Group-level Arousal and Valence Recognition in Static Images: Face, Body and Context

Wenxuan Mou, Oya Celiktutan, Hatice Gunes

Queen Mary University of London, UK

Motivation

Queen Mary University of London

Learning context

Bored??

Retail context

Like???

Enjoy????

Image Retrieval

Related works

Queen Mary University of London

From individual emotion to group emotion

- Group Happiness Intensity (Dhall et al. 2012)

Related works

Queen Mary University of London

	Data	Features	Classifier	Categories	No of Labelers
Dhall et.al FG'15	417 images	**Face** -Facial action unit -PHOG+LPQ **Scene context**	Multiple kernel learning	**Valence** Positive, neutral, negative	3/image
Ours	250 images	**Face** -Geometric -Local Appearance -Global Appearance **Body HOG** **Context**	KNN	**Arousal** High, medium, low **Valence** Positive, neutral, negative	15/image

Data

Queen Mary University of London

Our requirements

- Various emotions in the whole database
- A group of people on each image

Existing databases

- Static Facial Expressions in the Wild (SFEW)[1]
  - Various emotions
  - Single person
- Annotated Face in the Wild(AFW)[2]
  - Multiple people
  - No various emotions, 90% happy, no emotion annotation
- HAPpy PEople Images (HAPPEI) [3]
  - Multiple people
  - All are happy

[1] Dhall, et al. "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark." 2011.
[2] Zhu, et al"Face detection, pose estimation, and landmark localization in the wild.", 2012.
[3] Dhall, et al. Finding happiest moments in a social context, ACCV 2012.

Data Collection and Annotation

Queen Mary University of London

Our database

- Collected images from Internet
- Each image was annotated by 15 raters along Arousal-Valence

Data Collection and Annotation

Queen Mary University of London

Our database

- Collected images from Internet
- Each image was annotated by 15 raters along Arousal-Valence

