Engagement Estimation for an E-Learning Environment Application

Win Shwe Sin Khine*, Shinobu Hasegawa, Kazunori Kotani

winshwesinkhine@jaist.ac.jp, hasegawa@jaist.ac.jp, ikko@jaist.ac.jp

School of Information Science

Japan Advanced Institute of Science and Technology

The Thirteenth International Conference on Advances in Computer-Human Interactions (ACHI), November 22nd, 2020

* is the presenter

Presenter's Profile

Win Shwe Sin Khine
Ph.D. Candidate, JAIST
School of Information Science
Computer Vision Lab
B.C.Sc. (2017), M.C.Sc.(2020)
Email: winshwesinkhine@jaist.ac.jp

❑ Research Interests:

- Deep Learning, Machine Learning, Computer Vision, Human Emotion, Human Computer Interaction

❑ Past Experiences:

- 2017: Joined Galaxy Wave Technology Services Co.Ltd. as an internship student for software engineer position
- 2016: Joined Fujitsu ICT Lab – UIT as a trainee
- 2015: Joined OAS Corporation ICT Lab – UIT as a trainee

❑ Publications:

- Generation of Compound Emotions with Emotion Generative Adversarial Networks (EmoGANs), The SICE International Conference, 2020, Chiang Mai, Thailand.
- Engagement Estimation for an E-Learning Environment Application, the Thirteenth International Conference on Advances in Computer-Human Interactions, pp 51-56, March 22, 2020.

Outlines

- Introduction
- Methodology
- Experimental Results
- Conclusion

Introduction

- Starting from 1980s, student engagement becomes a significant concerns because of a large drop out rate, statistically between 20% and 60% according to R.W.Larson et.al [1].

- The reason is the students are extremely bored during lectures.

- Therefore, it is important to keep the good communication with students.

Photo credit: Adikos, creative commons

Introduction

Real classroom

Photo credit: superkimbo, creative commons

- In real environment, lecturers can recognize the students’ emotions through their facial expressions and adjust their teaching methods to improve their engagement levels.

Introduction

Virtual classroom

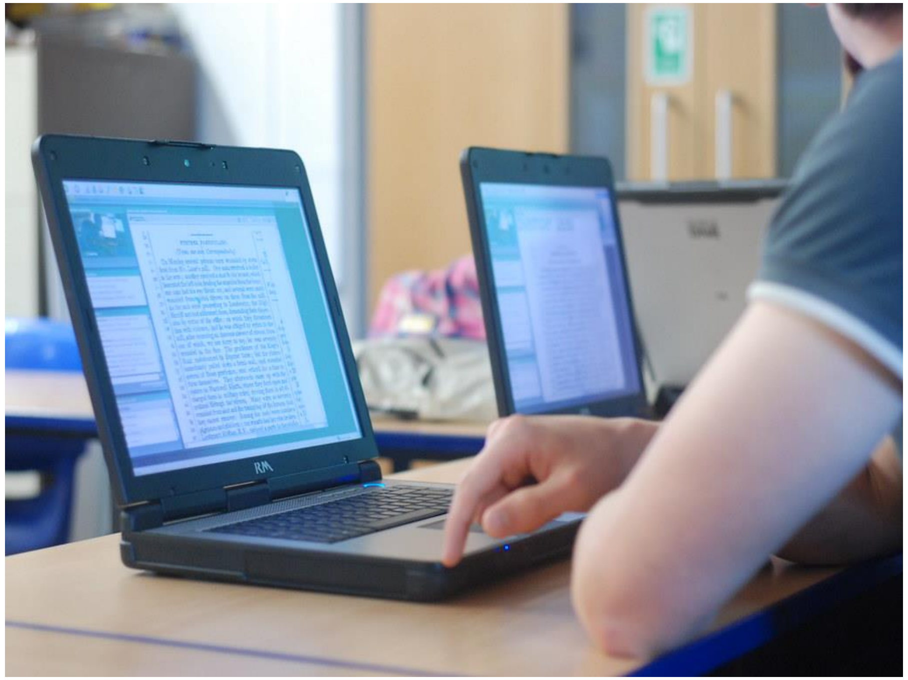


Photo credit: Mr Ush, creative commons

- In a virtual environment, it has difficulties to detect students' emotions because there is no interaction with students.

- The problems of virtual system motivate us to perform automatic engagement detection based on their facial expressions.

Introduction

- The purpose of this study is to make an improvement in virtual learning system and prevent the students to drop out from their lectures by recognition their engagement levels.

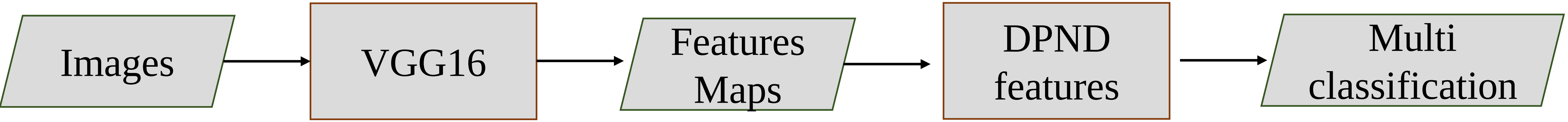
- To realize this purpose, we propose an automated engagement recognition system based on facial expression by using transfer learning technique.

Introduction

Authors	Frameworks	Advantages	Disadvantages
V.Mayya et.al [2]	Deep CNN	Extraction of specific features	Less generalization, Need huge amount of data, Over-fitting
D.K.Jain et.al [3]	Ext-DNN	Extraction of specific features	Less generalization, Need huge amount of data, Over-fitting
M.Sabri et.al [4]	Siamese and triplet Networks	More generalization	Manually selection of apex and onset frames
X.He et.al [5]	B-CNN, E-CNN	Assistant Learning	Poor recognition on less amount of data
J.Chen et.al [6]	DNN, SVM	Avoidance of over-fitting problem	Not end to end mode

