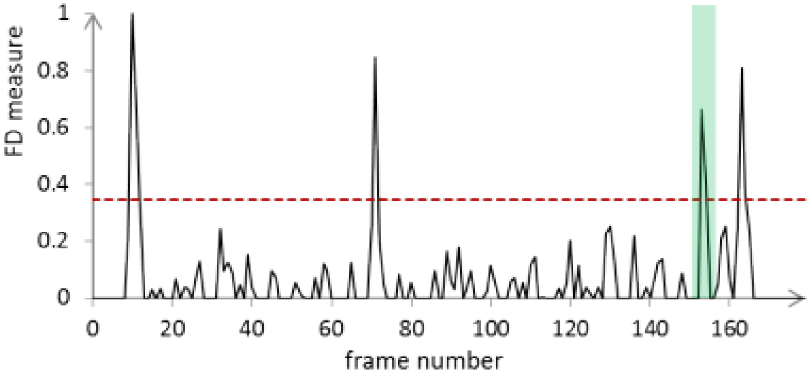


AUC values of the ME spotting experiments



	SMIC-E-HS	SMIC-E-VIS	SMIC-E-NIR	CASMEII
LBP	83.32%	84.53%	80.60%	92.98%
HOOF	69.41%	74.90%	73.23%	64.99%

- **Challenging scenarios**:
  - E.g. At the given threshold, one true ME spot, three false eye blink spots



Attempt: Exclude eye regions, eye-blink detector to exclude blinks…
FPR ↓ TPR ↓

Features for FD Analysis

- A majority of works that applied FD Analysis opted for different feature choices:

Feature	First Work to Use
LBP	Moilanen et al. (2014)
HOG	Davison et al. (2015)
MDMD	Wang et al. (2016)
3D HOG, Optical Flow (OF)	Davison et al. (2016)
HOOF	Li et al. (2017)
Riesz Pyramid	Duque et al. (2018)

- Other methods:
  - Optical flow vectors of small local regions, integrated into spatio-temporal regions to find onset/offset times **(Patel et al., 2015)**
  - Random walk model to compute probability of containing MEs **(Xia et al., 2016)**

Performance Metrics & Specific Settings

- ME spotting is akin to a binary detection task (present / not present)
- Typical detection performance metrics:
  - TPR
  - FPR
  - ROC / AUC

$$\text{TPR} = \frac{\text{Number of frames of correctly spotted MEs}}{\text{Total number of ground truth ME frames from all samples}}$$

$$\text{FPR} = \frac{\text{Number of incorrectly spotted frames}}{\text{Total number of non-ME frames from all samples}}$$

- The micro-interval $N = 0.5s$ is taken as the presumed maximum possible duration of MEs
  - Li et al. (2017)'s work used $N = 0.32s$ which corresponds to $N = 65$ for CASME II

Recent Approaches on Spotting Micro-Expression Sequences

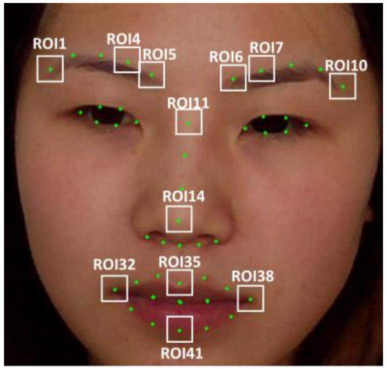
Jingting Li, Sujing Wang, He Ying, **Moi Hoon Yap**,
Catherine Soladie, Renaud Seguier

MELAB
Micro-Expression LABoratory

Chinese Academy of Sciences

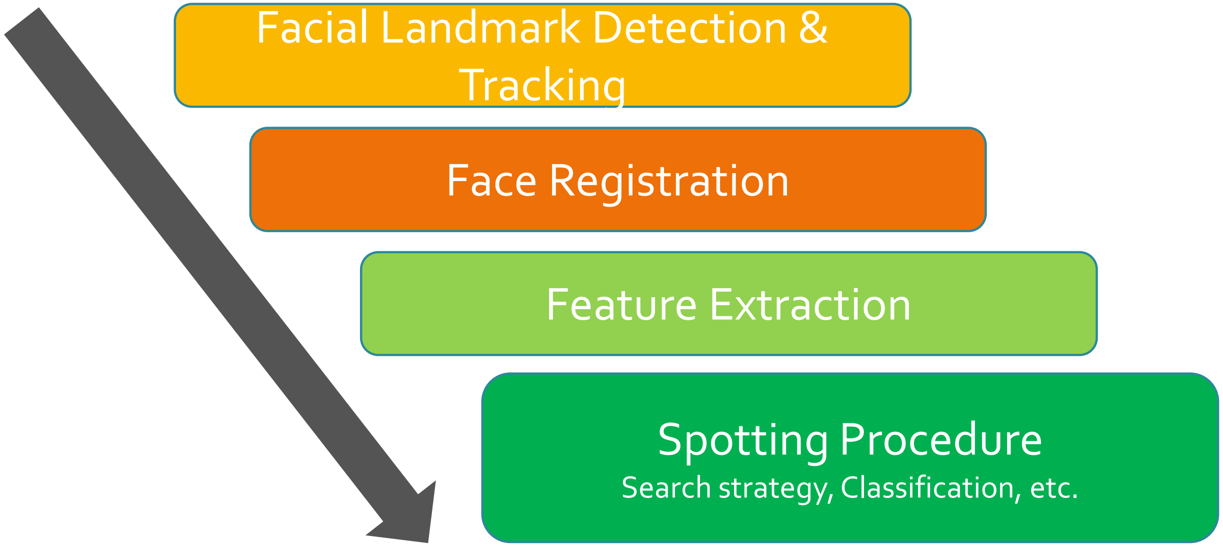
CentraleSupélec

Manchester
Metropolitan
University



ME Apex Spotting Pipeline

ME apex spotting process will follow these steps:



