The More the Merrier: Analysing the Affect of a Group of People in Images

Abhinav Dhall, Jyoti Joshi, Karan Sikka, Roland Goecke and Nicu Sebe

University of Canberra, Australian National University,
University of California San Diego, University of Trento

Inferring Affect

- Since we are dealing with images here, let’s pose it as a problem of expression analysis
- Methods can be categorised on the basis of:
  - Posed / Spontaneous
  - Discrete / Continuous
  - Lab-controlled / `in the wild’ (unconstrained)

Inferring Affect (2)

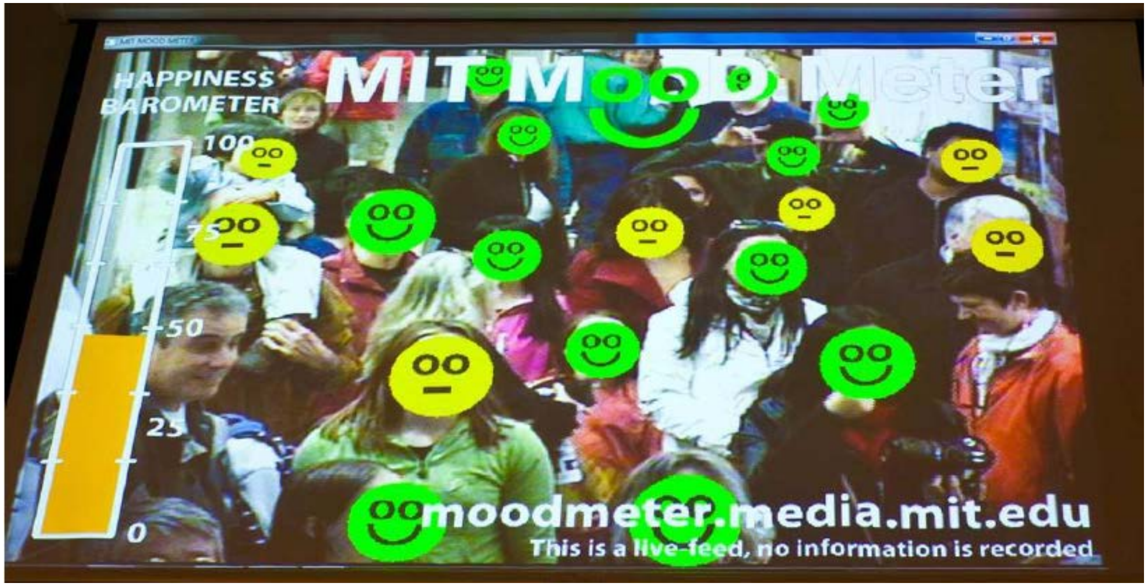
- Since we are dealing with images here, let's pose it as a problem of expression analysis
- Methods can be categorised on the basis of:
  - Posed / Spontaneous
  - Discrete / Continuous
  - Lab-controlled / `in the wild' (unconstrained)

Due to the large number of images being posted on the internet (1.8 B/day) from social events, there is another attribute for categorizing emotion recognition:

- *Single subject / Multiple subjects in the scene*
- Inferring the affect of a group of people from images

Prior Work

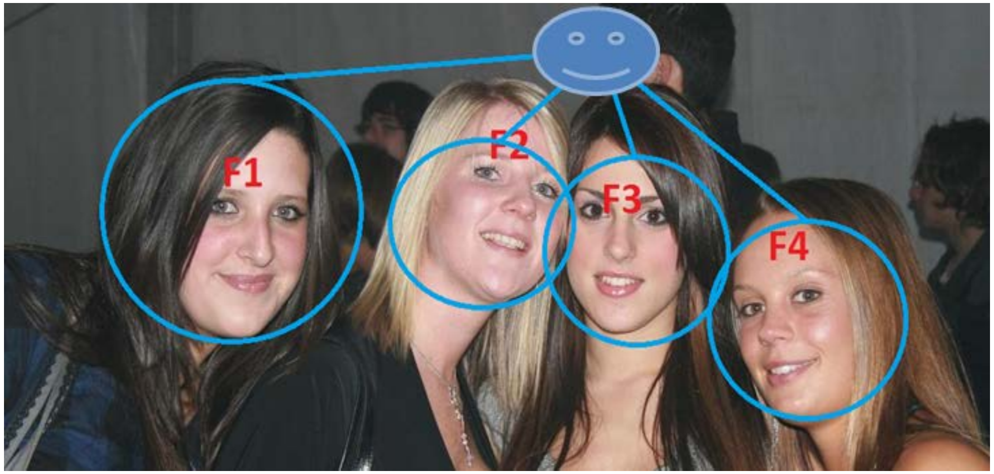
- Hoque et al. 2012
  - MIT Mood Meter – four cameras installed at different locations on MIT campus to infer mood of passers-by
  - Based on averaging of smile intensities



