# Face-GPS: A Comprehensive Technique for Quantifying Facial Muscle Dynamics in Videos


**Juni Kim**
Stanford Online High School
Redwood City, CA 94063

**Zhikang Dong**
Department of Applied Mathematics and Statistics
Stony Brook, NY 11794

**Paweł Polak**
Department of Applied Mathematics and Statistics
Institute for Advanced Computational Science
Stony Brook, NY, 11794

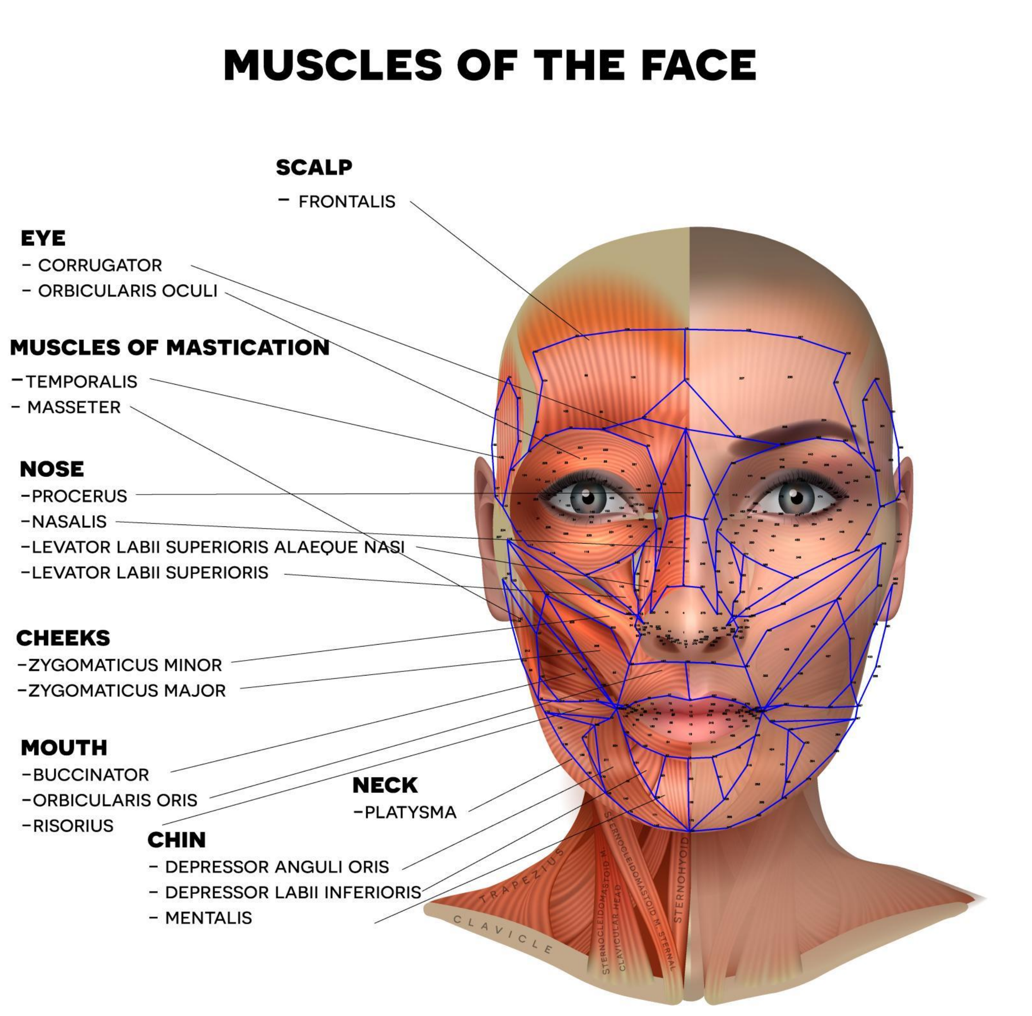


## Abstract

We introduce a novel method that combines differential geometry, kernels smoothing, and spectral analysis to quantify facial muscle activity from widely accessible video recordings, such as those captured on personal smartphones. Our approach emphasizes practicality and accessibility. It has significant potential for applications in national security and plastic surgery. Additionally, it offers remote diagnosis and monitoring for medical conditions such as stroke, Bell's palsy, and acoustic neuroma. Moreover, it is adept at detecting and classifying emotions, from the overt to the subtle. The proposed face muscle analysis technique is an explainable alternative to deep learning methods and a non-invasive substitute to facial electromyography (fEMG).