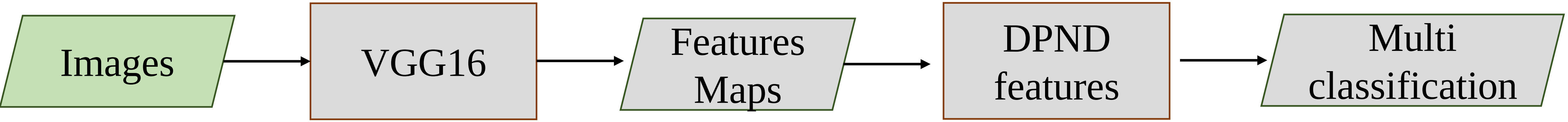
Table 1: Literature Reviews

Methodology



- VGG16 : Pretrained Face Model standing for Visual Geometry Group-16 (O.M.Parkhi et.al. [7])
- DPND = Deep Peak Netural Differences (J.Chen et.al. [8])

Methodology



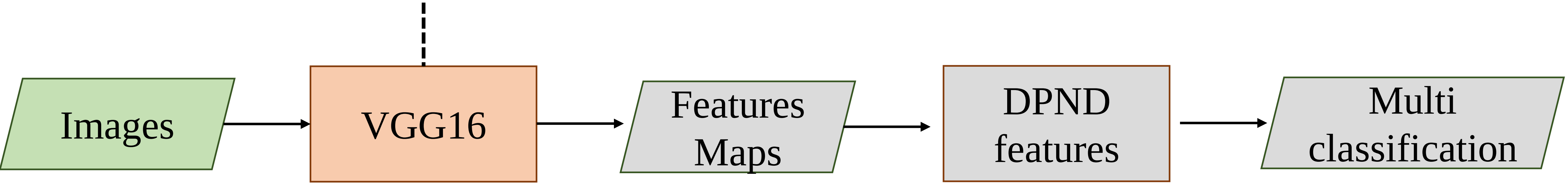
- ❖ VGG16 : Pretrained Face Model standing for Visual Geometry Group-16 (O.M.Parkhi et.al. [7])
- ❖ DPND = Deep Peak Netural Differences (J.Chen et.al. [8])

= Inputs/Outputs

= Process

Methodology

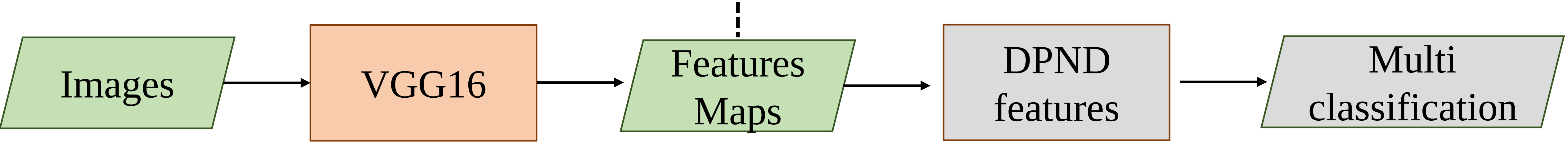
❖ VGG16 model achieved 98% accuracy in face recognition on large scaled dataset



❖ VGG16 : Pretrained Face Model standing for Visual Geometry Group-16 (O.M.Parkhi et.al. [7])
❖ DPND = Deep Peak Netural Differences (J.Chen et.al. [8])

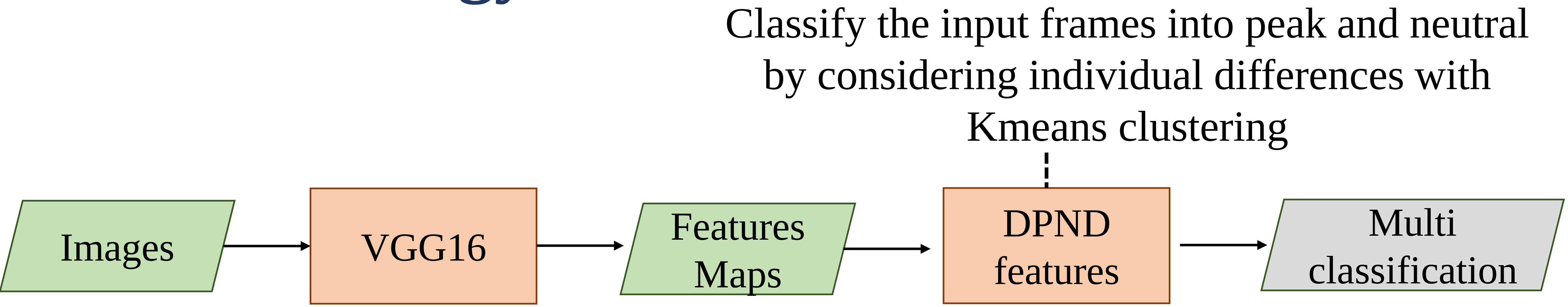
Methodology

- ❖ Extracted features from last two fully connected layers of VGG16 are classified by using Support Vector Machines (SVM) classifiers



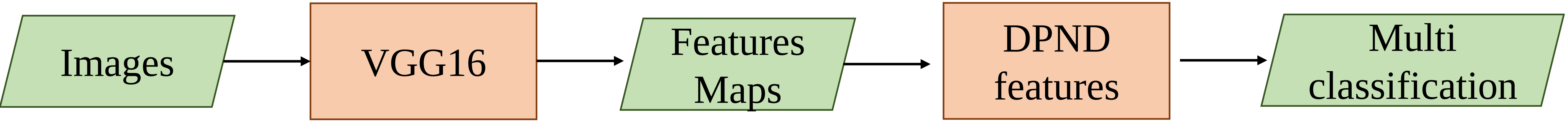
- ❖ VGG16 : Pretrained Face Model standing for Visual Geometry Group-16 (O.M.Parkhi et.al. [7])
- ❖ DPND = Deep Peak Netural Differences (J.Chen et.al. [8])

Methodology



❖ VGG16 : Pretrained Face Model standing for Visual Geometry Group-16 (O.M.Parkhi et.al. [7])
❖ DPND = Deep Peak Netural Differences (J.Chen et.al. [8])

Methodology



- ❖ VGG16 : Pretrained Face Model standing for Visual Geometry Group-16 (O.M.Parkhi et.al. [7])
- ❖ DPND = Deep Peak Netural Differences (J.Chen et.al. [8])