System snapshot from the original paper

Prior Work (2)

- Dhall et al. 2012 & 2015
  - Group happiness intensity analysis
  - Topic modelling of data-driven and manual attributes
  - Limited to positive emotion only
  - More details in the late morning session!

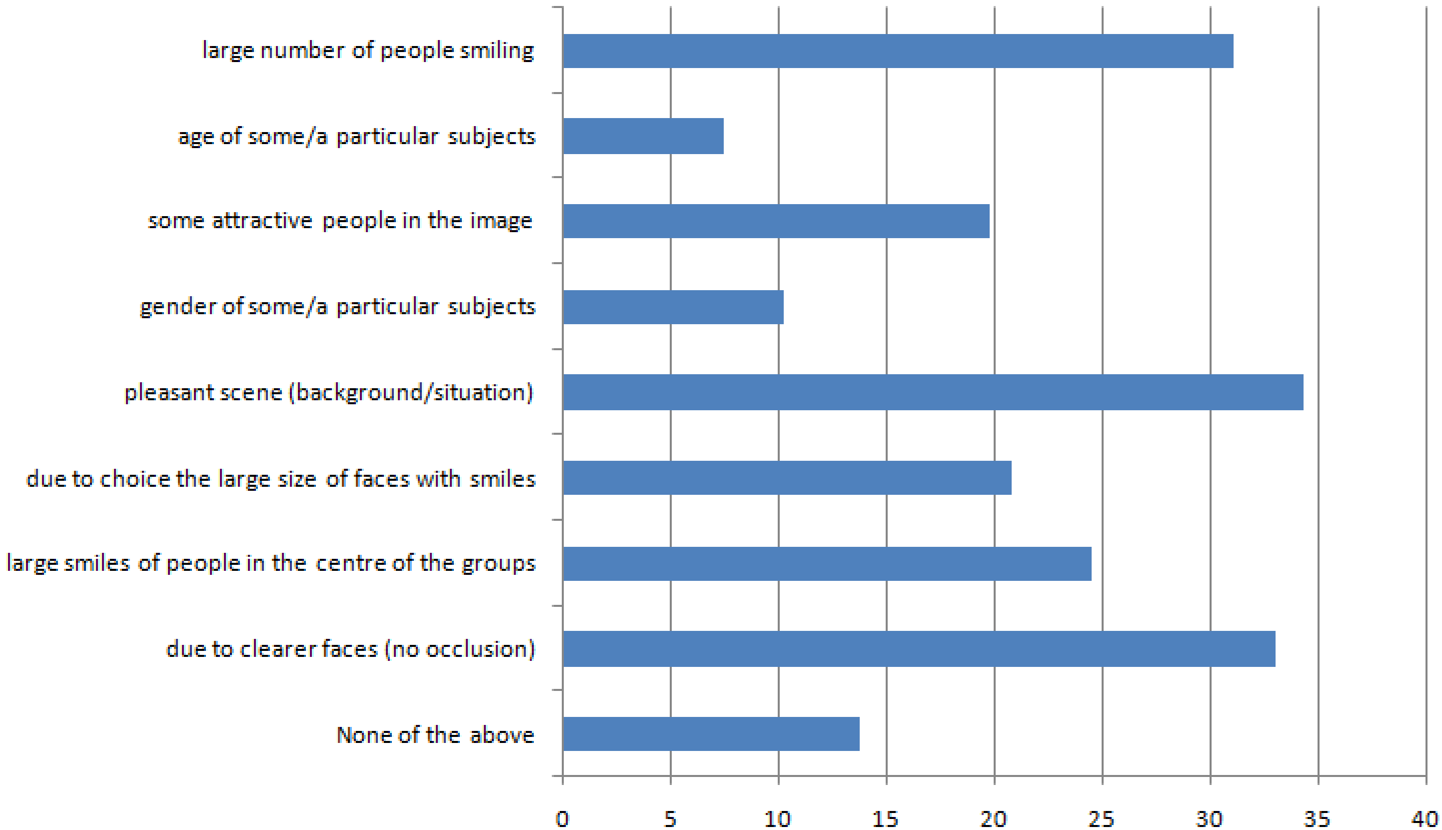


Group Affect

- Barsade et al. 1998 and Kelly et al. 2001
  - **Top-down component:**
    Overall emotion of group constructed by uniqueness of individual members' emotion expression.
  - **Bottom-up component:**
    Emotion emerging at the group level and followed by individual participants of the group.