Data Collection and Annotation

Queen Mary University of London

Annotation questionnaire

Valence

- Positive
- Neutral
- Negative

Activity

- High
- Medium
- Low

Data Collection and Annotation

Queen Mary University of London

Annotation questionnaire

Valence

- Positive
- Neutral
- Negative

Activity

- High
- Medium
- Low

Inter-rater agreement

	Cronbach's α
Arousal	0.85
Valence	0.96

Theories for group emotion

Queen Mary University of London

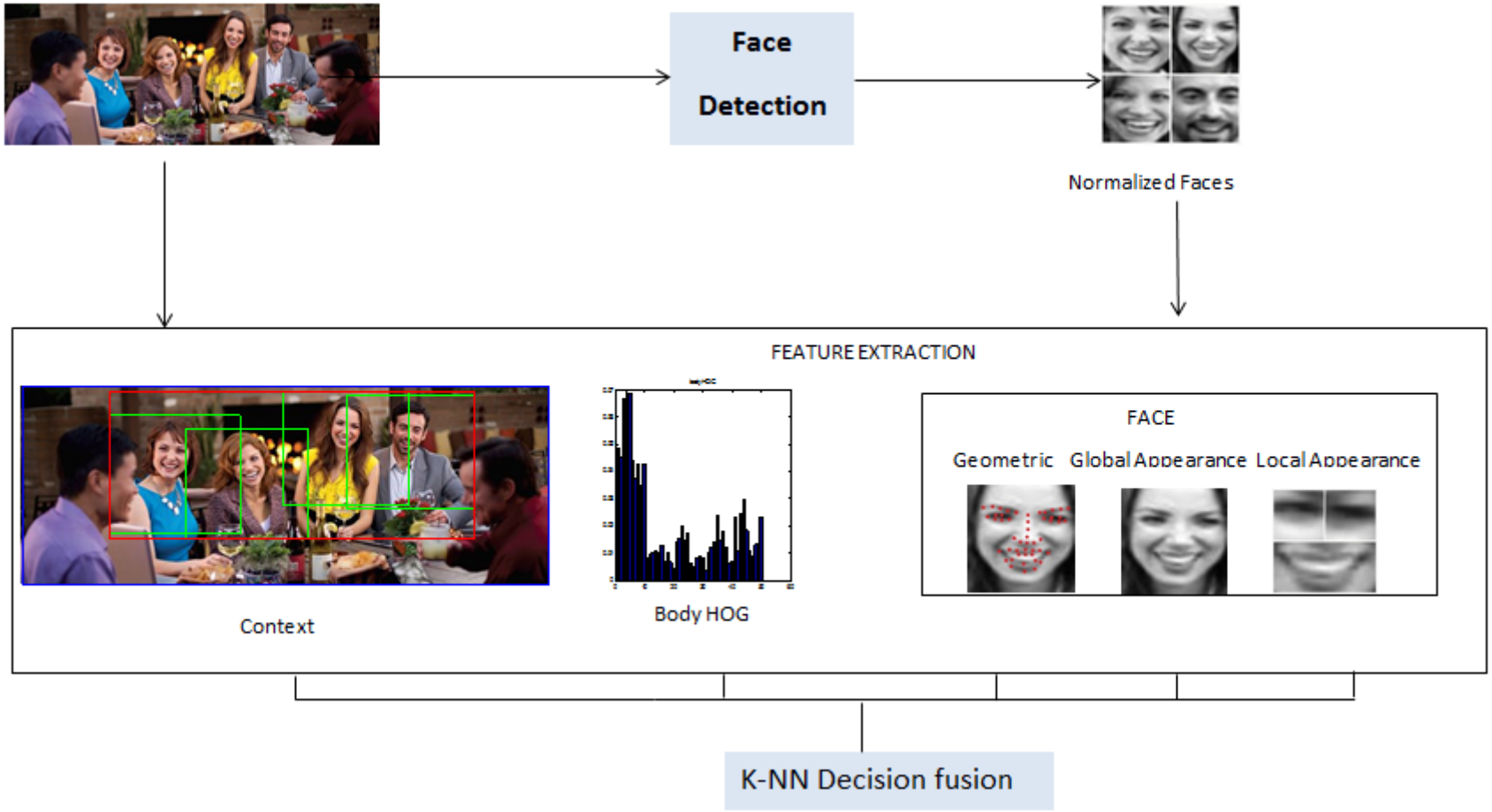
Top-down

- group-as-a-whole
- use group-level information

Bottom-up

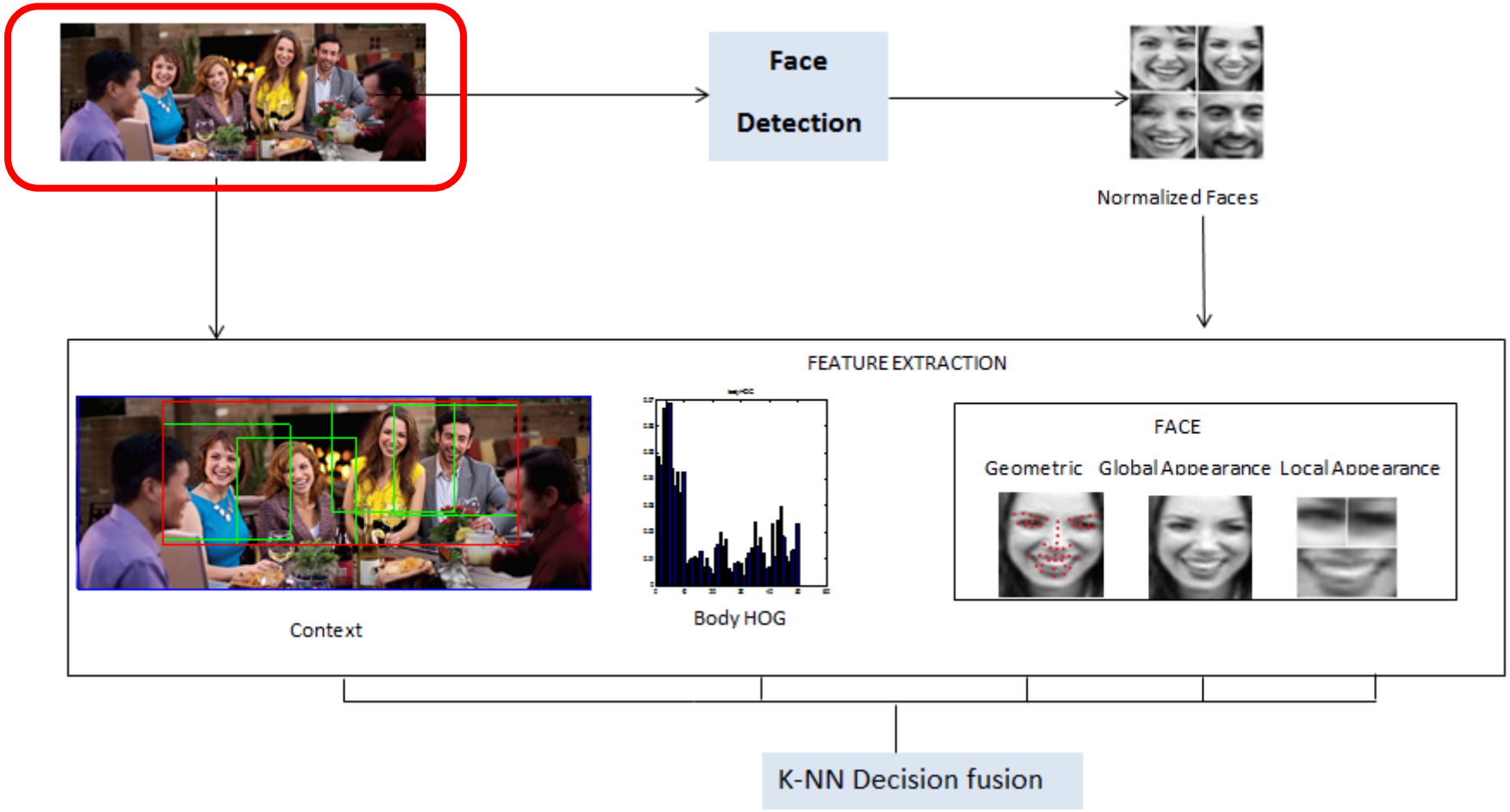
- group-as-sum-of-its-parts
- start with individuals