Facial Landmark Detection & Tracking

Face Registration

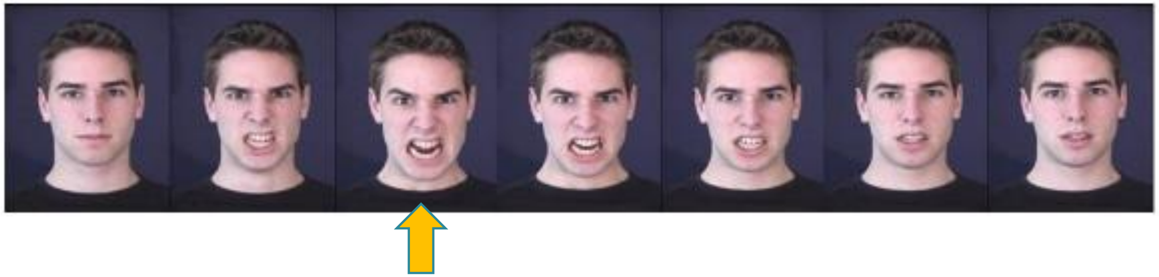
Feature Extraction

Spotting Procedure

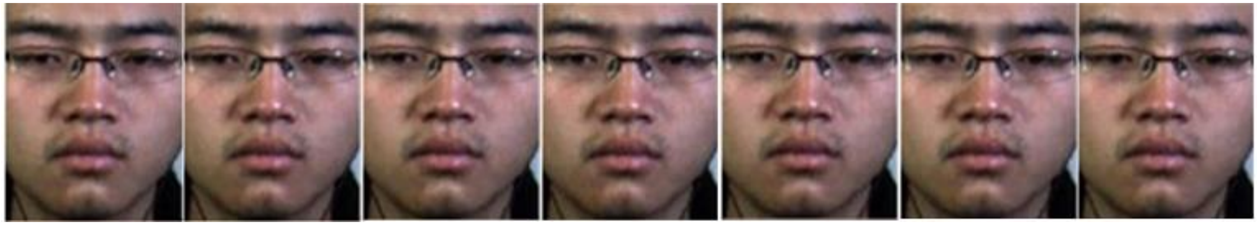
Search strategy, Classification, etc.

Spotting the apex frame

- **Can you tell which frame is the apex (with the strongest indication of emotion) ?**

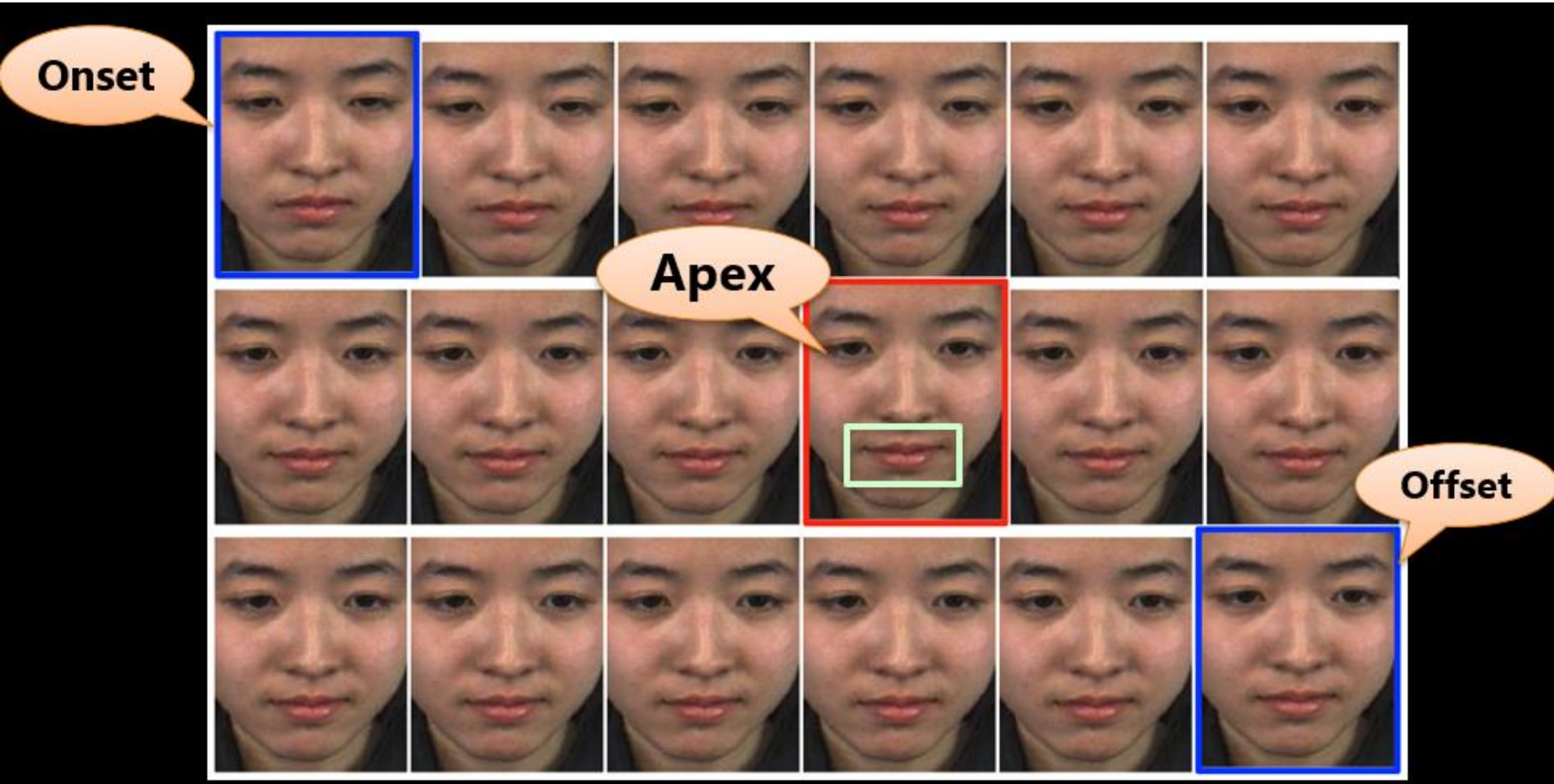


- **How about this ?**



🤔

What is the apex frame?



APEX = Instant indicating the peak intensity of an emotion

Why apex?

- Most expressive frame and better represent the emotion of entire video

- Provide useful information in behavioral psychology field for behavior analysis

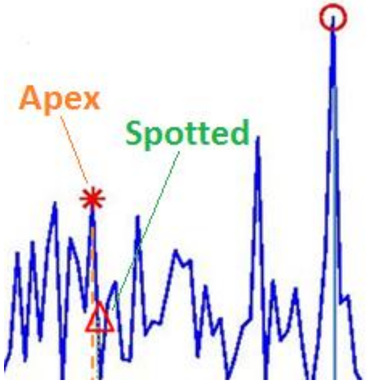
Spotting the apex frame

- **Apex frame spotting:**
  - Apex frame: the instant that is the most expression emotional state in the sequence
  - Motivation: The apex frame is a potentially useful piece of information that may be helpful to recognition (spotting sequences tend to have a high margin of error)
- The first work to attempt this: Yan et al. (2014)
  - CLM, LBP and OF were used to obtain features
  - Frame with largest feature magnitude is selected as the apex
  - However, is the **frame with the strongest magnitude** a good candidate?
  - Can the presence of **noise** (eye blinks, large movements) affect this criterion?