Experimental Results

<u>Dataset</u>

- DAiSEE: Dataset for Affective States in E-Environment [9]
- Includes 9068 videos with 10 seconds duration with 112 subjects.
- Includes four effective states such as Boredom, Confused, Engagement, and Cofusion.
- Indicates different levels of states, ranging from 0 to 3.
    - 0: “Very Low”, 1: “Low”, 2: “High”, 3: “Very High”

Experimental Results

<u>Preprocessing</u>

- Frame Conversion
  - ➢Converts the videos into frames by using FFMPEG
- Frame Selection
  - ➢Selects 0.005% of randomized samples from the original dataset
- Preprocessing of VGG-16
  - ➢Crop 224 patches, horizontally flipped, averages and scale

Experimental Results

Epochs	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	1.5751	0.4728	4.0148	0.4657
2	8.3836	0.4814	13.4151	0.4657
3	13.8168	0.4764	14.5101	0.4765
4	14.3863	0.4748	14.8962	0.4814
5	14.5621	0.4796	14.9514	0.4549
6	14.6212	0.4723	14.8178	0.4941
7	14.7216	0.4719	14.7799	0.4814
8	14.7823	0.4749	14.5372	0.4853
9	14.7558	0.4769	14.7691	0.4843
10	14.8142	0.4748	14.7804	0.4843

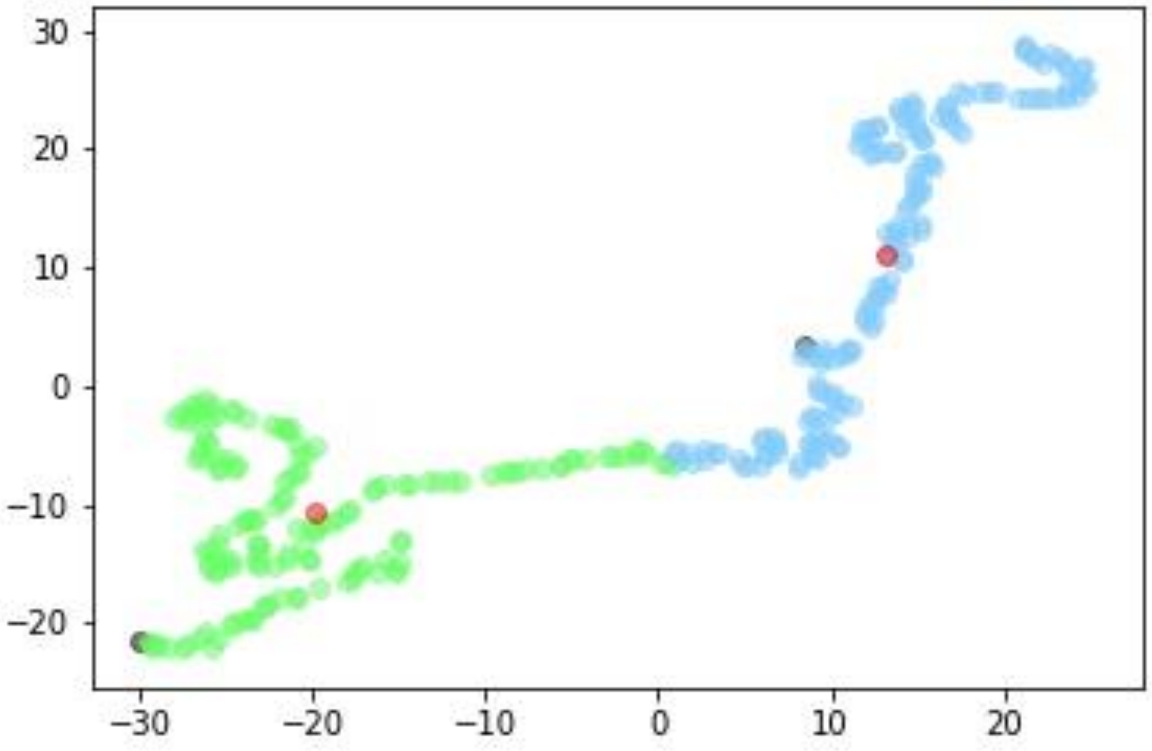
Table 2: Accuracy and loss values for deep representations from **‘fc6’** dense layers by fine-tuning VGG-16 model

Experimental Results

Epochs	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	11.1773	0.4644	13.8237	0.4716
2	12.5062	0.4667	13.7523	0.4637
3	12.4247	0.4700	13.5395	0.4696
4	12.4178	0.4666	13.6752	0.4892
5	13.6556	0.4664	14.6021	0.4716
6	13.9989	0.4658	14.2492	0.4824
7	13.9343	0.4745	14.1655	0.4657
8	14.0053	0.4686	14.2087	0.4745
9	13.9698	0.4690	14.1892	0.4706
10	13.9838	0.4667	14.3439	0.4853

Table 3: Accuracy and loss values for deep representations from **'fc7'** dense layers by fine-tuning VGG-16 model