- Survey (Dhall et al. 2015)
  - **Global affect:**
    Scene, clothes, neighbours
  - **Local affect:**
    Facial expression and other facial attributes such as occlusion, age, gender etc.

Survey

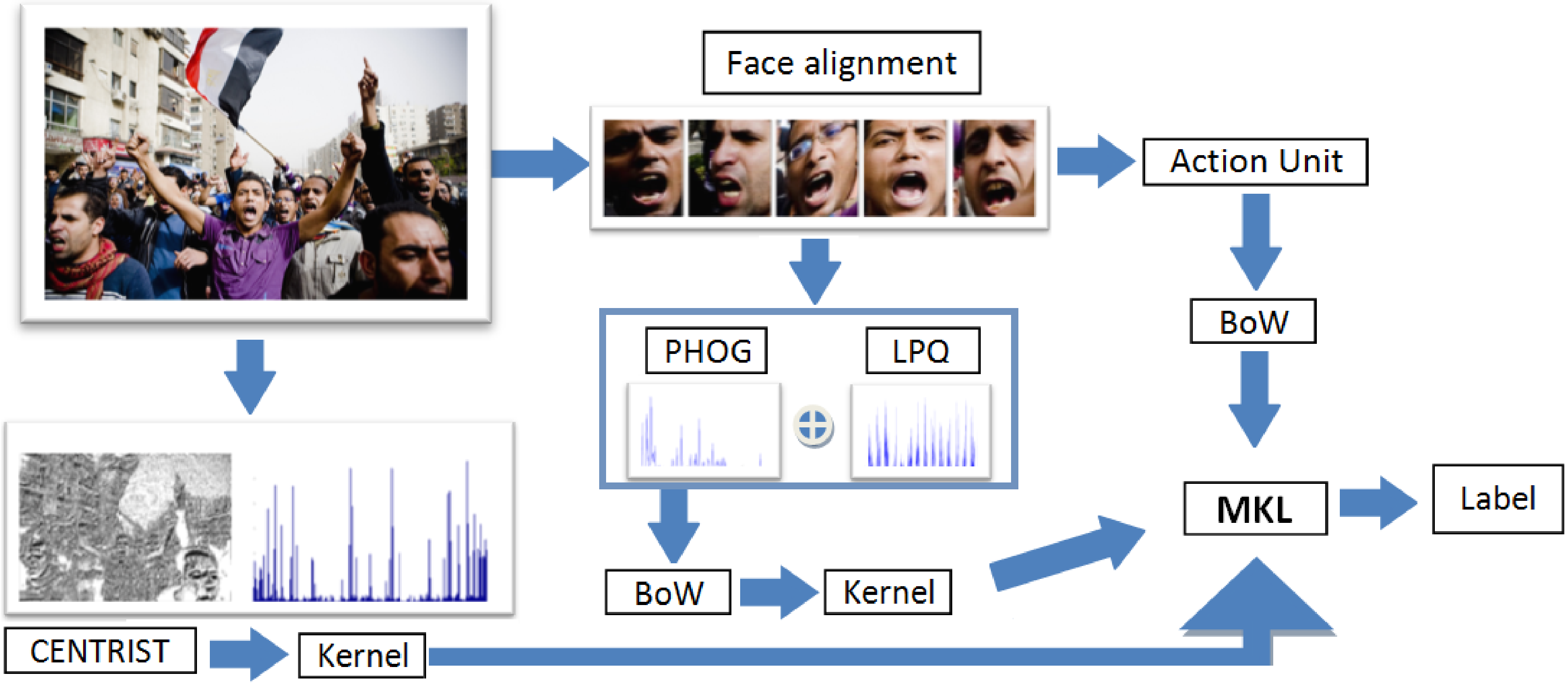