Our Methodology

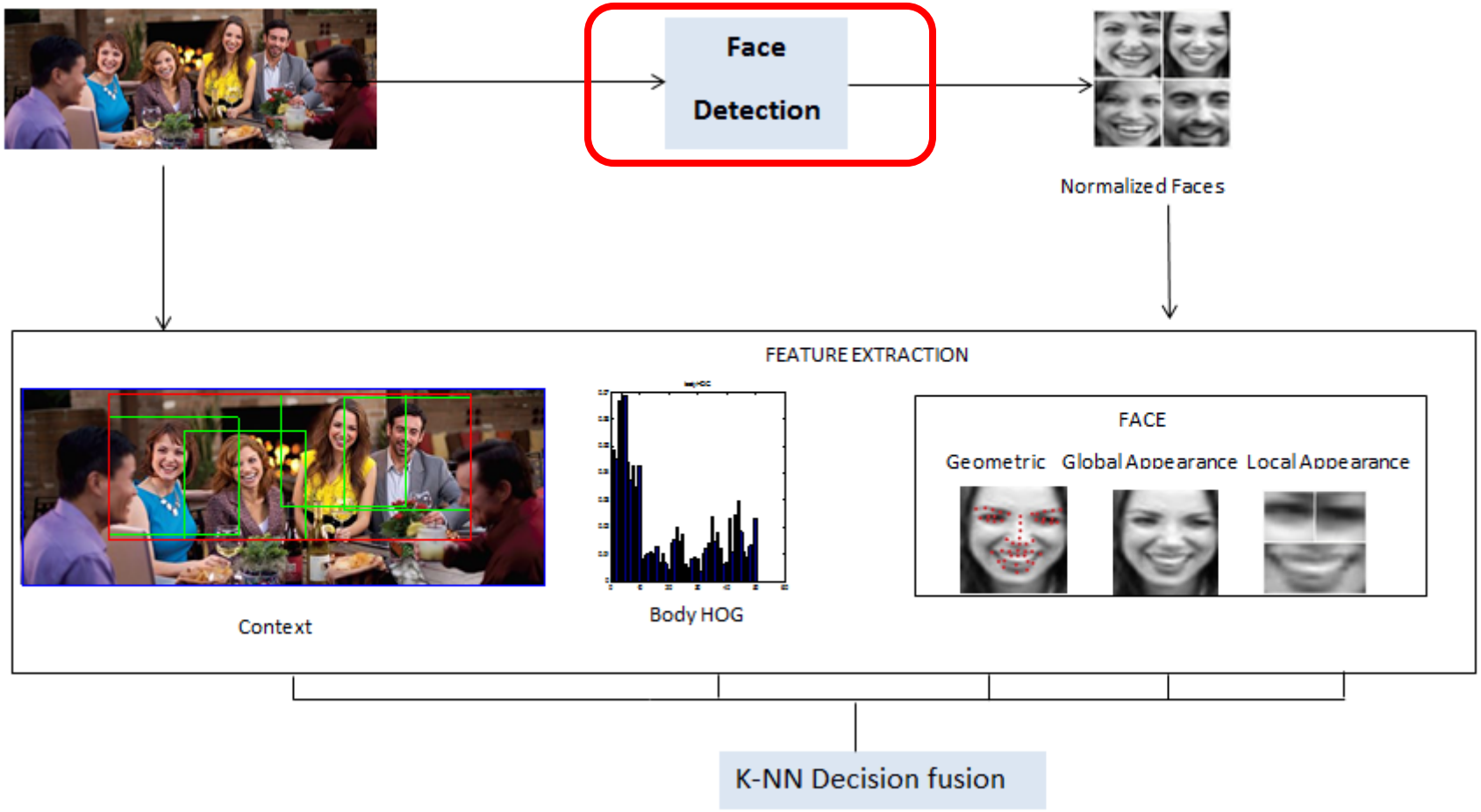


Our Methodology

Queen Mary University of London

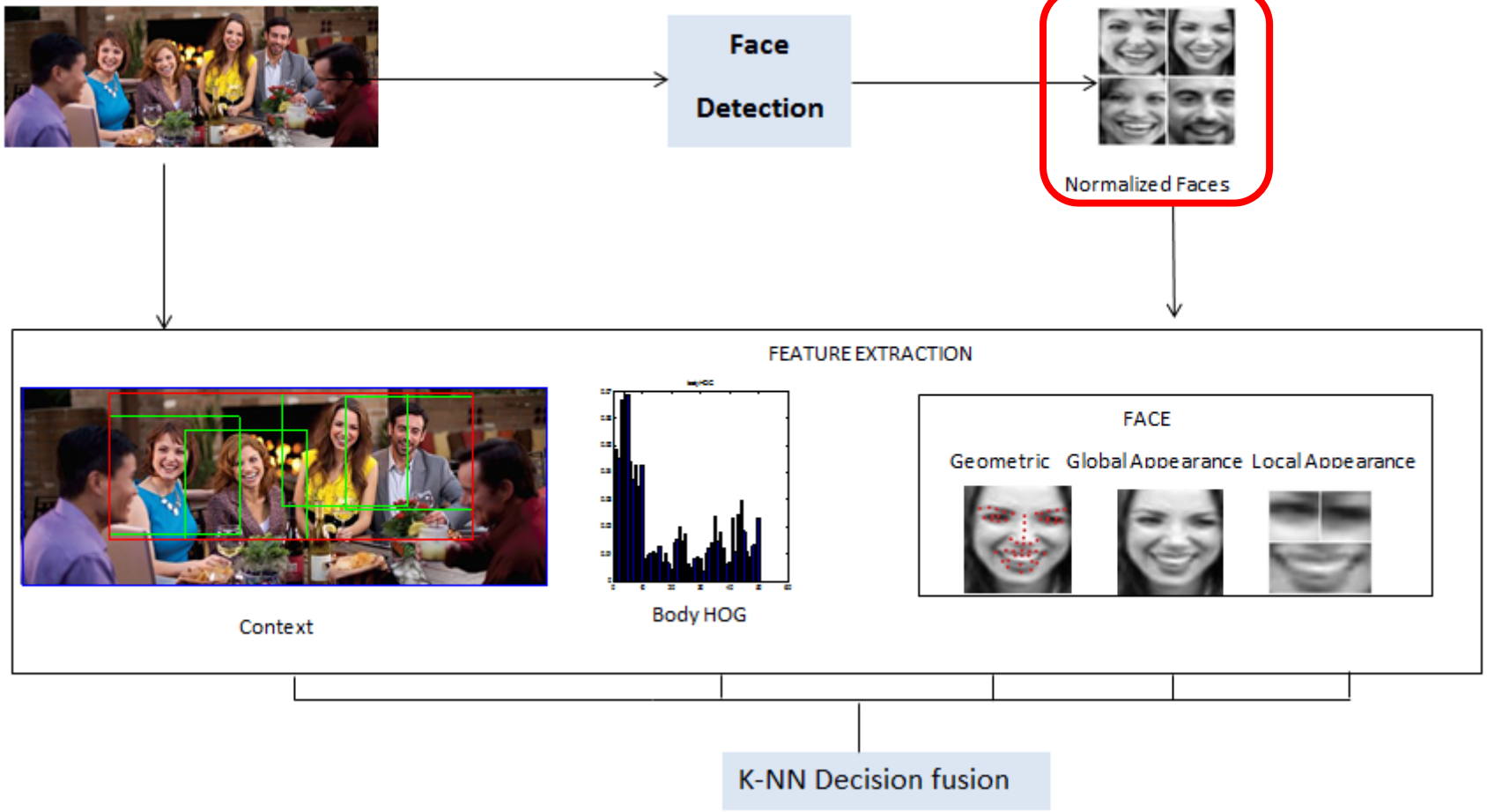


Our Methodology

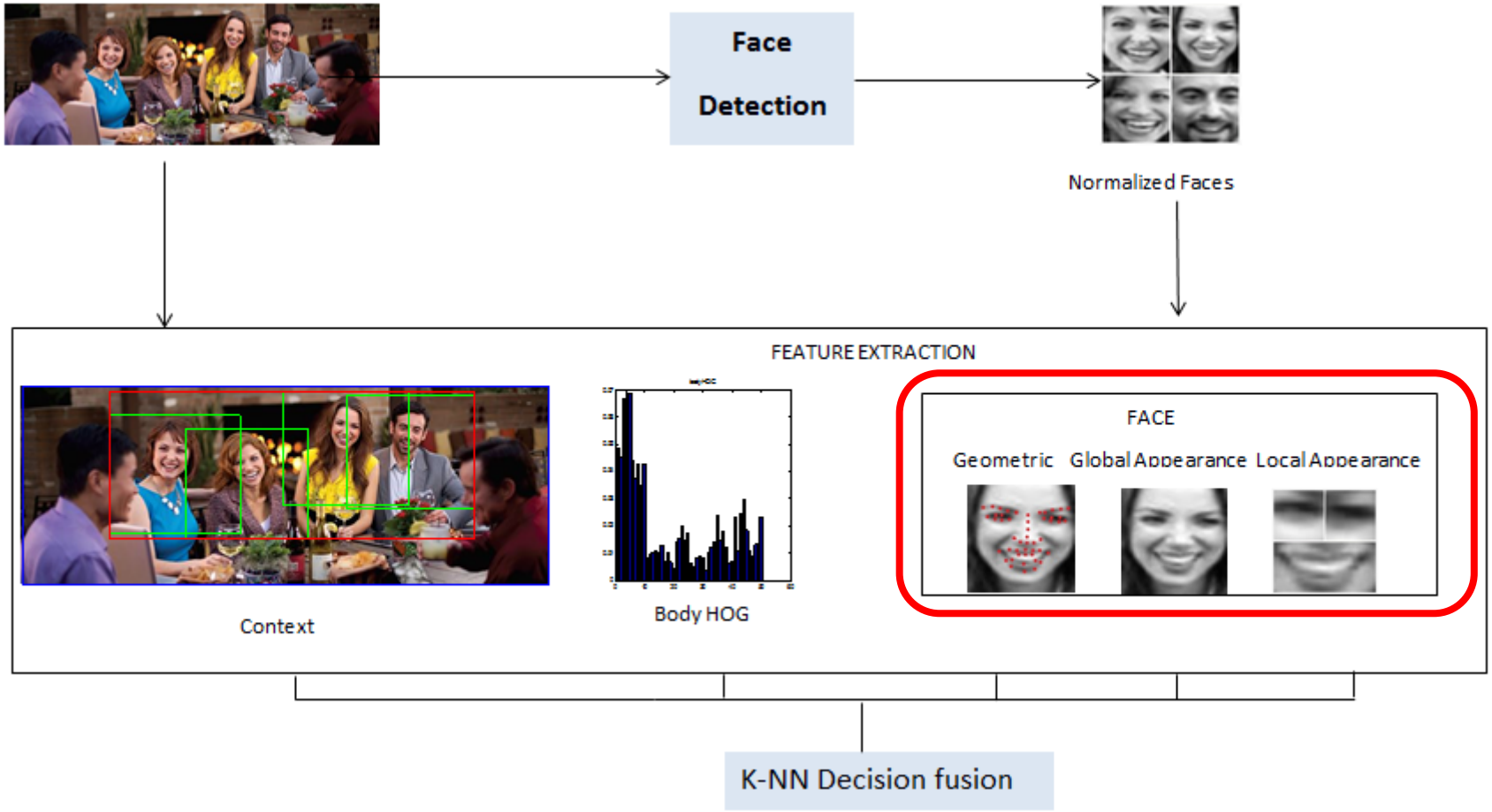


Our Methodology

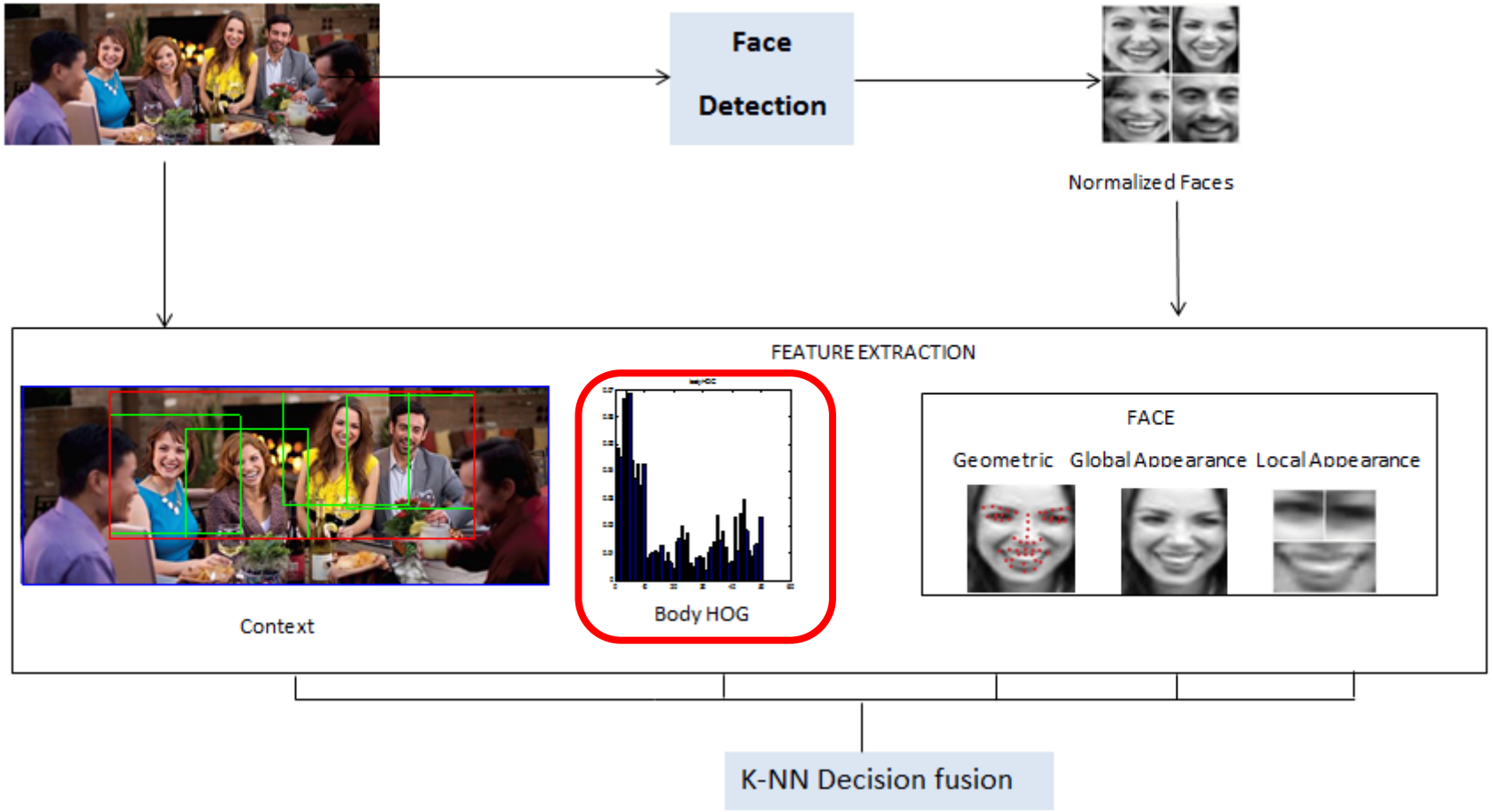
Queen Mary University of London



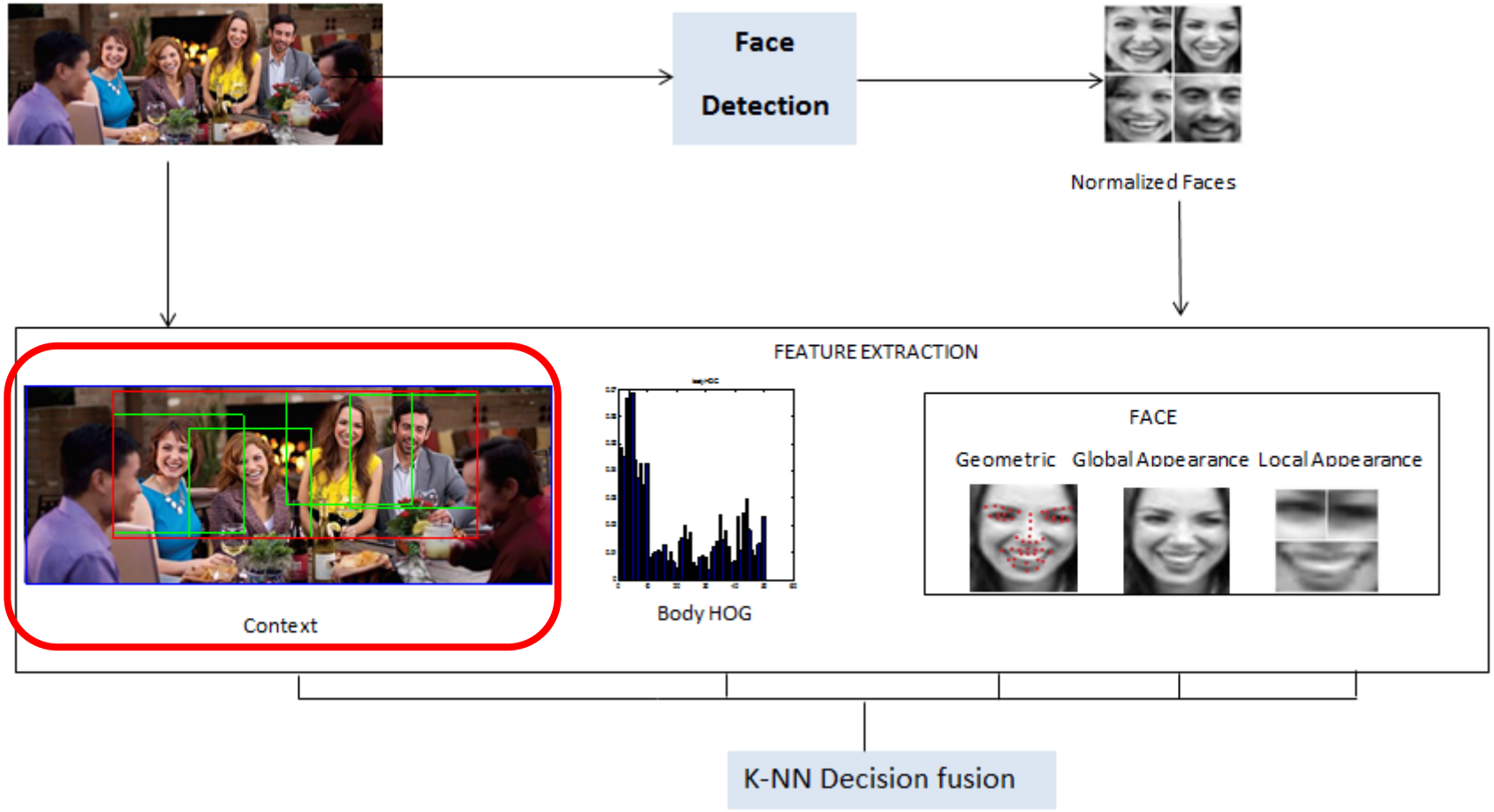
Our Methodology



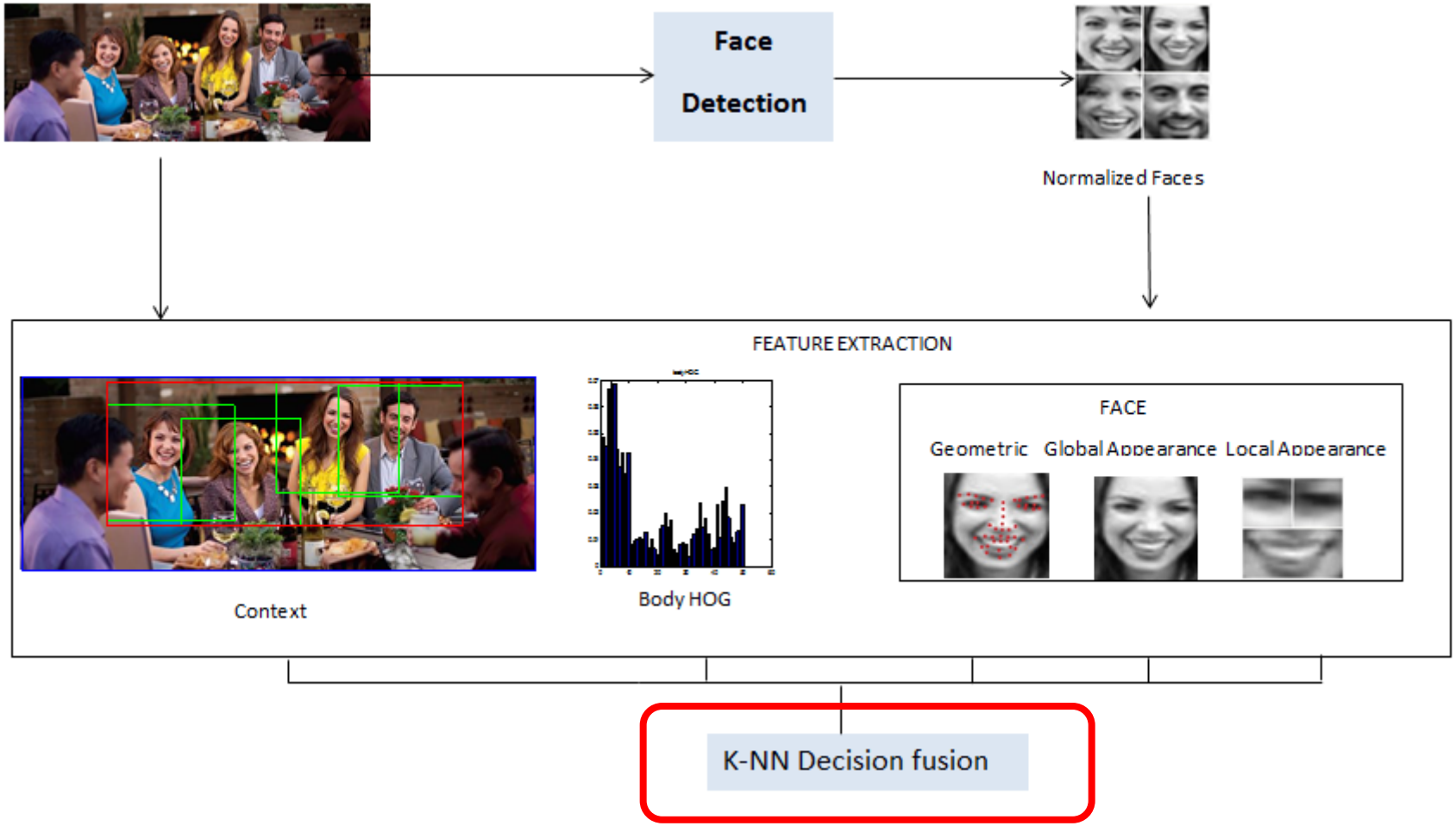
Our Methodology



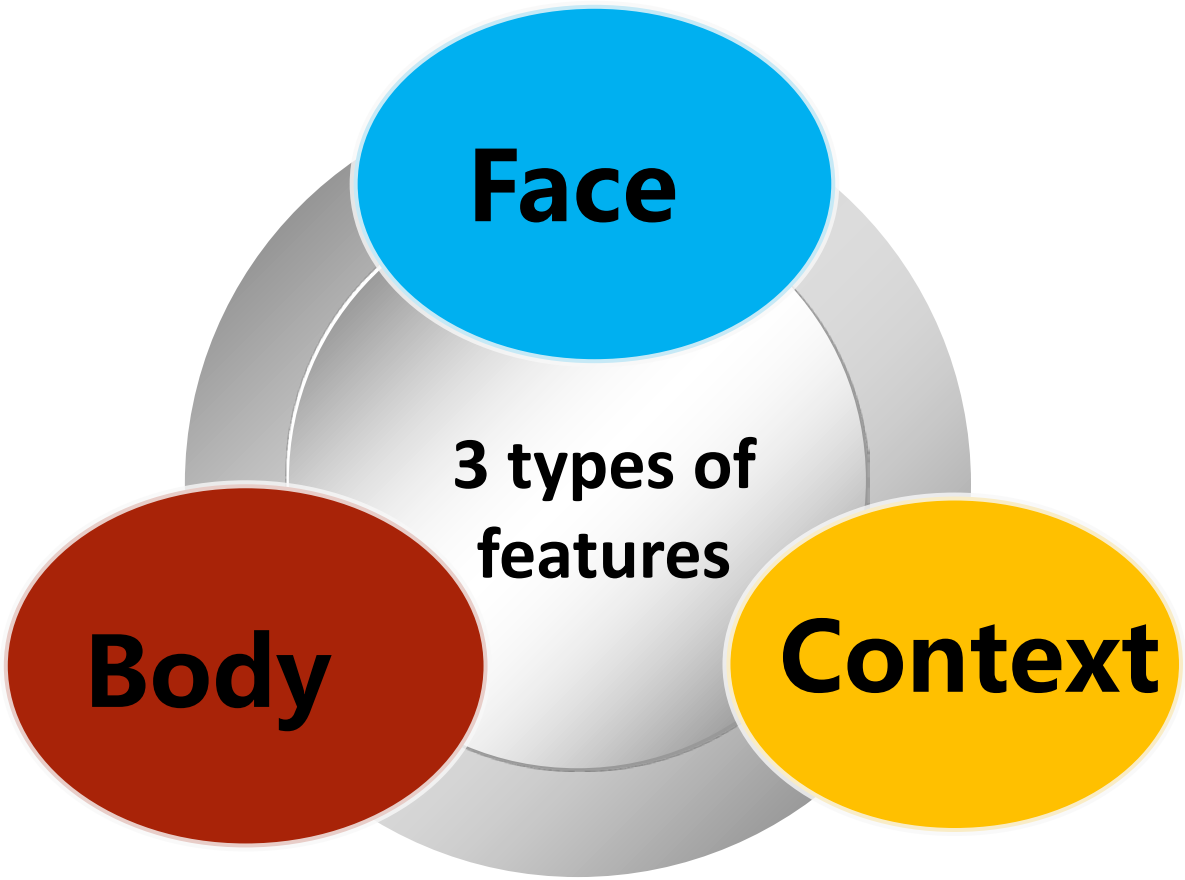
Our Methodology



Our Methodology



Feature Extraction

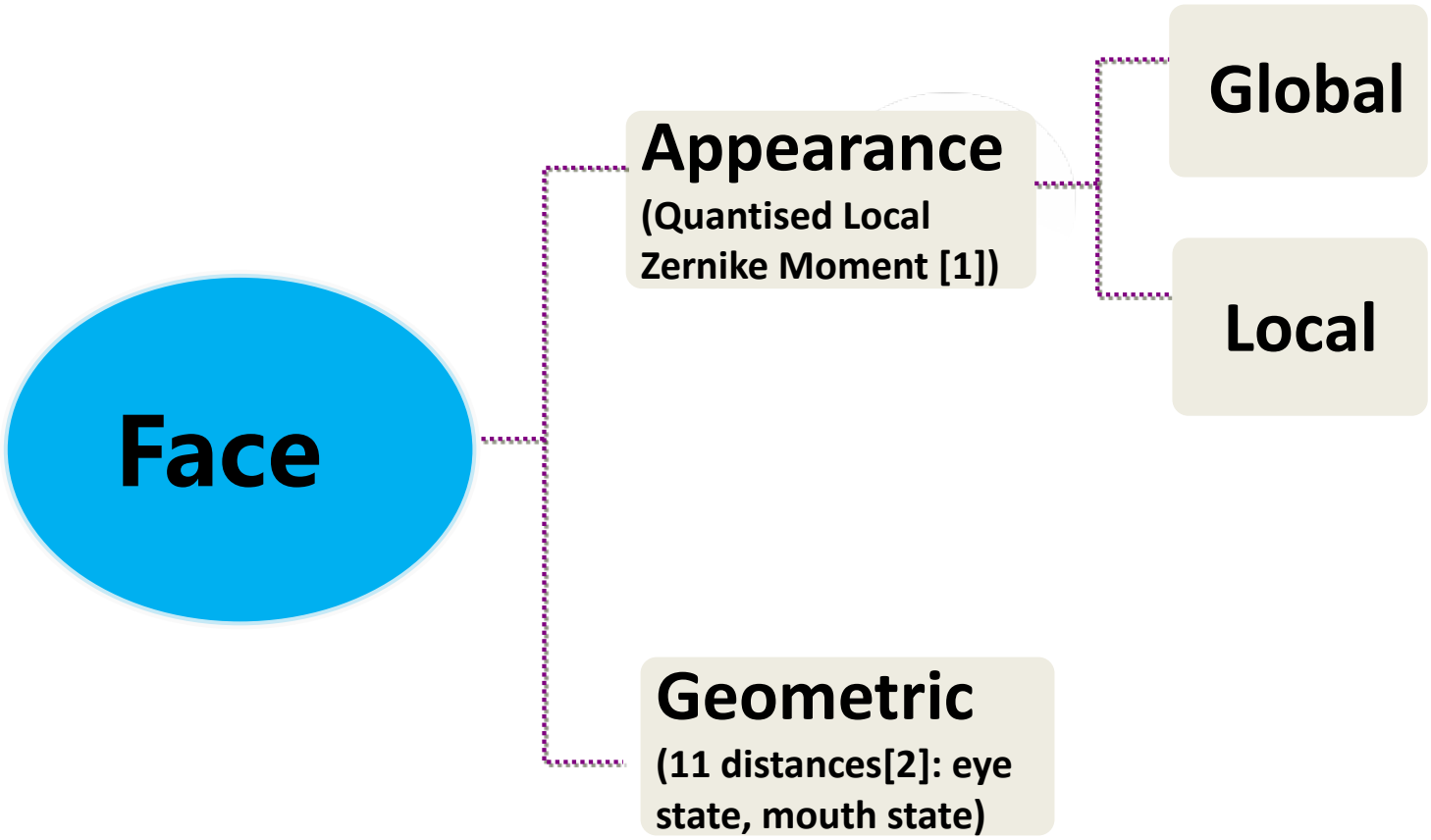