Automatic Apex Frame Spotting in Micro-Expression Database

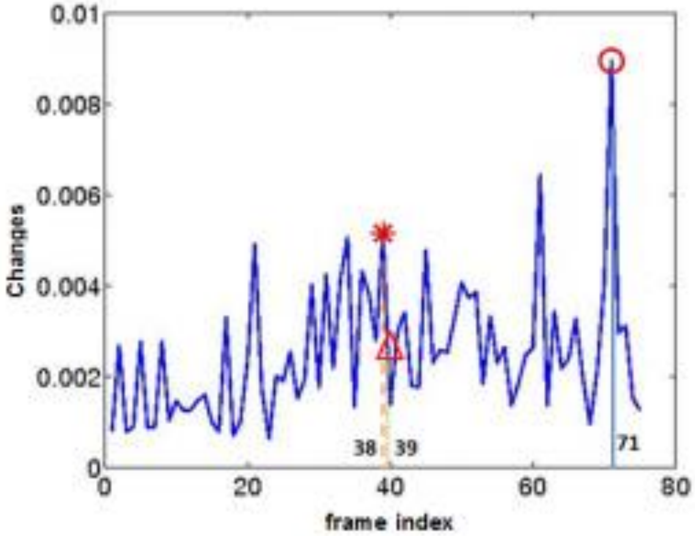
ACPR 2015

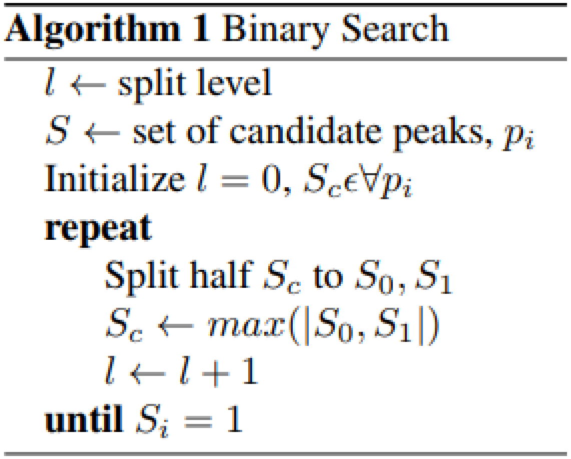
Sze-Teng Liong, John See, KokSheik Wong, Anh Cat Le Ngo,
Yee-Hui Oh, Raphael C.W. Phan



Exhaustive binary search strategy

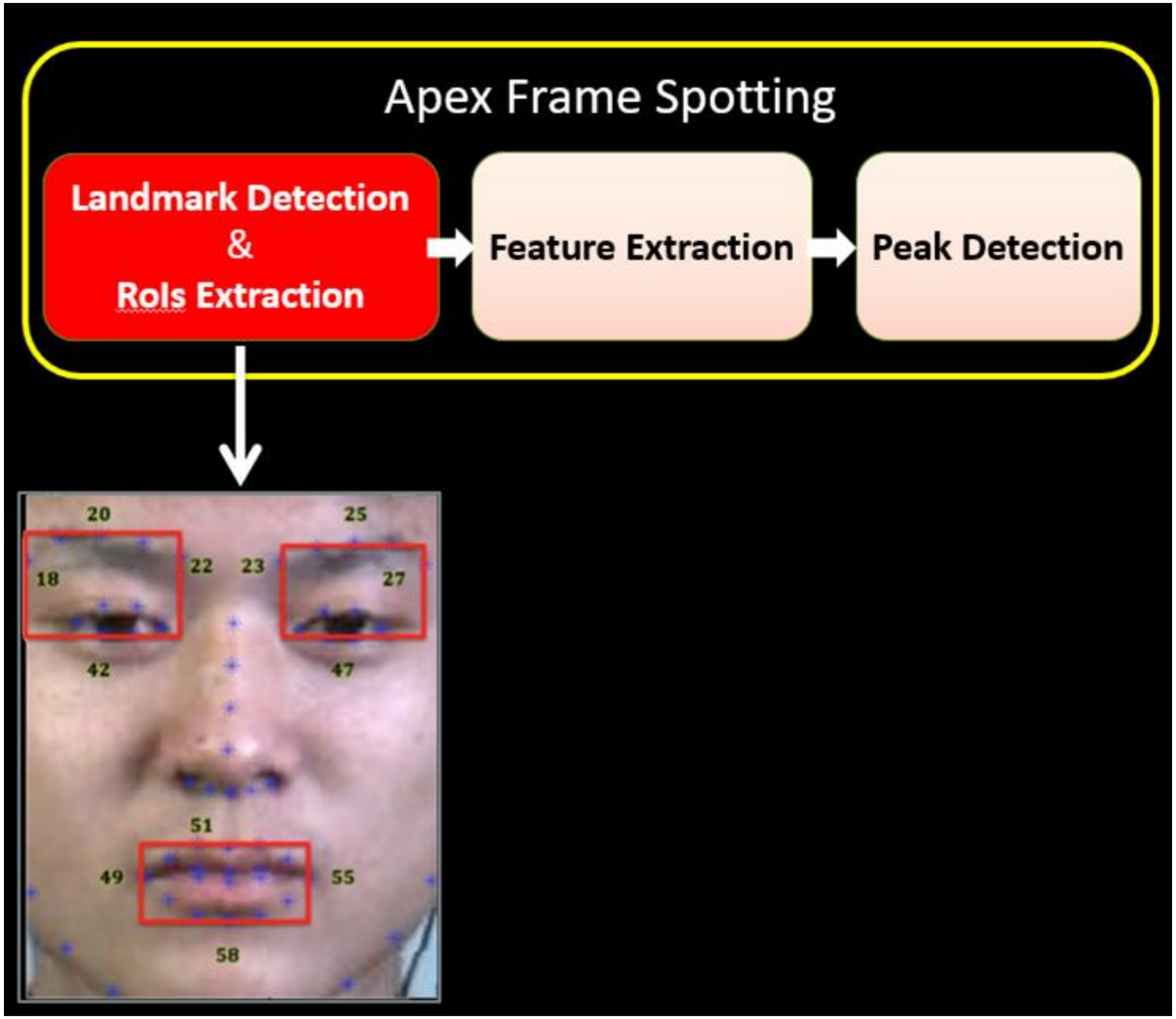
- This is as simple as it gets….
- Exhaustive binary search strategy recursively searches through the set of candidate peaks by further splitting the partition that possesses the larger sum of feature magnitudes. Search stops when the final partition contains only 1 candidate peak, hence can no longer be split.





```
**Algorithm 1** Binary Search
    $l \leftarrow$ split level
    $S \leftarrow$ set of candidate peaks, $p_i$
    Initialize $l = 0$, $S_c \epsilon \forall p_i$
    **repeat**
        Split half $S_c$ to $S_0, S_1$
        $S_c \leftarrow max(|S_0, S_1|)$
        $l \leftarrow l + 1$
    **until** $S_i = 1$
```