## Introduction

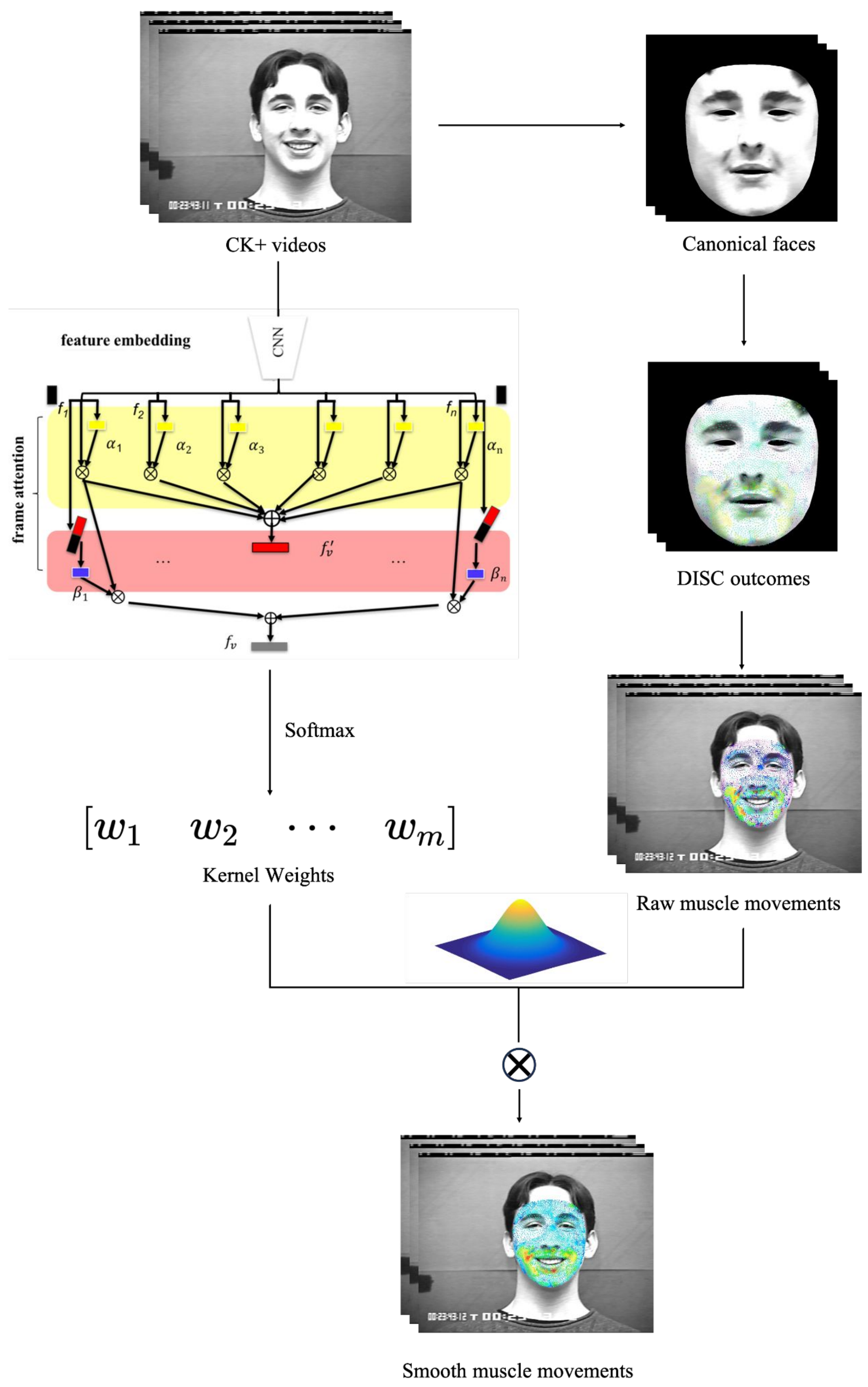
As a fundamental aspect of human communication, facial expressions convey emotions, feelings, and personal identities. Traditional facial muscle movement analysis mainly employs facial electromyography (fEMG) to detect muscle contractions and relaxations. However, fEMG's need for specialized equipment and expertise renders it inflexible and unsuitable for quick prediagnosis. As an alternative, the Facial Action Coding System (FACS) uses visualization-based methods to categorize facial actions into Action Units (AUs). Despite its capability to capture distinct facial expressions, FACS is time-consuming, subject to bias, and unsuitable for large sample studies. To mitigate these limitations, researchers have explored automated scoring systems employing techniques like probabilistic likelihood classifiers, and Dynamic Bayesian Networks for AU modeling. Although deep learning approaches have shown promising results, their lack of explainability poses challenges for those without domain expertise. Although research has sought to interpret facial recognition outcomes, these studies have neither considered the importance of facial muscle movements for clinical applications nor effectively eliminated the confounding effects of background noise and head movements.