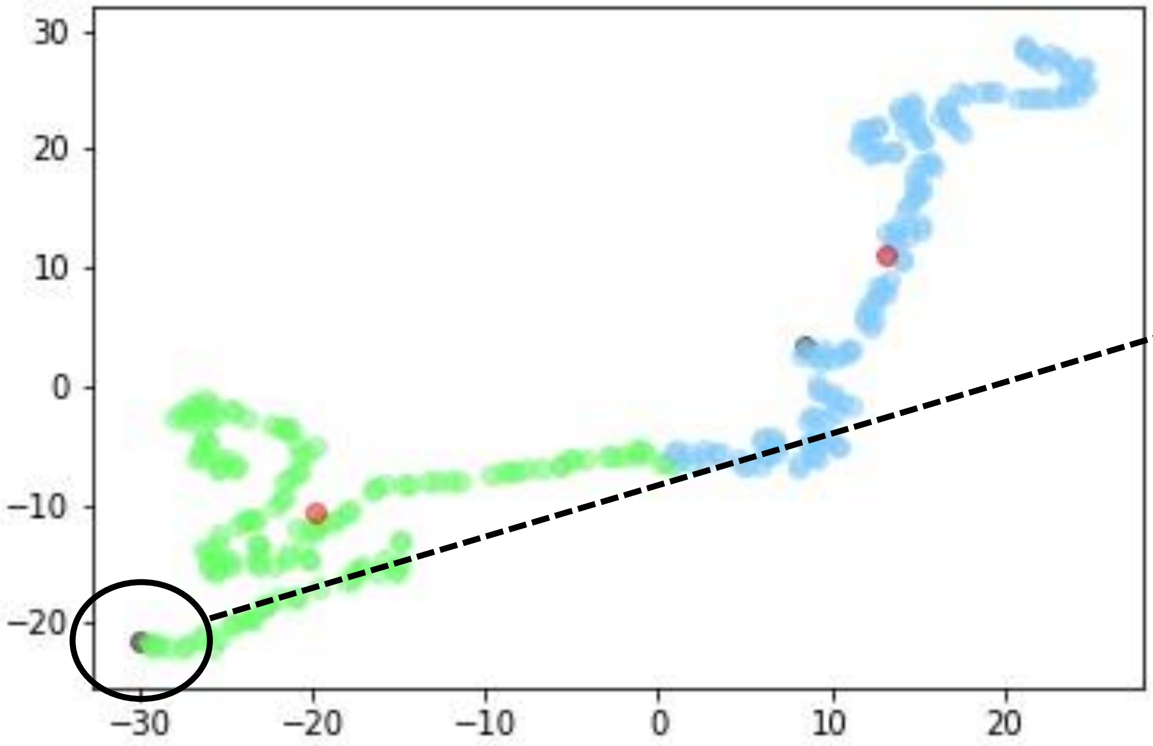
Experimental Results

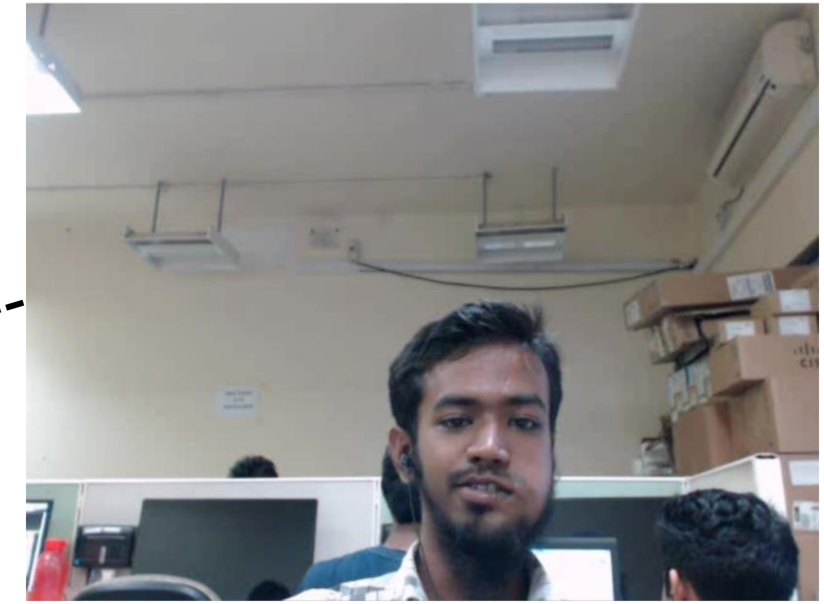


- Red Circle: Centers
- Black Circle: Samples
- Green Circle: Cluster 0
- Blue Circle: Cluster 1

Figure 1: Kmeans clustering results for peak and neutral frames for single person