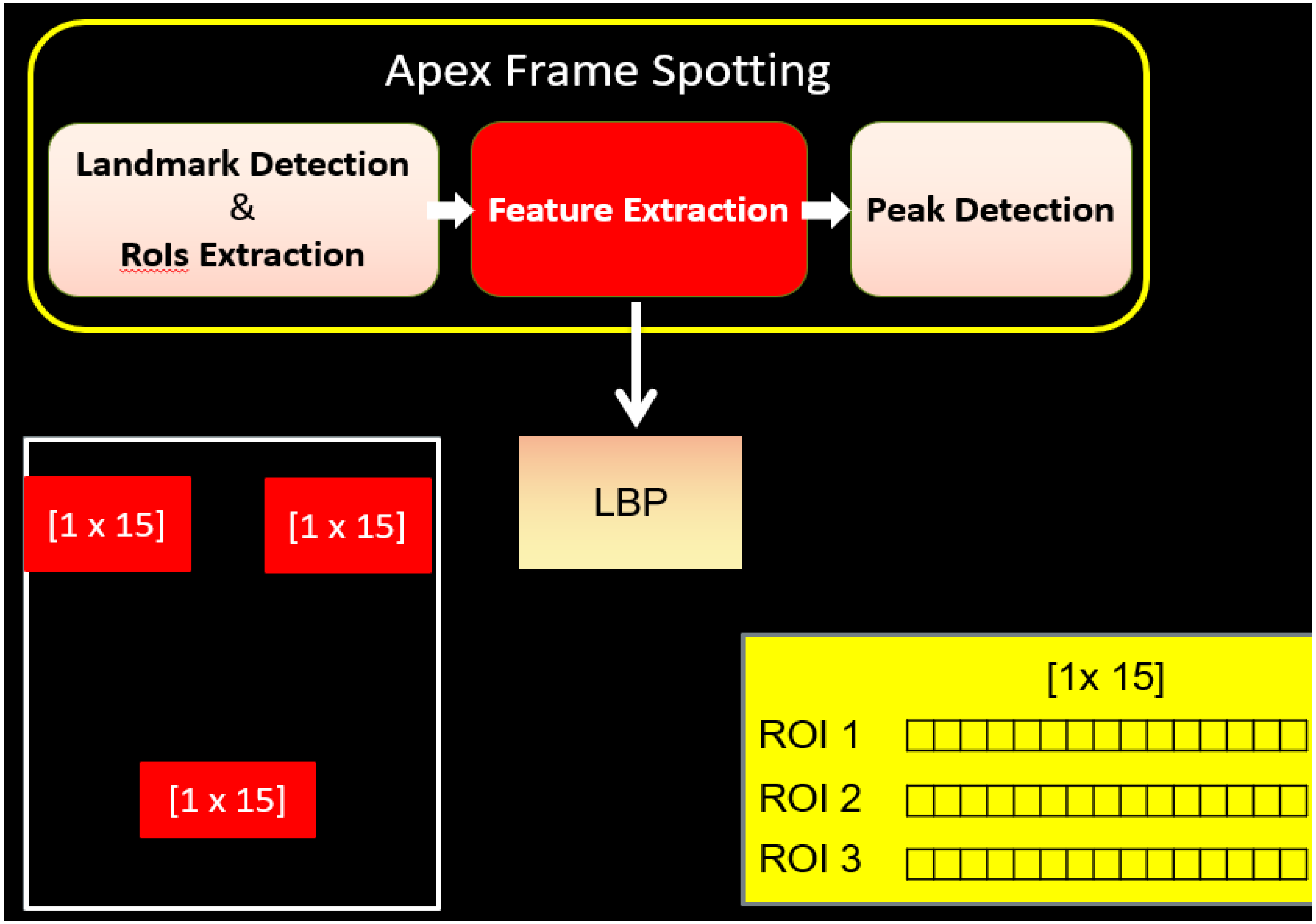
Apex Frame Spotting



Liong, S. T., See, J., Wong, K., Le Ngo, A. C., Oh, Y. H., & Phan, R. (2015). Automatic apex frame spotting in micro-expression database. In 2015 3rd IAPR Asian conference on pattern recognition (ACPR) (pp. 665-669)