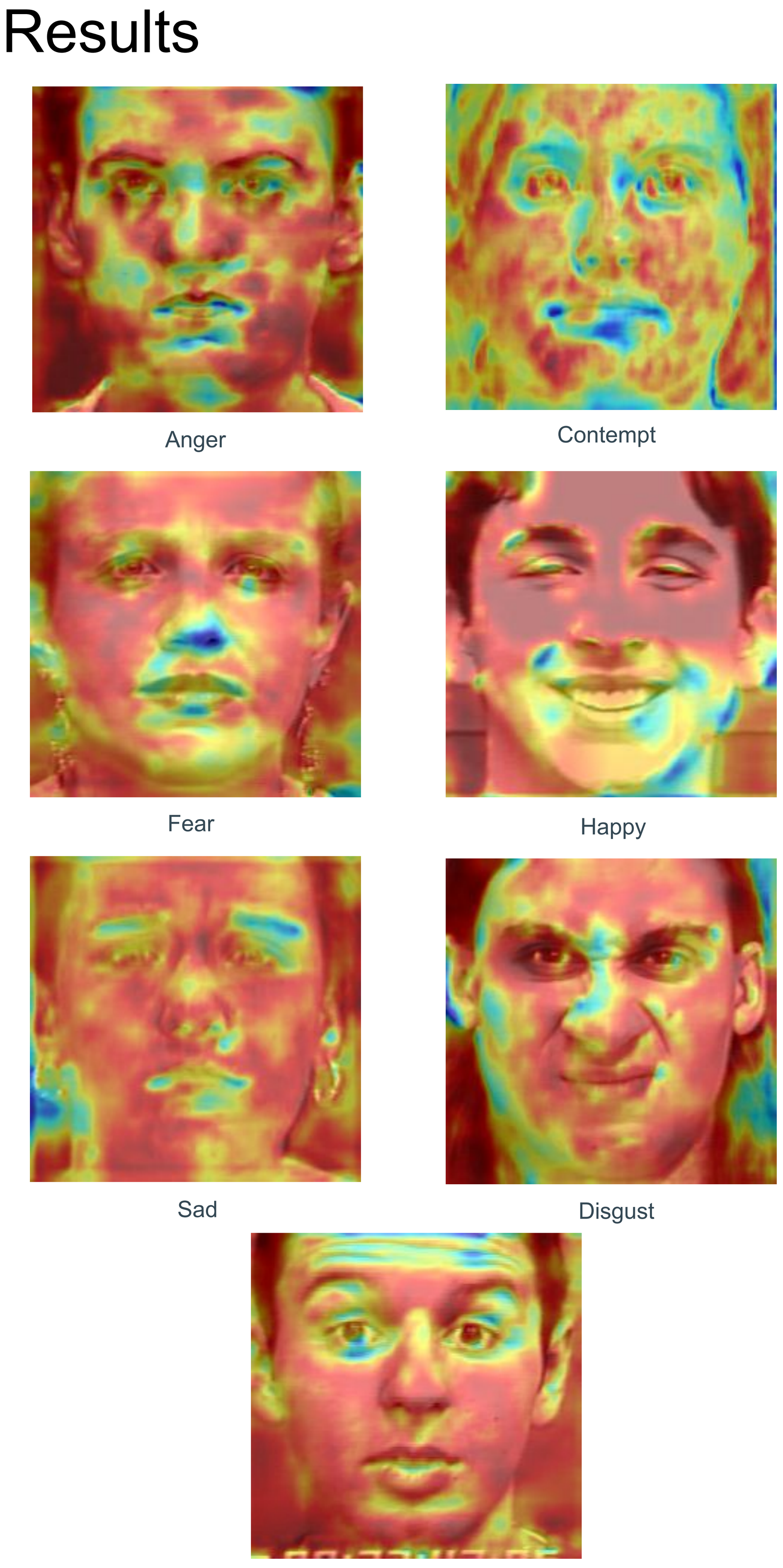
## Methodology

Initially, we extract face manifolds from video frames and convert them into a canonical face representation to minimize the effects of background and head movements. Next, we apply the Lucas-Kanade algorithm to create a vector field that estimates muscle movements between frames in the canonical face representation. To make these results more interpretable, this vector field is further smoothed using multiple kernels and is then overlaid on the original video. We evaluate this pipeline using 327 videos from the CK+ Dataset, each featuring individuals exhibiting one of seven basic emotions.