Experimental Results

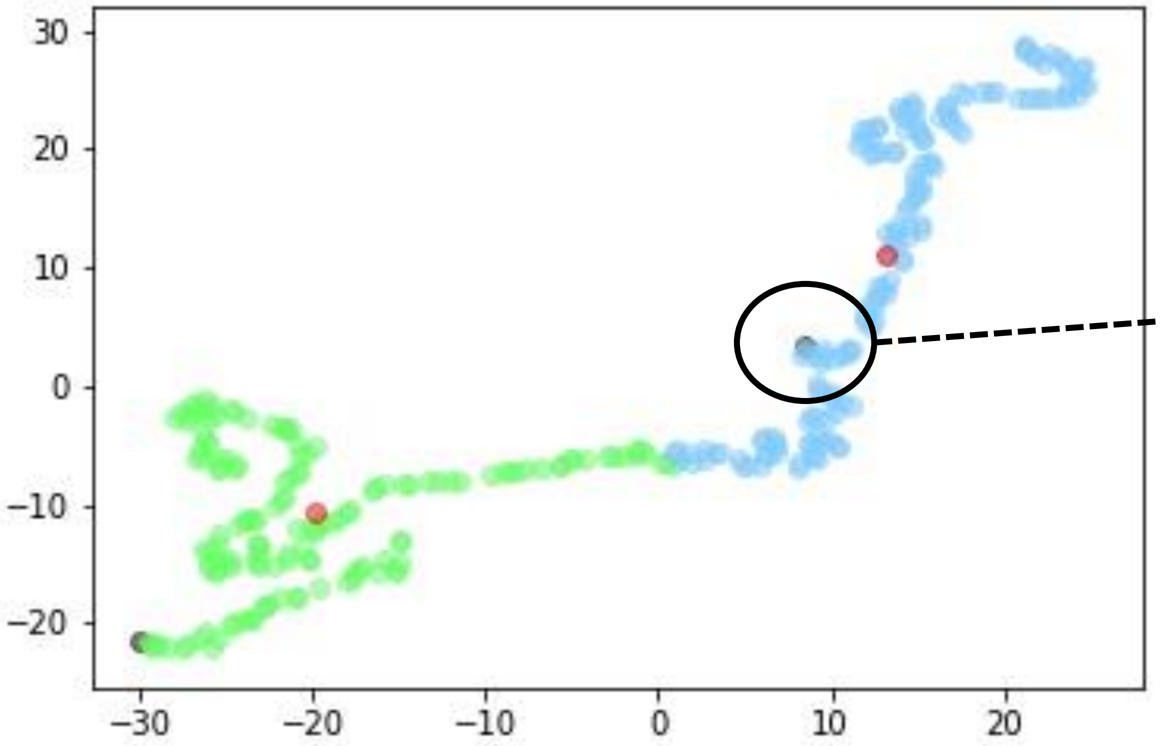


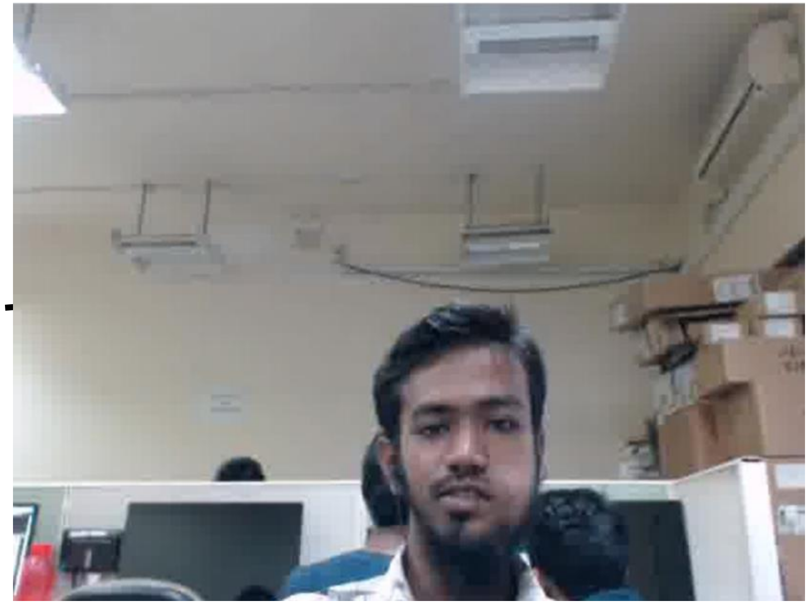


Sample from Cluster 0

Figure 1: Kmeans clustering results for peak and neutral frames for single person