Attributes as mentioned in the survey

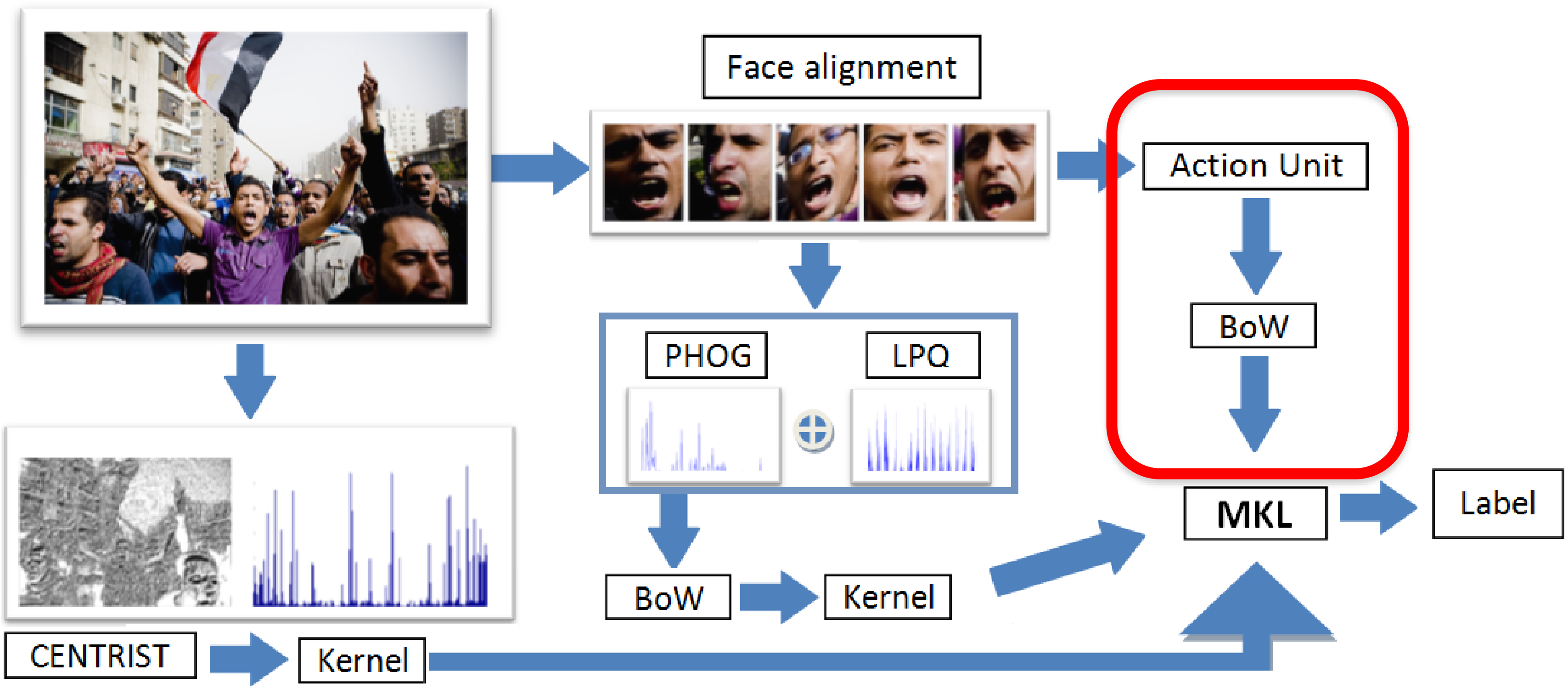
Data

- Group Affect Database
  - Keywords based search from Flickr, Google Images, HAPPEI database
  - Positive – Neutral – Negative classes (3 human annotators)