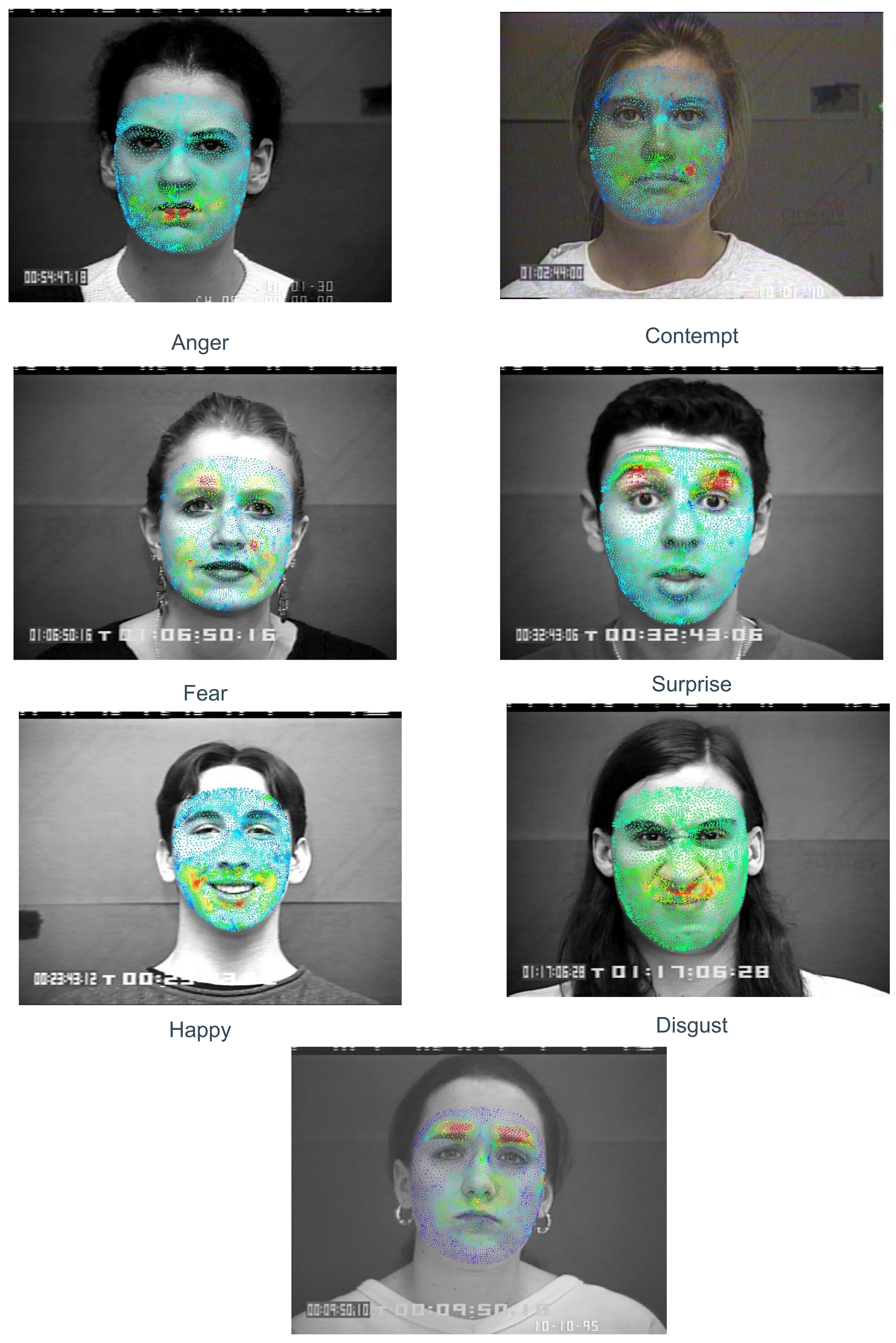



The overall pipeline for our method

## Results




Visualizations of FAN (without Face-GPS) using the Grad-CAM package.