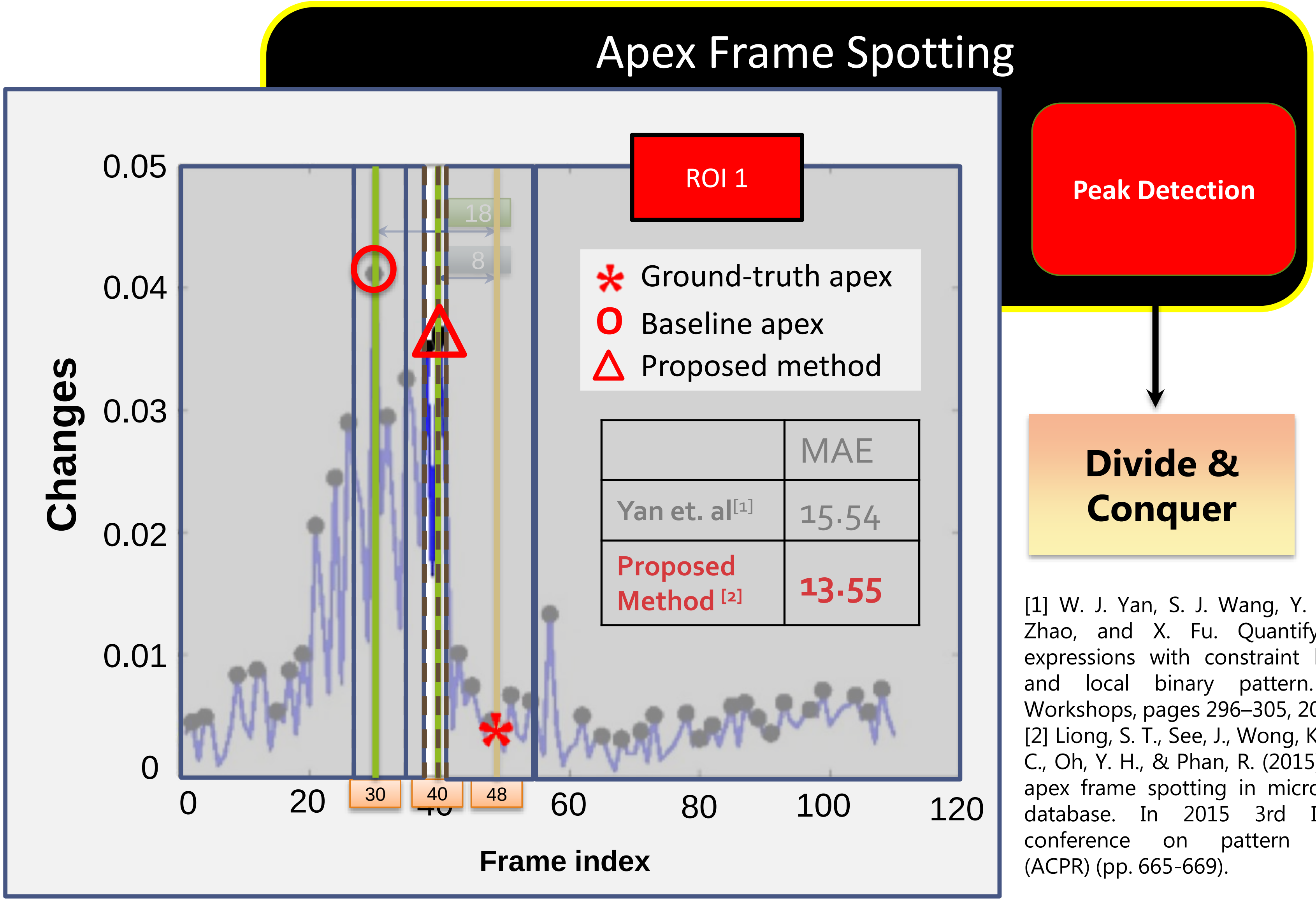
Apex Frame Spotting

- Each ROI, compute LBP features



Liong, S. T., See, J., Wong, K., Le Ngo, A. C., Oh, Y. H., & Phan, R. (2015). Automatic apex frame spotting in micro-expression database. In 2015 3rd IAPR Asian conference on pattern recognition (ACPR) (pp. 665-669)

Apex Frame Spotting

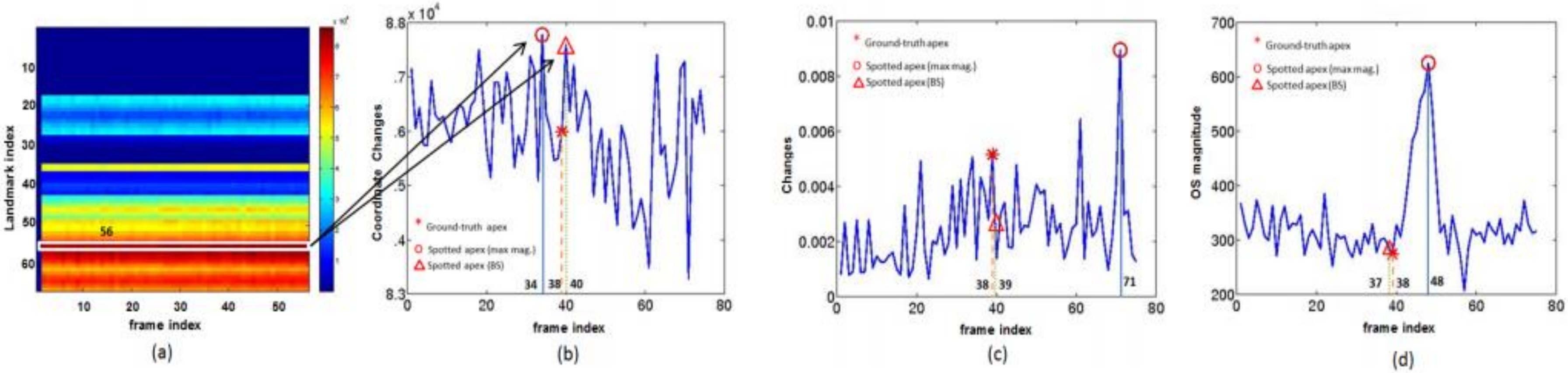


	MAE
Yan et. al[1]	15.54
Proposed Method [2]	13.55

[1] W. J. Yan, S. J. Wang, Y. H. Chen, G. Zhao, and X. Fu. Quantifying micro-expressions with constraint local model and local binary pattern. In ECCV Workshops, pages 296–305, 2014.

[2] Liong, S. T., See, J., Wong, K., Le Ngo, A. C., Oh, Y. H., & Phan, R. (2015). Automatic apex frame spotting in micro-expression database. In 2015 3rd IAPR Asian conference on pattern recognition (ACPR) (pp. 665-669).

Searching for the apex frame



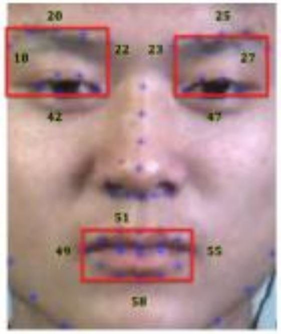
- The density heatmap shows the landmark # that contains the most changes in the video
- The proposed binary search strategy is able to get as close as possible to the g/t apex instead of doing the “greedy”, i.e. going for the max magnitude