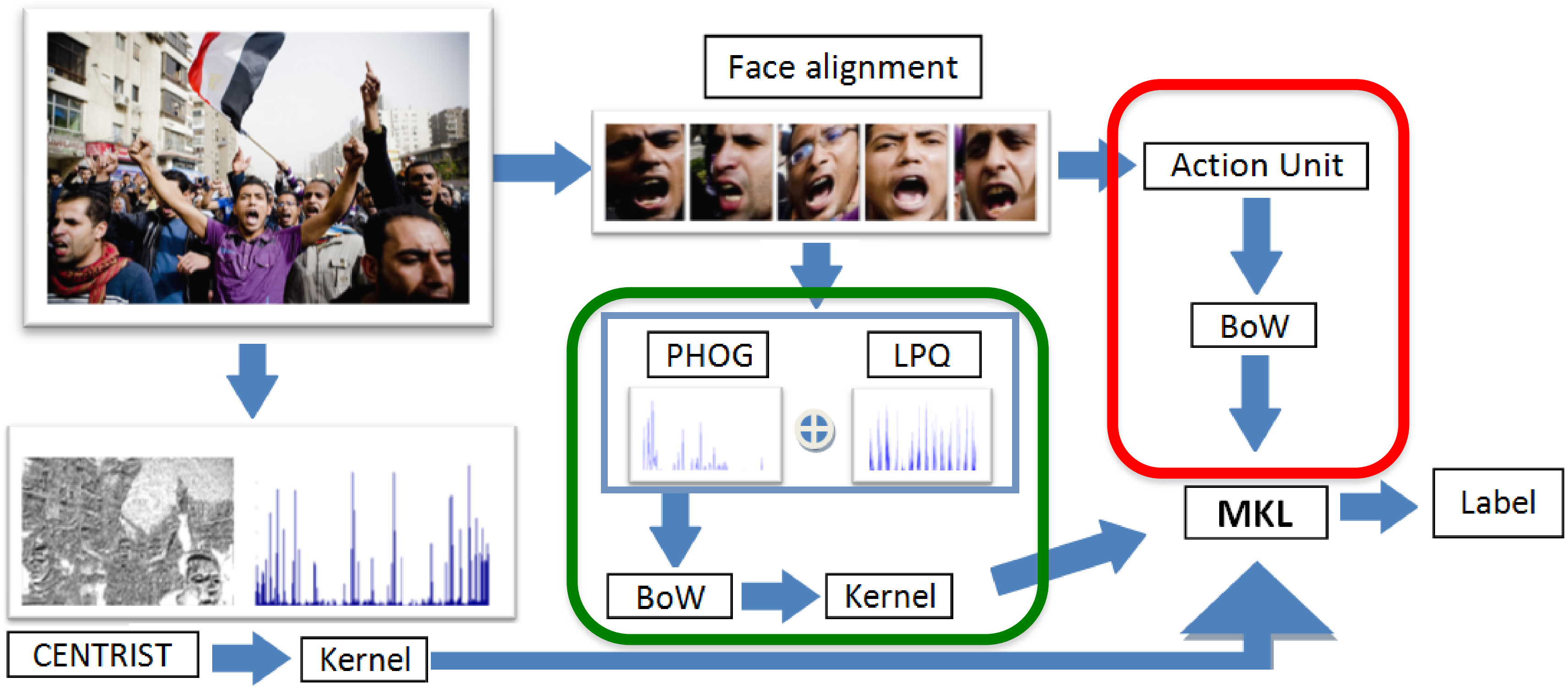
Proposed Framework



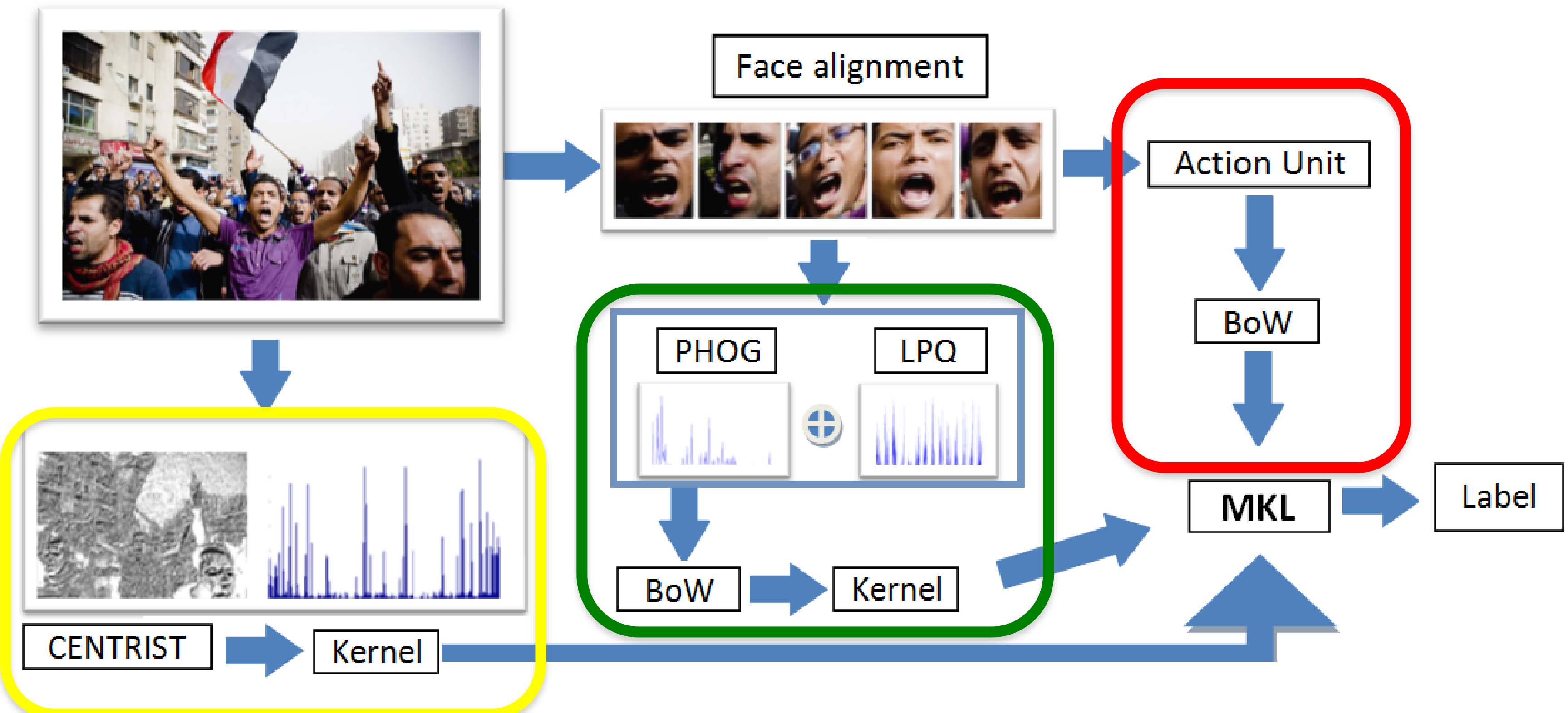
Proposed Framework



Proposed Framework



Proposed Framework



Face Analysis

Bottom-up component

- Face detection using Mixture of Pictorial Structures (Zhu and Ramanan 2012)
- Facial Action Unit (AU) (CERT toolbox)
- The group is modelled as a Bag of Words (BoW_AU)
- Each face in a group is a word
- Group of people is a document