Experimental Results





Sample from Cluster 1

Figure 1: Kmeans clustering results for peak and neutral frames for single person

Experimental Results

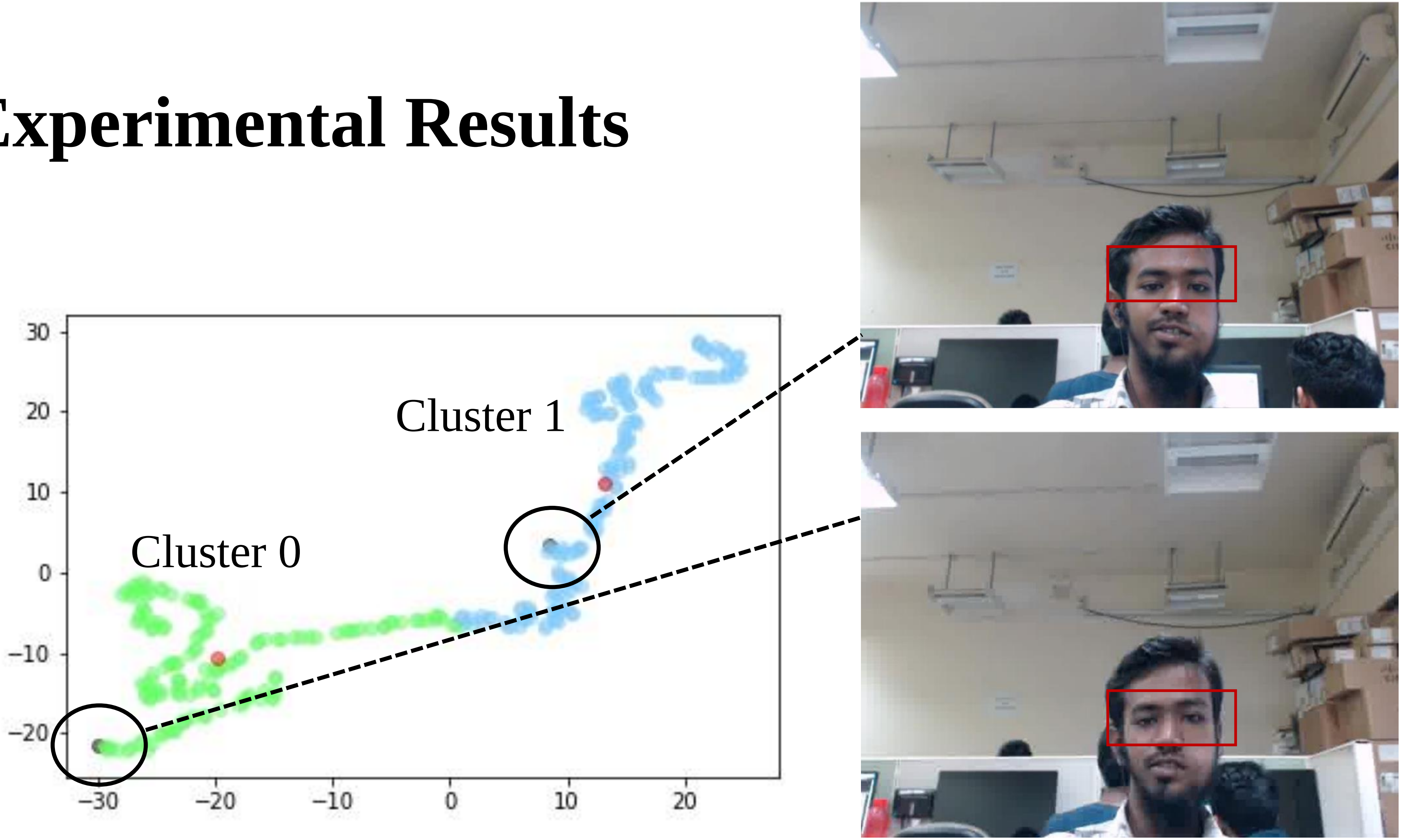
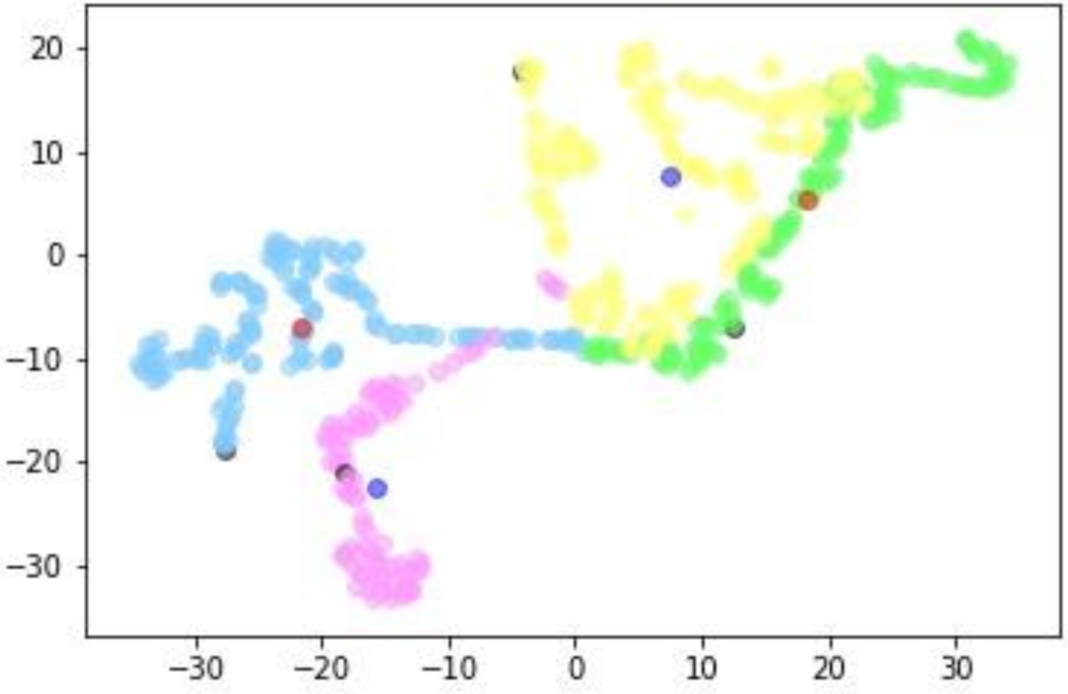