Feature Extraction

Queen Mary University of London

Face

Feature Extraction



[1] Sariyanidi, et al. "Local Zernike Moment Representation for Facial Affect Recognition." 2013.
[2] Akakın, et al. "Robust classification of face and head gestures in video." 2011.

Feature Extraction

- **Face**
  - **Appearance** (Quantised Local Zernike Moment [1])
    - **Global**
    - **Local**
  - **Geometric** (11 distances[2]: eye state, mouth state)

[1] Sariyanidi, et al. "Local Zernike Moment Representation for Facial Affect Recognition." 2013.
[2] Akakın, et al. "Robust classification of face and head gestures in video." 2011.

Feature Extraction

Face

Appearance
(Quantised Local Zernike Moment [1])

Global

Local

Geometric
(11 distances[2]: eye state, mouth state)

[1] Sariyanidi, et al. "Local Zernike Moment Representation for Facial Affect Recognition." 2013.
[2] Akakın, et al. "Robust classification of face and head gestures in video." 2011.

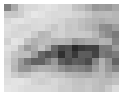
Feature Extraction

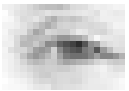
Face

Appearance
(Quantised Local Zernike Moment [1])

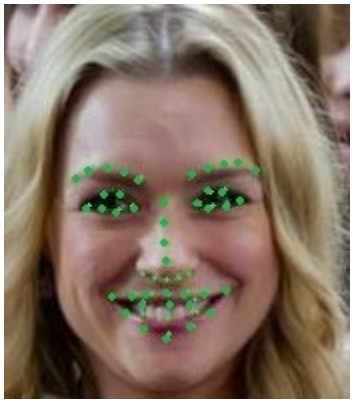
Global

Local



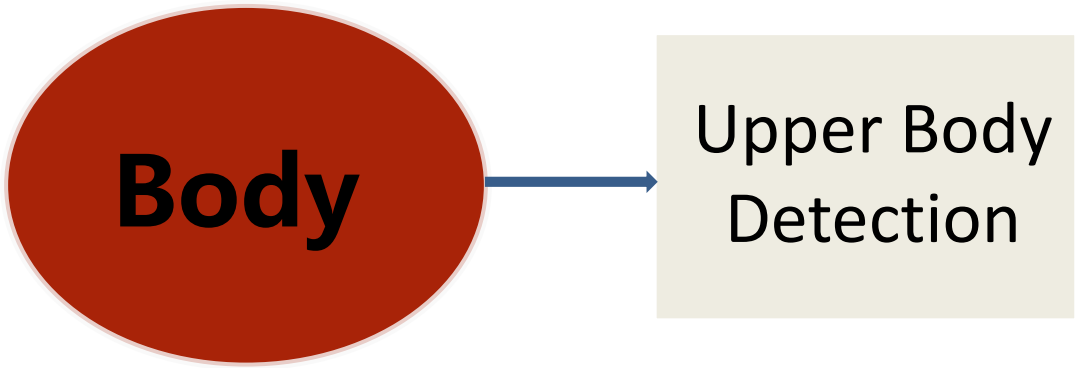


Geometric
(11 distances[2]: eye state, mouth state)



[1] Sariyanidi, et al. "Local Zernike Moment Representation for Facial Affect Recognition." 2013.
[2] Akakın, et al. "Robust classification of face and head gestures in video." 2011.

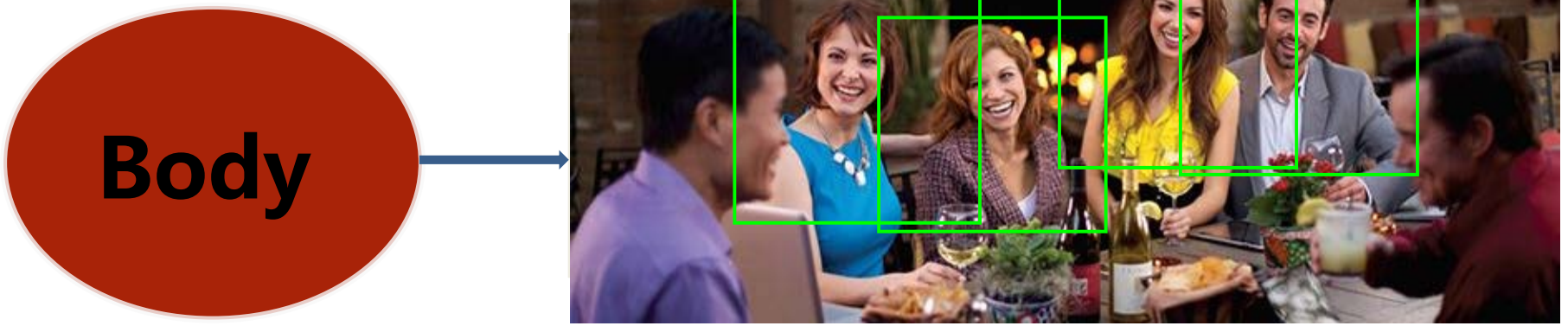
Feature Extraction



[1] Bosch, et al. "Representing shape with a spatial pyramid kernel." *Proceedings of the 6th ACM international conference on Image and video retrieval*. ACM, 2007.