Results

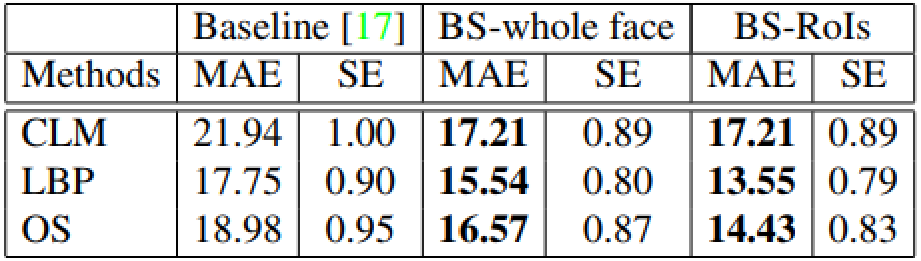
- **Performance metrics:**

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|e_i|$$

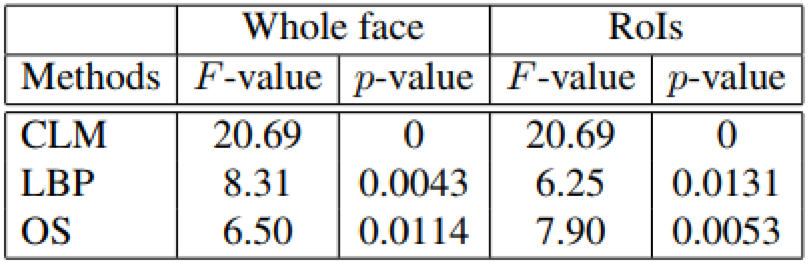
$$SE = \frac{\sigma}{\sqrt{n}},$$



- **Results:**

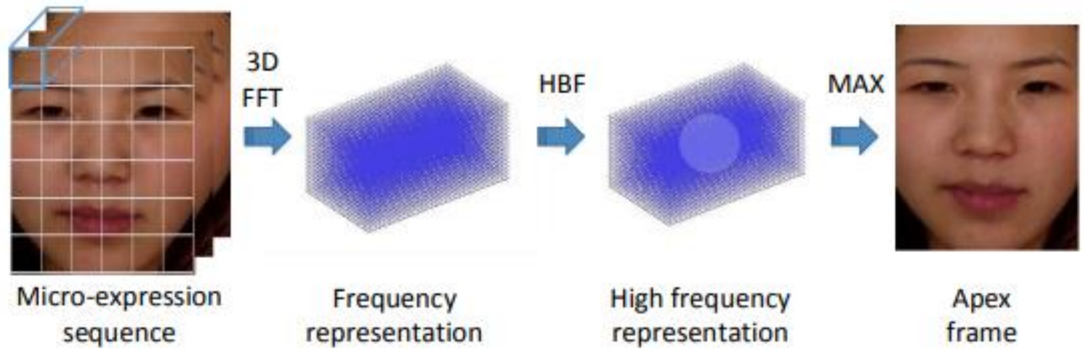


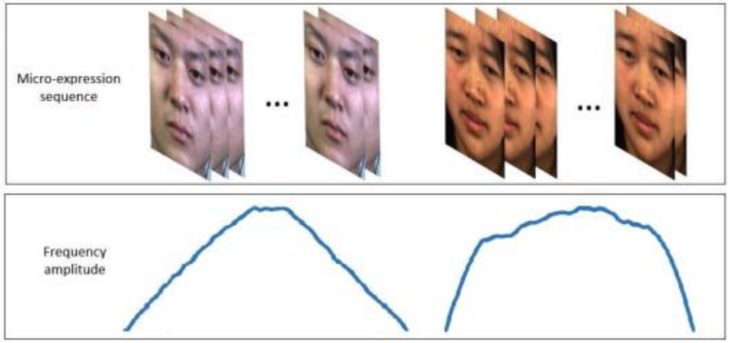
	Baseline [17]		BS-whole face		BS-RoIs	
Methods	MAE	SE	MAE	SE	MAE	SE
CLM	21.94	1.00	**17.21**	0.89	**17.21**	0.89
LBP	17.75	0.90	**15.54**	0.80	**13.55**	0.79
OS	18.98	0.95	**16.57**	0.87	**14.43**	0.83



	Whole face		RoIs	
Methods	F-value	p-value	F-value	p-value
CLM	20.69	0	20.69	0
LBP	8.31	0.0043	6.25	0.0131
OS	6.50	0.0114	7.90	0.0053

Searching for Apex using Frequency Amplitude





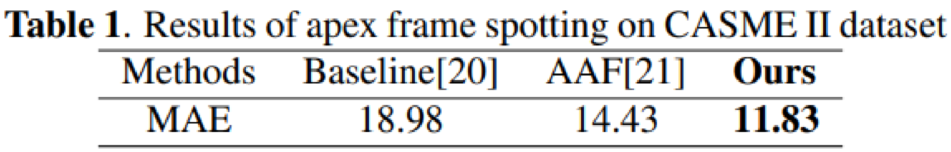


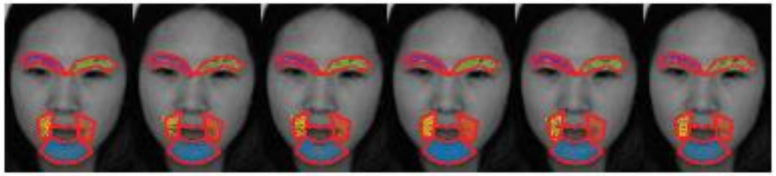
Table 1. Results of apex frame spotting on CASME II dataset

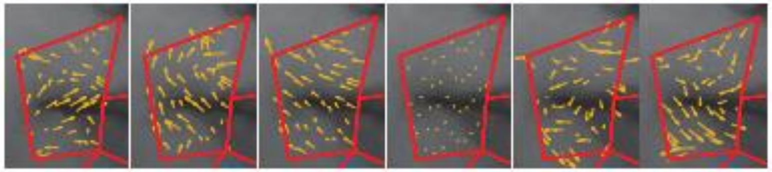
Methods	Baseline[20]	AAF[21]	**Ours**
MAE	18.98	14.43	**11.83**

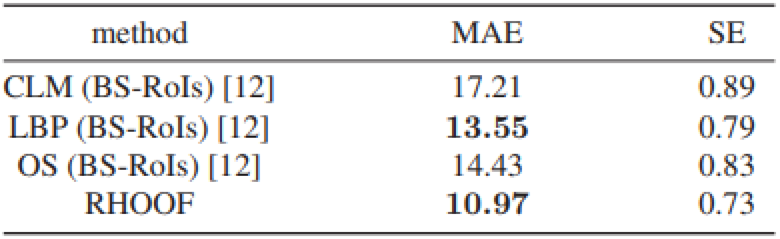
- Introduce the use of frequency amplitude (via High-frequency Band Filter) to track the likely location of apex frame
- The freq. amp. of all 36 pre-determined blocks are accumulated for each frame interval ➔ max is chosen as “peak interval”
- Middle frame of peak interval is chosen as apex

Li, Y., Huang, X., & Zhao, G. (2018). Can micro-expression be recognized based on single apex frame?. In 2018 25th IEEE International Conference on Image Processing (ICIP) (pp. 3094-3098)

Searching for Apex using Region-based HOOF (RHOOF)







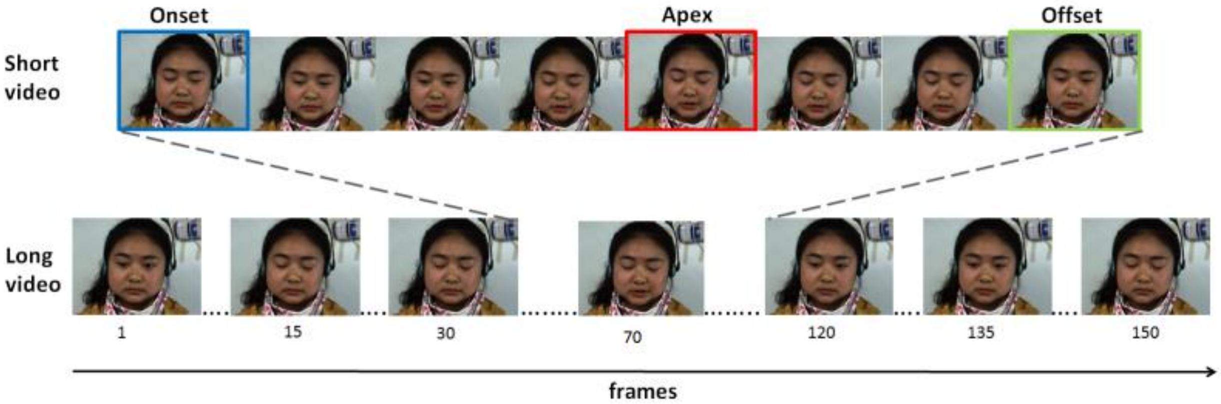
method	MAE	SE
CLM (BS-RoIs) [12]	17.21	0.89
LBP (BS-RoIs) [12]	**13.55**	0.79
OS (BS-RoIs) [12]	14.43	0.83
RHOOF	**10.97**	0.73