Figure 1: Kmeans clustering results for peak and neutral frames for single person

Experimental Results



- Red or Dark Blue Circle: Centers
- Black Circle: Samples
- Green and Yellow Circle: Cluster 0
- Blue and Pink Circle: Cluster 1

Figure 2: Kmeans clustering results for peak and neutral frames for multiple persons

Experimental Results

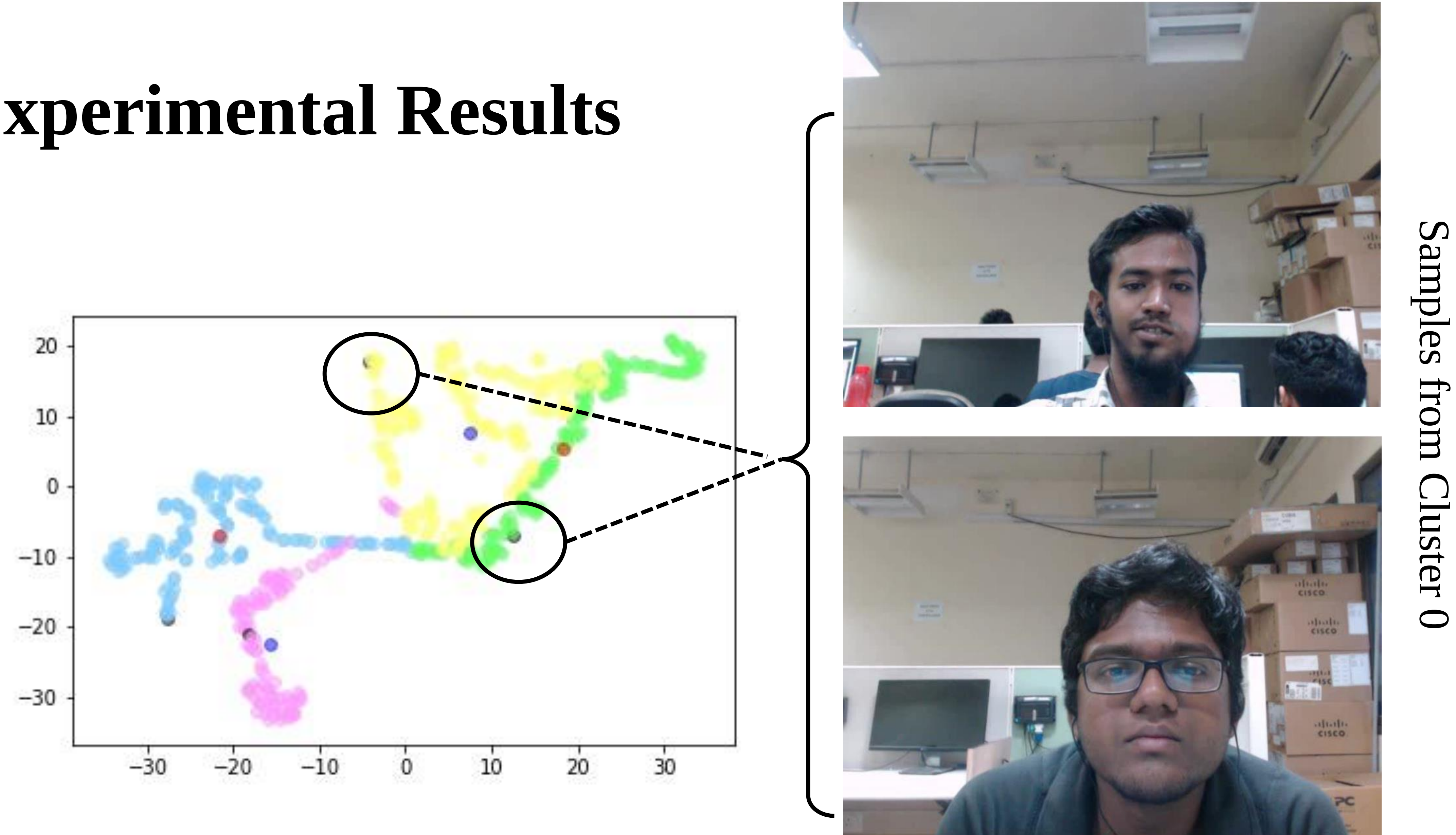
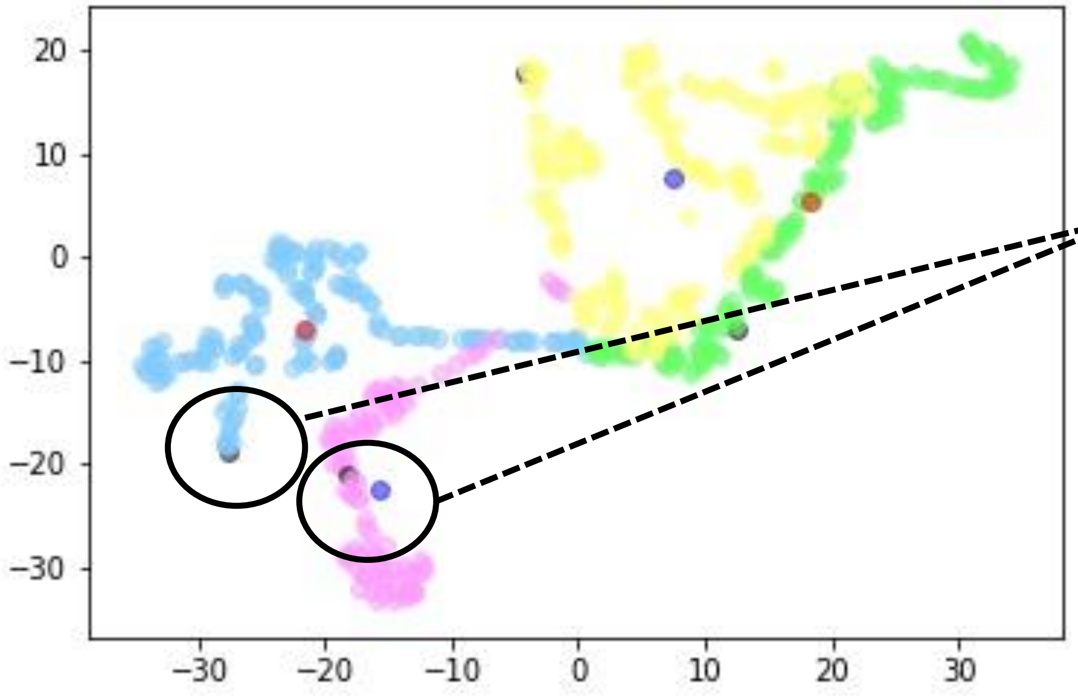
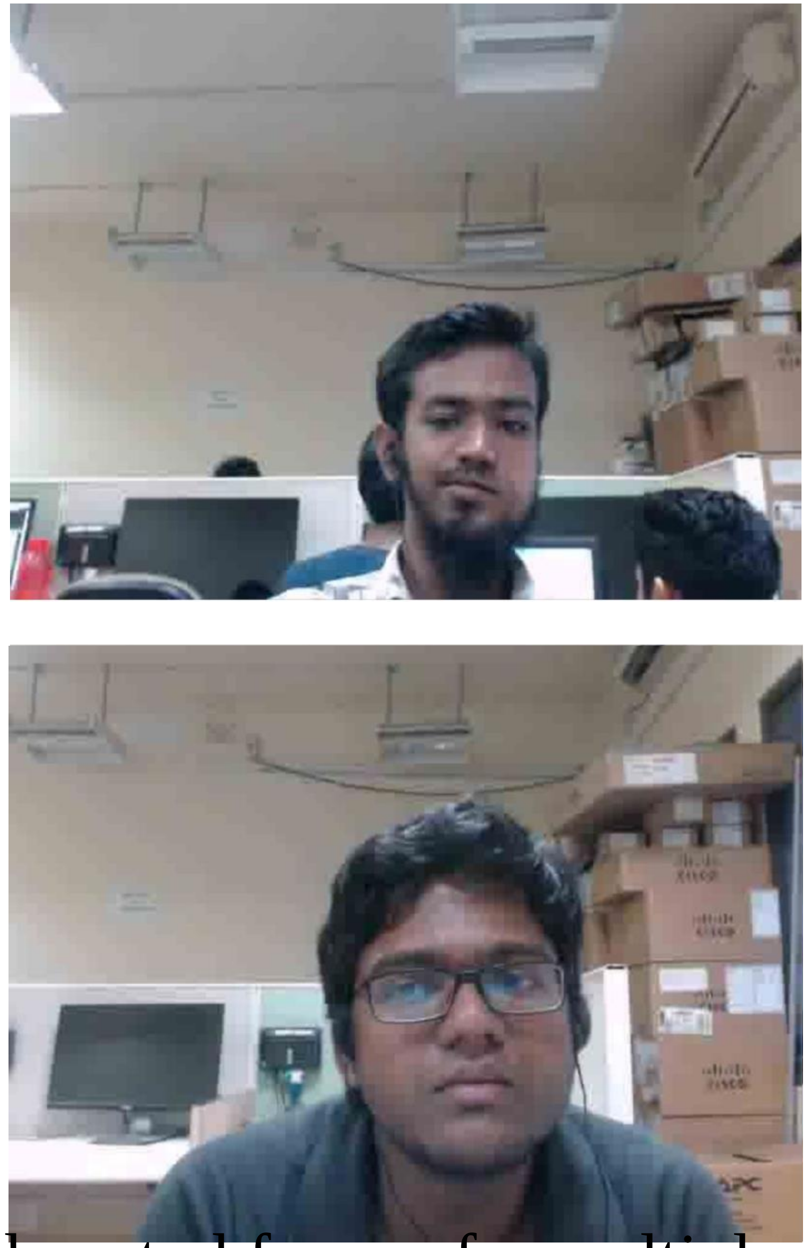


Figure 2: Kmeans clustering results for peak and neutral frames for multiple persons