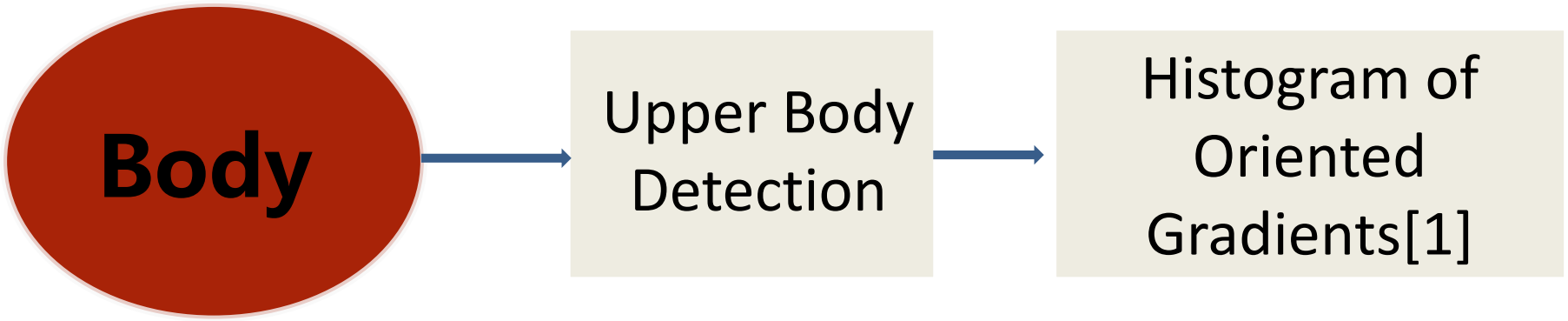
Feature Extraction

Queen Mary University of London



[1] Bosch, et al. "Representing shape with a spatial pyramid kernel." *Proceedings of the 6th ACM international conference on Image and video retrieval*. ACM, 2007.

Feature Extraction



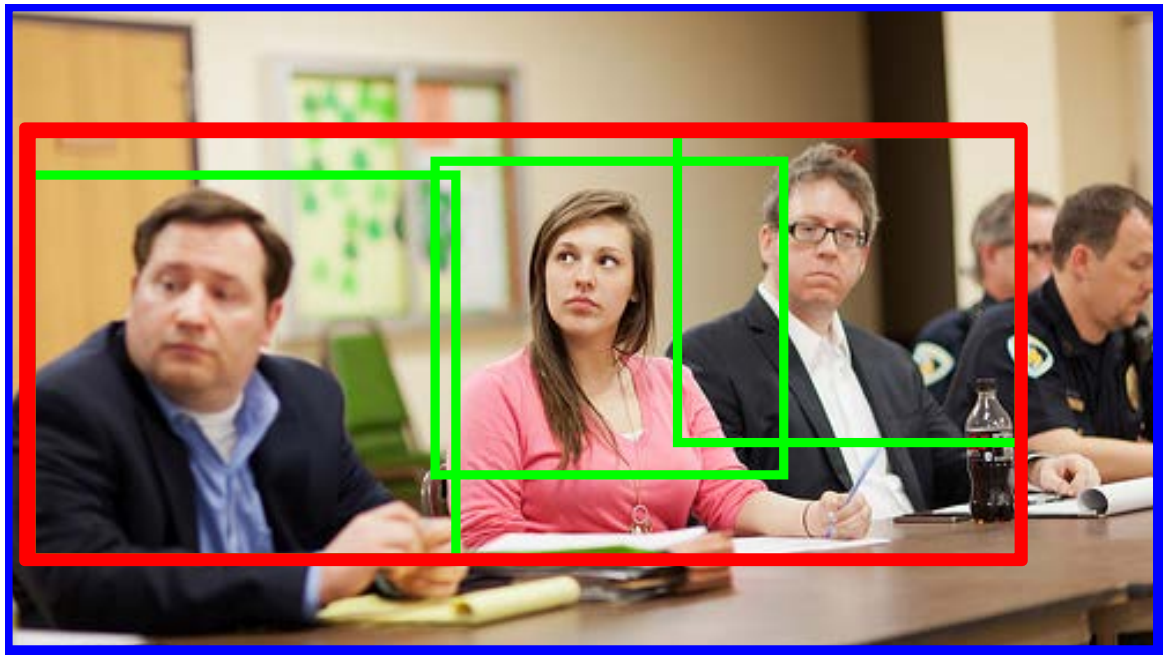
[1] Bosch, et al. "Representing shape with a spatial pyramid kernel." *Proceedings of the 6th ACM international conference on Image and video retrieval*. ACM, 2007.

Feature Extraction

Queen Mary University of London

Context

1. Bodies' relative scale and location w.r.t bounding box
2. Bounding box's relative scale and location w.r.t Image

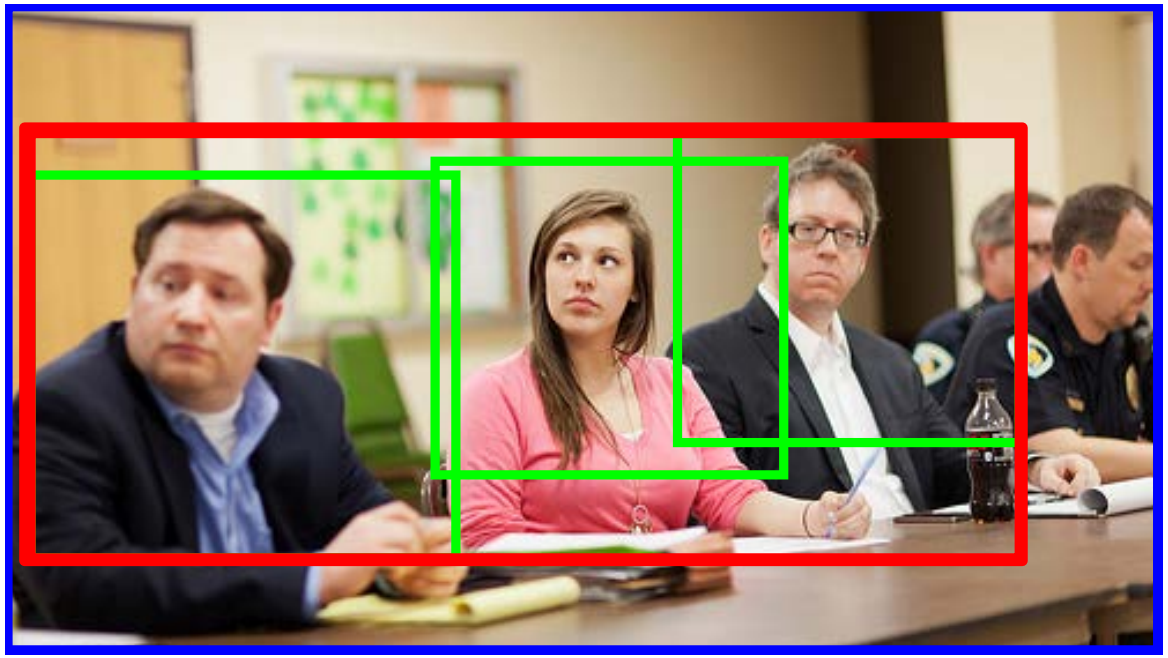


Feature Extraction

Queen Mary University of London

Context

1. Bodies' relative scale and location w.r.t bounding box
2. Bounding box's relative scale and location w.r.t Image



Experiment and Analysis

Queen Mary University of London

Representation divided images into 3 groups based on number of faces, **2** faces**, 3** faces and **4-more** faces.