- Optical flow fields (Pyramid LK) computed for 5 ROIs
- HOOF is computed for each ROI, binned based on 8 directions
- Region priority (based on statistics): Eyebrow → Lip corner → Chin

Ma, H., An, G., Wu, S., & Yang, F. (2017). A region histogram of oriented optical flow (RHOOF) feature for apex frame spotting in micro-expression. In *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)* (pp. 281-286)

Spotting apex from long videos

- Given a long untrimmed video, find the apex frame.



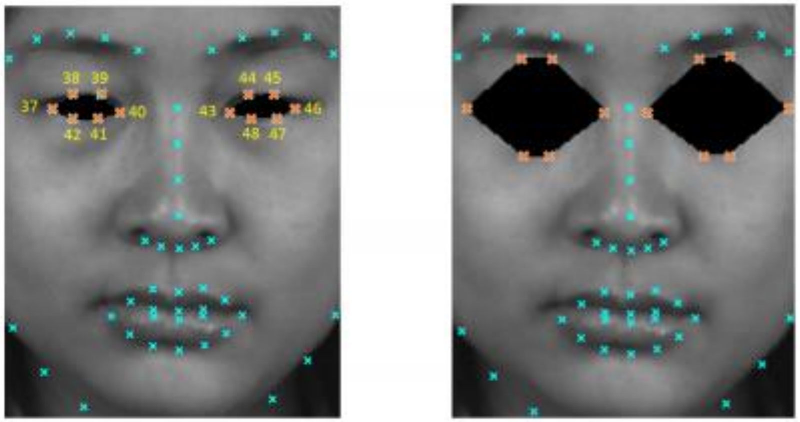
- Performance metric
  - MAE / SE
  - **Apex Spotting Rate (ASR):** The success rate of spotting the apex within the short video

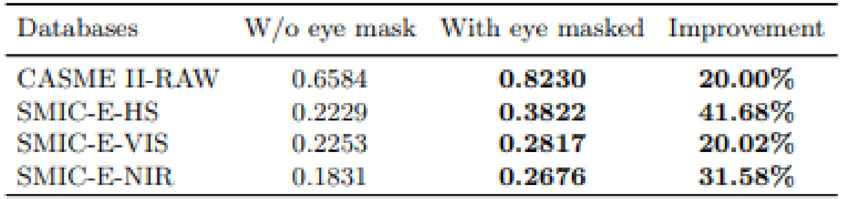
$$ASR = \frac{1}{M}\sum_{j=1}^{M}\alpha$$

$$\alpha = \begin{cases} 1 & if\ f^* \in (f_{j,onset}, f_{j,offset}) \\ 0 & otherwise \end{cases}$$

Spotting apex from long videos – Vital Pre-processes

- Eye Masking: A crucial step
  - Avoid erroneous detection caused by motions more intense than a ME, i.e. eye blinks, squints