Experimental Results



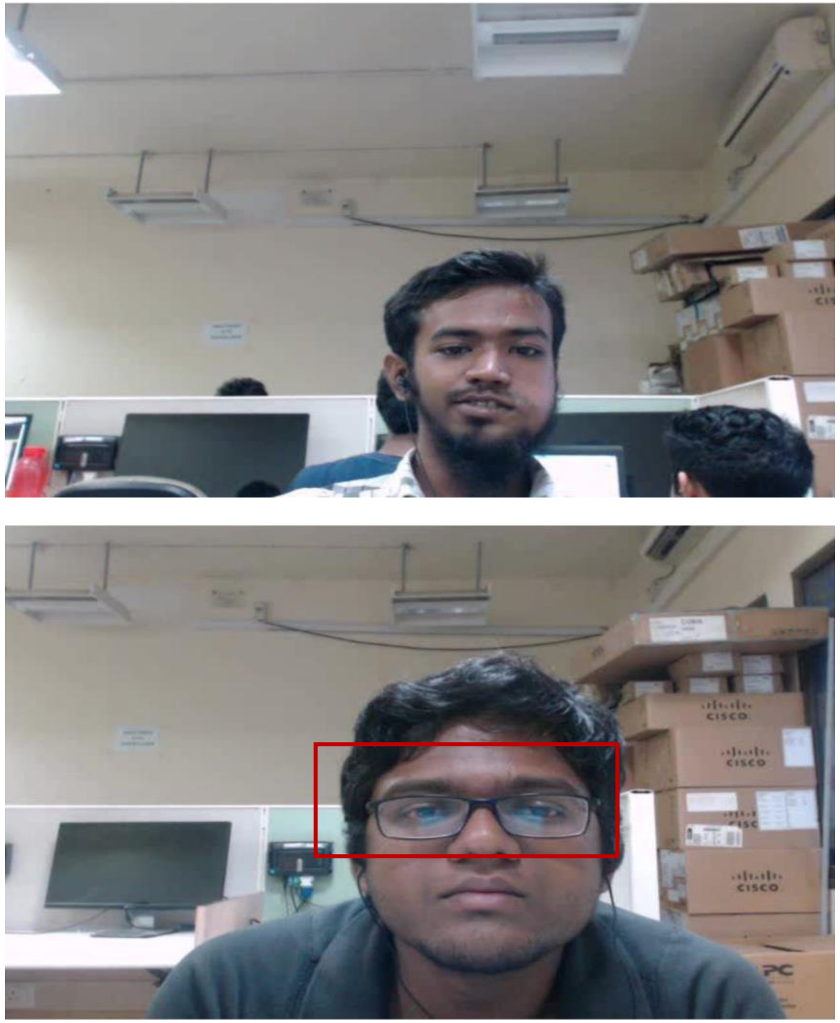


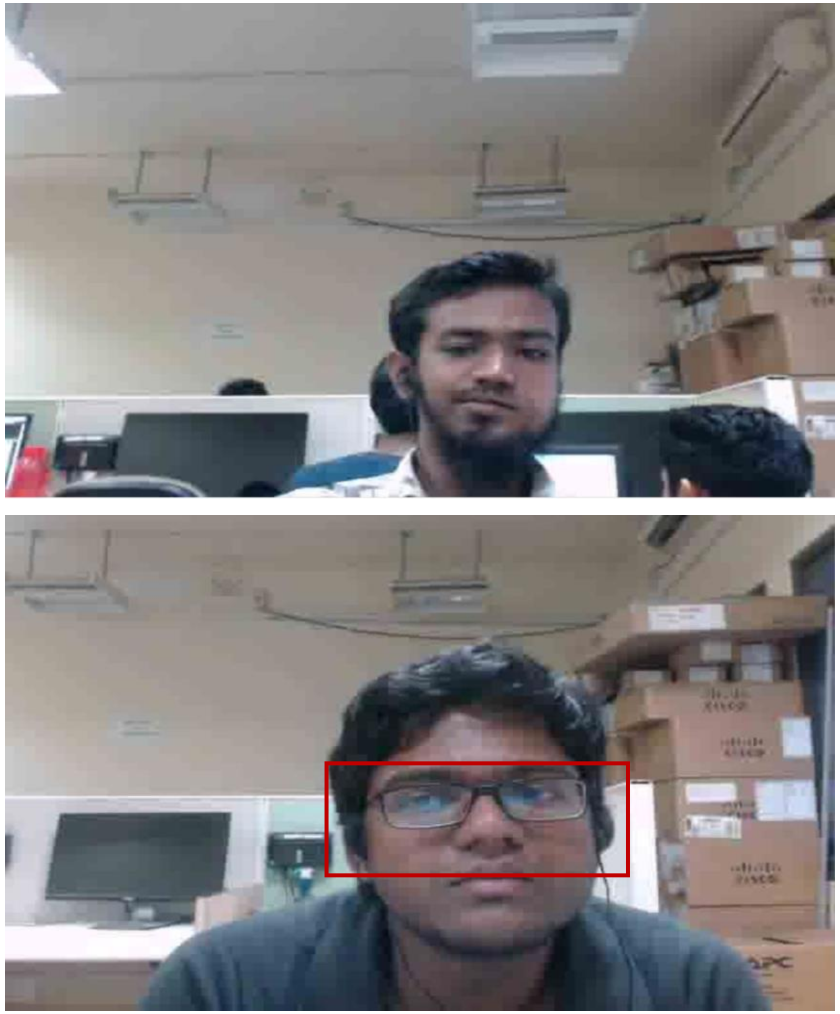
Samples from Cluster 1

Figure 2: Kmeans clustering results for peak and neutral frames for multiple persons

Experimental Results

Samples from Cluster 0





Samples from Cluster 1

Figure 3: Comparison results of samples

Conclusions

- In this study, we proposed the engagement levels estimation based on the facial features by using transfer learning technique.
- We also considered the individual differences in expressing the engagement levels.
- In the future, we will make an improvement in accuracy according to our proposed method.

References

1. Reed W Larson and Maryse H Richards. “Boredom in the middle school years:Blaming schools versus blaming students”. In:American journal of education99.4 (1991), pp. 418–443.
2. Veena Mayya, Radhika M Pai, and MM Manohara Pai. “Automatic facial expres-sion recognition using DCNN”. In:Procedia Computer Science93 (2016), pp. 453–461.
3. Deepak Kumar Jain, Pourya Shamsolmoali, and Paramjit Sehdev. “Extended deepneural network for facial emotion recognition”. In:Pattern Recognition Letters120(2019), pp. 69–74.
4. Motaz Sabri and Takio Kurita. “Facial expression intensity estimation using Siameseand triplet networks”. In:Neurocomputing313 (2018), pp. 143–154.
5. Xuanyu He and Wei Zhang. “Emotion recognition by assisted learning with con-volutional neural networks”. In:Neurocomputing291 (2018), pp. 187–194.
6. Jingying Chen, Ruyi Xu, and Leyuan Liu. “Deep peak-neutral difference featurefor facial expression recognition”. In:Multimedia Tools and Applications77.22(2018), pp. 29871–29887.