Face Analysis (II)

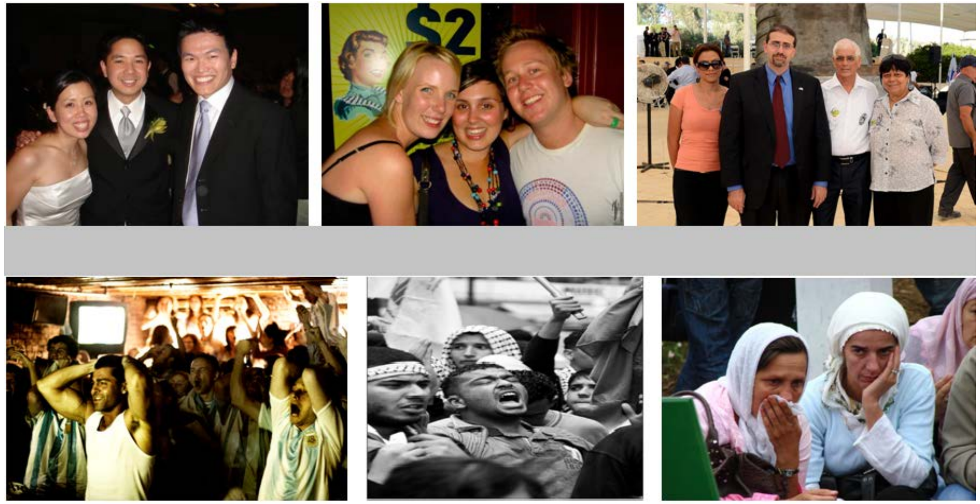
Bottom-up component

- Based on the survey: various attributes about the subjects in the group that effect the perception of the affect of a group.
- Age, gender, attractiveness, facial features: glasses, moustaches etc.
- Compute low-level features:
  - Pyramid of Histogram of Gradients (PHOG), Bosch et al. 2007
  - Local Phase Quantization (LPQ), Ojansivu et al. 2008
- Bag of Words: BoW_LL