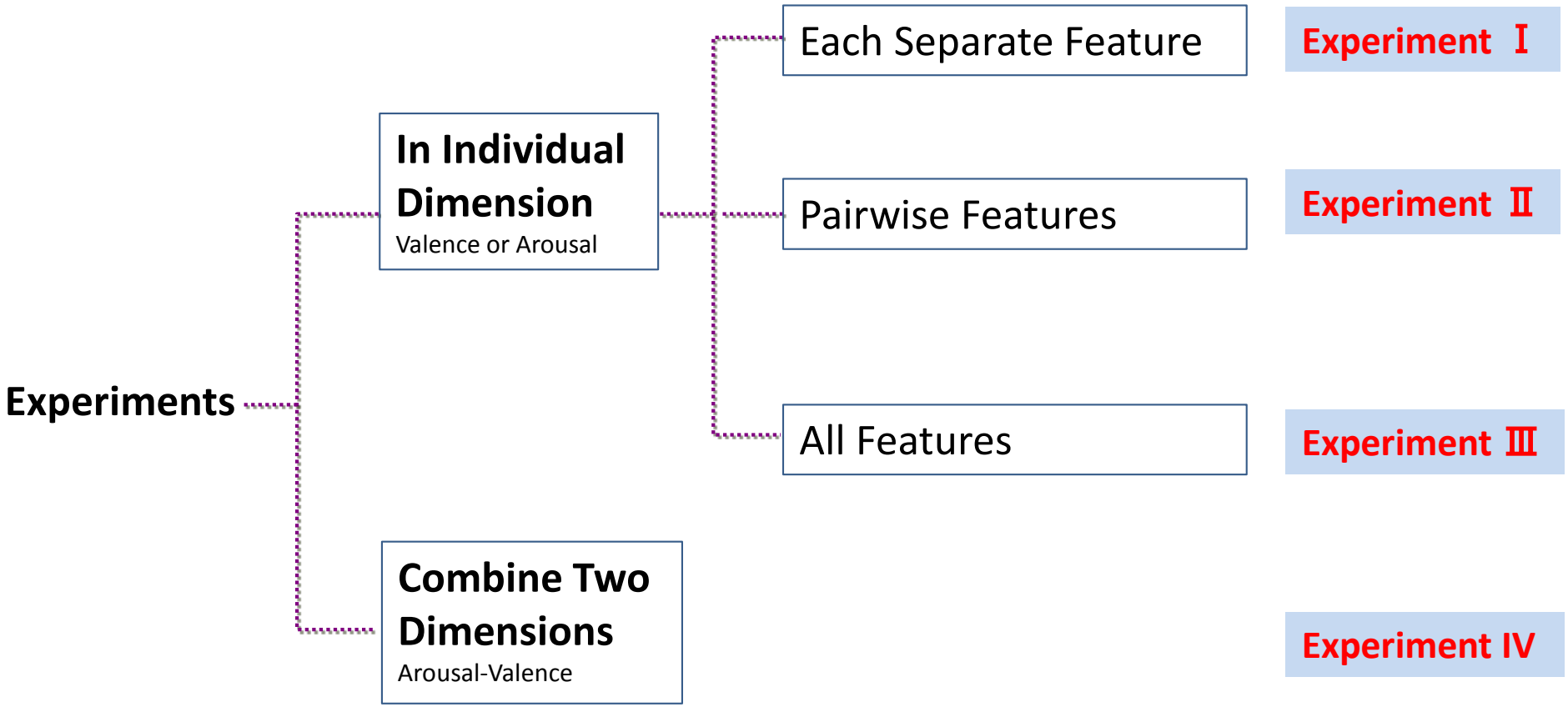
Classifier K-Nearest Neighbor

Experiment and Analysis

Queen Mary University of London

Representation divided images into 3 groups based on number of faces, **2** faces**, 3** faces and **4-more** faces.

Classifier K-Nearest Neighbor



Experiment and Analysis

Queen Mary University of London

Experiment Ⅰ: Different features independently

Faces / %	2 faces		3 faces		>= 4 faces		Mean	
	valence	arousal	valence	arousal	valence	arousal	valence	arousal
Geometric	**63**	56	48	55	**48**	49	**53**	**53**
Global QLZM	53	**57**	46	40	44	**52**	47	49
Part QLZM	37	33	**69**	**61**	28	43	44	45
Context	39	35	30	33	44	35	37	34
Body HOG	38	23	33	35	37	38	35	32

Experiment and Analysis

Queen Mary University of London

Experiment Ⅱ: Pairwise features decision fusion

V: valence A: Arousal

Geometric	Global QLZM	Local QLZM	Context	Body HOG	2 faces / %		3 faces / %		>=4faces / %		Mean / %	
Face features					V	A	V	A	V	A	V	A
√	√				**55**	**57**	43	57	51	52	**49**	**55**
√		√			51	47	**56**	59	41	47	**49**	51
√			√		43	40	53	53	43	49	46	47
√				√	46	44	46	**62**	46	45	46	50
	√		√		39	36	38	33	**51**	33	42	34
	√			√	40	32	41	35	46	40	42	35
		√	√		36	30	33	31	53	40	40	33
		√		√	39	32	41	38	38	43	39	37
			√	√	31	40	33	31	43	35	35	35
	√	√			40	44	**56**	53	41	**56**	45	51

Experiment and Analysis

Queen Mary University of London

Experiment Ⅲ: All features with decision fusion

V: valence A: Arousal

Face features	Context	Body HOG	2 faces / %		3 faces / %		>=4faces / %		Mean / %	
			V	A	V	A	V	A	V	A
√			47	**51**	53	51	44	54	48	52
√	√		47	44	48	57	57	**59**	50	**53**
√		√	**55**	49	56	46	48	56	53	50
√	√	√	45	40	**58**	**64**	**60**	50	**54**	51

Face features: geometric, local appearance, global appearance

Experiment and Analysis

Queen Mary University of London

Experiment results IV: Combined Dimension

Face features	Context	Body HOG	2 faces / %	3 faces / %	>=4faces / %	Mean / %
√			28.33	40	31.11	33.15
	√		21.6	8	20	16.53
		√	18.33	20	26.67	21.67
√	√		**31.67**	40	**44.44**	**38.70**
√		√	28.33	**56**	35.56	**39.96**
√	√	√	28.33	44	35.56	**35.96**

V: valence A: Arousal

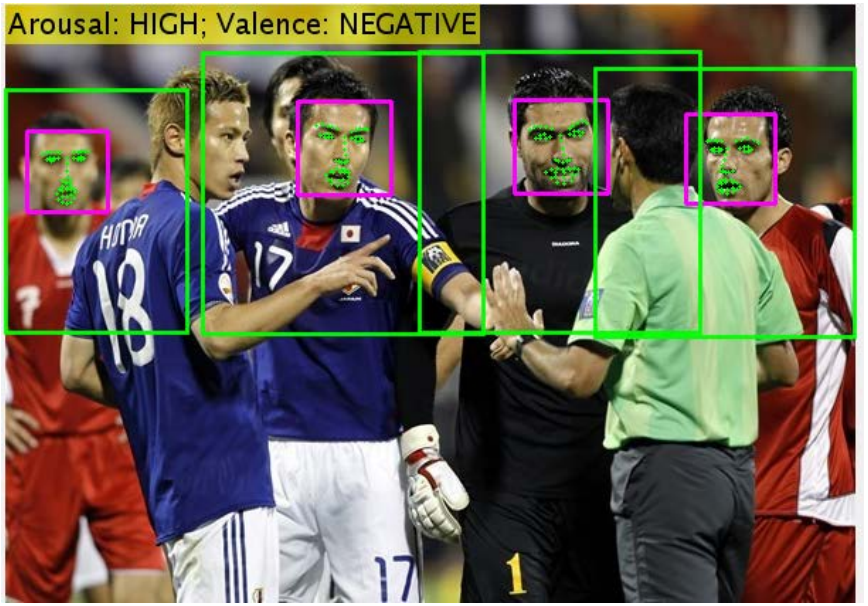
Face features: geometric, local appearance, global appearance

Representative Results

Representative Results

Representative Results

Queen Mary University of London



Conclusion and Future Work

Queen Mary University of London

Conclusion

1. A new database for group emotion detection
2. Automatic group emotion detection along Arousal and Valence dimensions
3. A context feature and body-HOG feature used for group emotion detection

Conclusion and Future Work

Queen Mary University of London

Conclusion

1. A new database for group emotion detection
2. Automatic group emotion detection along Arousal and Valence dimensions
3. A context feature and body-HOG feature used for group emotion detection

Future work

1. Feature selection before fusion and classification
2. Experiment with more sophisticated classifiers
3. Test on a larger database

Thank you!

Questions?