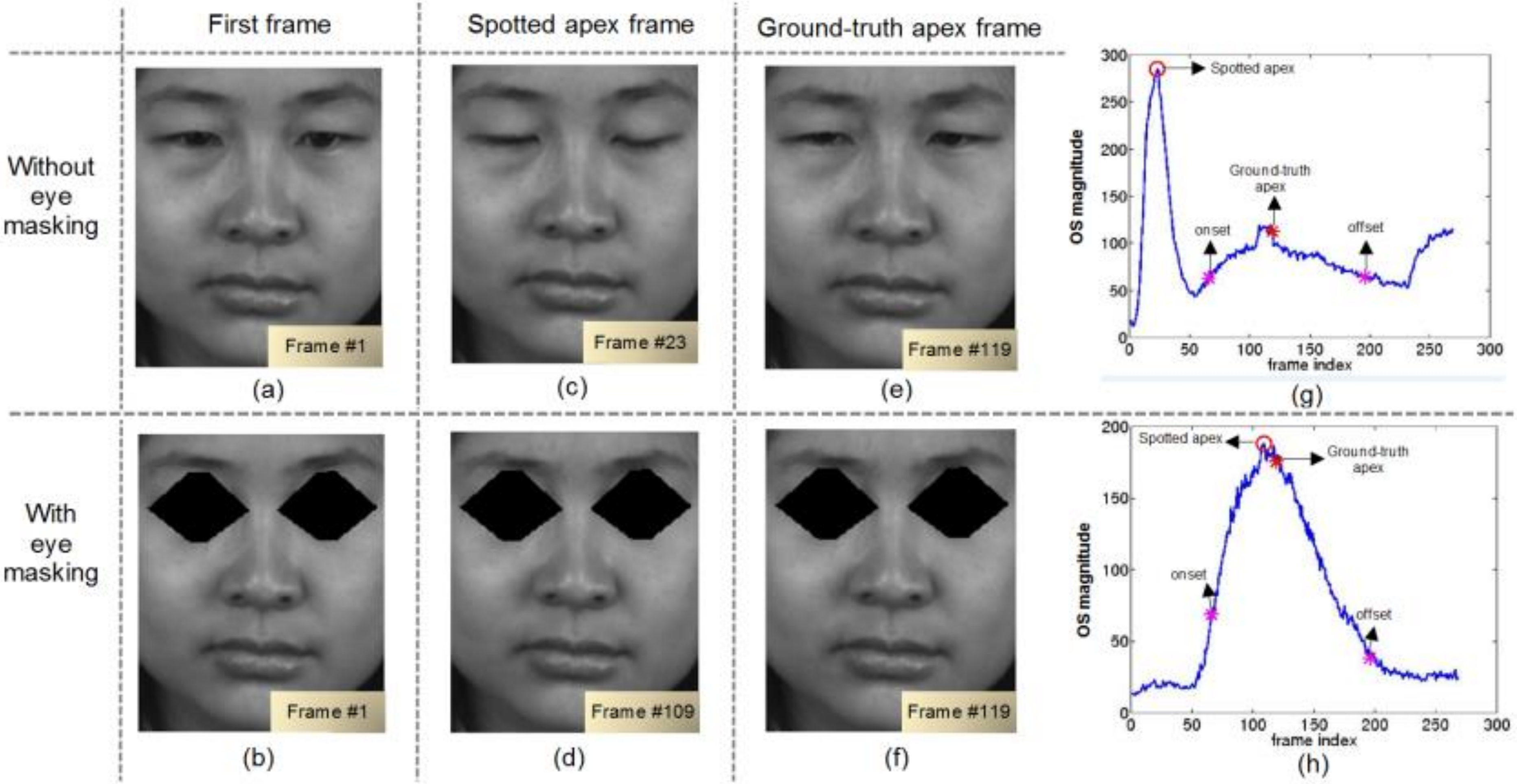


Databases	W/o eye mask	With eye masked	Improvement
CASME II-RAW	0.6584	**0.8230**	**20.00%**
SMIC-E-HS	0.2229	**0.3822**	**41.68%**
SMIC-E-VIS	0.2253	**0.2817**	**20.02%**
SMIC-E-NIR	0.1831	**0.2676**	**31.58%**

- ROI Selection: Equally important for better filtering of motions
  - Most contributions come from (1) eye and eyebrow region, (2) mouth regions, rather than the whole face
  - Cropping of ROI based on landmarks ➔ Pseudo-widening of ROIs ➔ block-based representation within ROI for more precise local features

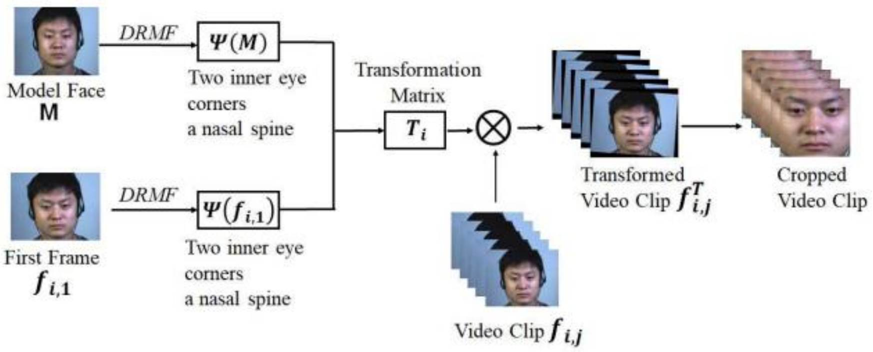
Liong, S. T., See, J., Wong, K., Le Ngo, A. C., Oh, Y. H., & Phan, R. (2015). Automatic apex frame spotting in micro-expression database. In 2015 3rd IAPR Asian conference on pattern recognition (ACPR) (pp. 665-669).

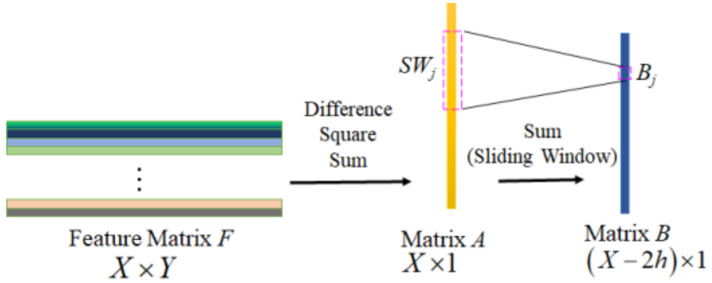
Spotting apex from long videos – Vital Pre-processes

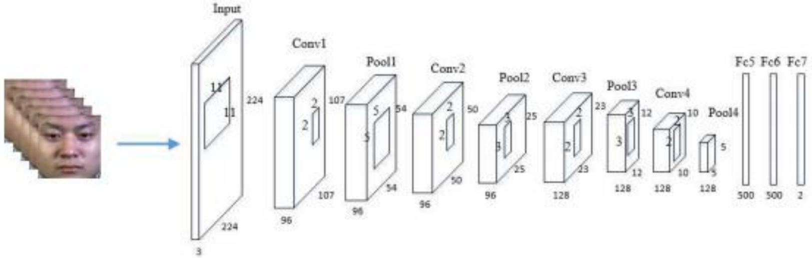


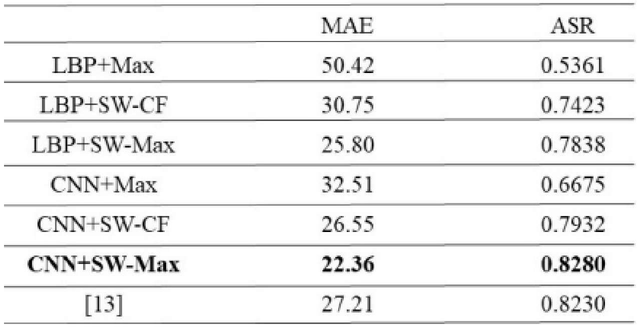
Liong, S. T., See, J., Wong, K., Le Ngo, A. C., Oh, Y. H., & Phan, R. (2015). Automatic apex frame spotting in micro-expression database. In 2015 3rd IAPR Asian conference on pattern recognition (ACPR) (pp. 665-669).

Spotting apex from long videos – ConvNet as Feature Extractor









	MAE	ASR
LBP+Max	50.42	0.5361
LBP+SW-CF	30.75	0.7423
LBP+SW-Max	25.80	0.7838
CNN+Max	32.51	0.6675
CNN+SW-CF	26.55	0.7932
CNN+SW-Max	**22.36**	**0.8280**
[13]	27.21	0.8230

CASME II

- CNN is used as a feature extractor to extract frame-wise features (FC6)
- Subsequent representations (A, B) are derived from this feature matrix F via a series of transformations (sum-of-square difference, sliding window with summation)
- Largest value in sliding window SW is the apex after finding largest value in B

Zhang, Z., Chen, T., Meng, H., Liu, G., & Fu, X. (2018). SMEConvNet: A convolutional neural network for spotting spontaneous facial micro-expression from long videos. *IEEE Access*, *6*, 71143-71151.

Spotting apex from long videos: Summary

- Spotting of apex frame can be useful towards recognition.
  - Find apex -> use apex to recognise the emotional class
- Spotting of apex frame still has room for improvement
  - Particularly when there's large amounts of noise to handle
  - Doing this consistently across different datasets remains a challenge
- Spotting of apex frame is rarely attempted for long videos with multiple MEs or a combination of macro- and micro-expressions.
  - Room for future research!

End of Part 3

Questions?