Scene Analysis

Top-down component

- GIST descriptor: Scene_GIST
- CENTRIST descriptor: Scene_CENTRIST
- Descriptors compute statistics at global level
- Take into consideration both the scene background and information that may define the situation such as clothes



Fusion

- Fusion is performed between the scene and face features
- Moshe Bar's (2004) scene context model:
  - Low-resolution holistic representation ➔ similar to scene descriptor
  - Detailed object-level representation ➔ face analysis

Feature Fusion and Multiple Kernel Learning

- BoW_AU
- BoW_LL
- Scene_CENTRIST or Scene_GIST

Experiments

Feature	Positive	Neutral	Negative	Final
BoW_AU	70.93	33.33	37.93	50.43
BoW_LL	76.74	56.66	06.90	50.98
Scene_GIST	52.32	38.33	31.03	42.16
Scene_CENTRIST	50.00	45.00	39.65	45.58

Classification accuracy (%) feature wise

- High-level features based on AU perform similar to the low-level feature combination
- Classification accuracy for the Negative class is lower
- Why? Negative affect images in the database have a higher number of non-frontal faces and occlusions (e.g. protest rallies)

Experiments

Feature	Positive	Neutral	Negative	Final
BoW_AU	70.93	33.33	37.93	50.43
BoW_LL	76.74	56.66	06.90	50.98
Scene_GIST	52.32	38.33	31.03	42.16
Scene_CENTRIST	50.00	45.00	39.65	45.58

Classification accuracy (%) feature wise

- High-level features based on AU perform similar to the low-level feature combination
- Classification accuracy for the Negative class is lower
- Why? Negative affect images in the database have a higher number of non-frontal faces and occlusions (e.g. protest rallies)

Experiments

Feature	Positive	Neutral	Negative	Final
BoW_AU	70.93	33.33	37.93	50.43
BoW_LL	76.74	56.66	06.90	50.98
Scene_GIST	52.32	38.33	31.03	42.16
Scene_CENTRIST	50.00	45.00	39.65	45.58

Classification accuracy (%) feature wise

- High-level features based on AU perform similar to the low-level feature combination
- Classification accuracy for the Negative class is lower
- Why? Negative affect images in the database have a higher number of non-frontal faces and occlusions (e.g. protest rallies)

Experiments

Feature	Positive	Neutral	Negative	Final
BoW_AU	70.93	33.33	37.93	50.43
BoW_LL	76.74	56.66	06.90	50.98
Scene_GIST	52.32	38.33	31.03	42.16
Scene_CENTRIST	50.00	45.00	39.65	45.58

Classification accuracy (%) feature wise

- High-level features based on AU perform similar to the low-level feature combination
- Classification accuracy for the Negative class is lower
- Why? Negative affect images in the database have a higher number of non-frontal faces and occlusions (e.g. protest rallies)

Experiments

Feature	Positive	Neutral	Negative	Final
BoW_AU	70.93	33.33	37.93	50.43
BoW_LL	76.74	56.66	06.90	50.98
Scene_GIST	52.32	38.33	31.03	42.16
Scene_CENTRIST	50.00	45.00	39.65	45.58

Classification accuracy (%) feature wise

- High-level features based on AU perform similar to the low-level feature combination
- Classification accuracy for the Negative class is lower
- Why? Negative affect images in the database have a higher number of non-frontal faces and occlusions (e.g. protest rallies)

Experiments

Feature	Positive	Neutral	Negative	Final
BoW_AU	70.93	33.33	37.93	50.43
BoW_LL	76.74	56.66	06.90	50.98
Scene_GIST	52.32	38.33	31.03	42.16
Scene_CENTRIST	50.00	45.00	39.65	45.58

Classification accuracy (%) feature wise

- High-level features based on AU perform similar to the low-level feature combination
- Classification accuracy for the Negative class is lower
- Why? Negative affect images in the database have a higher number of non-frontal faces and occlusions (e.g. protest rallies)

Experiments

Feature	Positive	Neutral	Negative	Final
BoW_AU	70.93	33.33	37.93	50.43
BoW_LL	76.74	56.66	06.90	50.98
Scene_GIST	52.32	38.33	31.03	42.16
Scene_CENTRIST	50.00	45.00	39.65	45.58

Classification accuracy (%) feature wise

- High-level features based on AU perform similar to the low-level feature combination
- Classification accuracy for the Negative class is lower
- Why? Negative affect images in the database have a higher number of non-frontal faces and occlusions (e.g. protest rallies)

Experiments (2)

Feature	Positive	Neutral	Negative	Final
BoW_LL + BoW_AU + Scene_GIST	63.95	38.33	46.55	51.47
BoW_LL + BoW_AU	86.04	31.66	20.68	51.47
BoW_LL + BoW_AU+ Scene_CENTRIST	51.12	48.33	44.82	48.52
MKL - BoW_LL + BoW_AU + Scene_GIST	82.55 (0.0083)	78.33 (0.7993)	50.00 (0.1924)	67.15
MKL - BoW_LL + BoW_AU + Scene_CENTRIST	83.72 (0.0085)	80.00 (0.7976)	31.03 (0.1938)	67.64

Feature fusion and MKL based classification accuracy (%) performance

Experiments (2)

Feature	Positive	Neutral	Negative	Final
BoW_LL + BoW_AU + Scene_GIST	63.95	38.33	46.55	51.47
BoW_LL + BoW_AU	86.04	31.66	20.68	51.47
BoW_LL + BoW_AU+ Scene_CENTRIST	51.12	48.33	44.82	48.52
MKL - BoW_LL + BoW_AU + Scene_GIST	82.55 (0.0083)	78.33 (0.7993)	50.00 (0.1924)	67.15
MKL - BoW_LL + BoW_AU + Scene_CENTRIST	83.72 (0.0085)	80.00 (0.7976)	31.03 (0.1938)	67.64

Feature fusion and MKL based classification accuracy (%) performance

Experiments (2)

Feature	Positive	Neutral	Negative	Final
BoW_LL + BoW_AU + Scene_GIST	63.95	38.33	46.55	51.47
BoW_LL + BoW_AU	86.04	31.66	20.68	51.47
BoW_LL + BoW_AU+ Scene_CENTRIST	51.12	48.33	44.82	48.52
MKL - BoW_LL + BoW_AU + Scene_GIST	82.55 (0.0083)	78.33 (0.7993)	50.00 (0.1924)	67.15
MKL - BoW_LL + BoW_AU + Scene_CENTRIST	83.72 (0.0085)	80.00 (0.7976)	31.03 (0.1938)	67.64

Feature fusion and MKL based classification accuracy (%) performance

Experiments (2)

Feature	Positive	Neutral	Negative	Final
BoW_LL + BoW_AU + Scene_GIST	63.95	38.33	46.55	51.47
BoW_LL + BoW_AU	86.04	31.66	20.68	51.47
BoW_LL + BoW_AU+ Scene_CENTRIST	51.12	48.33	44.82	48.52
MKL - BoW_LL + BoW_AU + Scene_GIST	82.55 (0.0083)	78.33 (0.7993)	50.00 (0.1924)	67.15
MKL - BoW_LL + BoW_AU + Scene_CENTRIST	83.72 (0.0085)	80.00 (0.7976)	31.03 (0.1938)	67.64

Feature fusion and MKL based classification accuracy (%) performance

Experiments (2)

Feature	Positive	Neutral	Negative	Final
BoW_LL + BoW_AU + Scene_GIST	63.95	38.33	46.55	51.47
BoW_LL + BoW_AU	86.04	31.66	20.68	51.47
BoW_LL + BoW_AU+ Scene_CENTRIST	51.12	48.33	44.82	48.52
MKL - BoW_LL + BoW_AU + Scene_GIST	82.55 (0.0083)	78.33 (0.7993)	50.00 (0.1924)	67.15
MKL - BoW_LL + BoW_AU + Scene_CENTRIST	83.72 (0.0085)	80.00 (0.7976)	31.03 (0.1938)	67.64

Feature fusion and MKL based classification accuracy (%) performance

Conclusions

- A new framework for inferring the affect of a group of people
- A new labelled database containing images of groups of people
- Top-down component – scene descriptors
- Bottom-up component – face analysis
- MKL based fusion framework

Future Work

- Extension to videos
- Adding body pose information
- Adding intensities to data to emulate valence-arousal labelling
- Database extension (currently 800 images)

Thank you

Questions?

- EmotiW 2015 Challenge and Workshop at ICMI 2015
  (9-13 Nov 2015, Seattle)
  http://icmi.acm.org/2015/index.php?id=challenges
- Hiring three Assistant Professors at the University of Canberra