Facial muscle tracking with our kernel smoothing technique, from the CK+ Dataset

| Model | Average Accuracy |
|---|---|
| Face-GPS without FAN | 85.0% |
| Face-GPS with FAN | 86.1% |

Accuracies of an XGBoost Classifier trained on displacement vectors from DISC in predicting emotions from the CK+ Dataset. They show the effectiveness of the kernel smoothing technique to provide informative features for predictive models and analysis.

## Conclusion

We present an end-to-end approach for dynamically quantifying facial muscle movements. Our method assesses these movements by tracking pixel displacements on a corresponding canonical face, allowing for accurate measurement even when the face is in motion or turned sideways. We develop a multi-kernel smoothing method to enhance the interpretability of face recognition deep learning models, highlighting the movements of specific muscle groups while filtering out noise from video recordings. Despite these advancements, capturing the facial manifold accurately, especially at its boundaries, remains a challenge and an area for future refinement. We also plan to improve this methodology to apply our kernels more precisely to the contours of facial muscles.

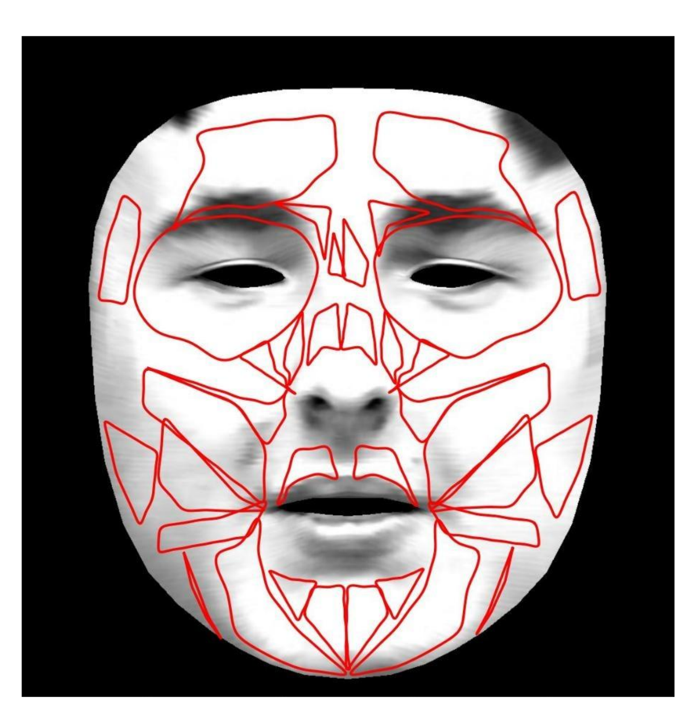

## References

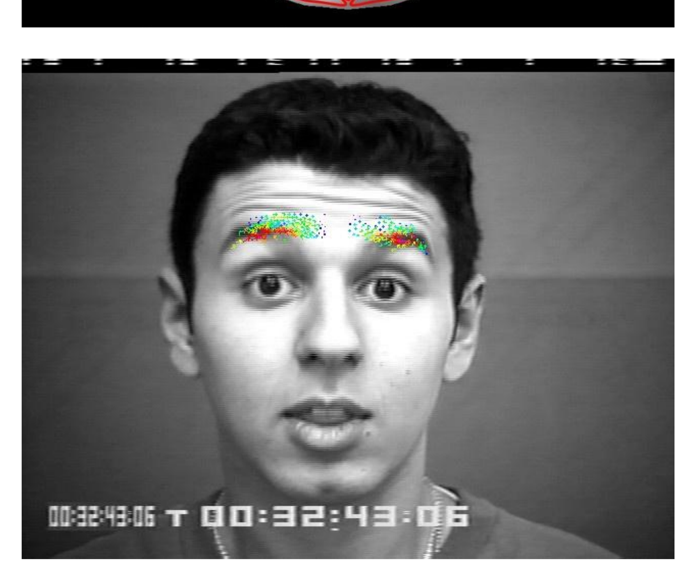
P. Ekman and W. V. Friesen, "Facial action coding system," In: Environmental Psychology & Nonverbal Behavior, 1978.

T. Yadav, M. M. U. Atique, H. F. Azgomi, J. T. Francis, and R. T. Faghih, "Emotional valence tracking and classification via state-space analysis of facial electromyography," in 2019 53rd Asilomar Conference on Signals, Systems, and Computers. IEEE, 2019, pp. 2116–2120

D. Meng et al. Frame attention networks for facial expression recognition in videos. In 2019 IEEE international conference on image processing (ICIP), pages 3866–3870. IEEE, 2019

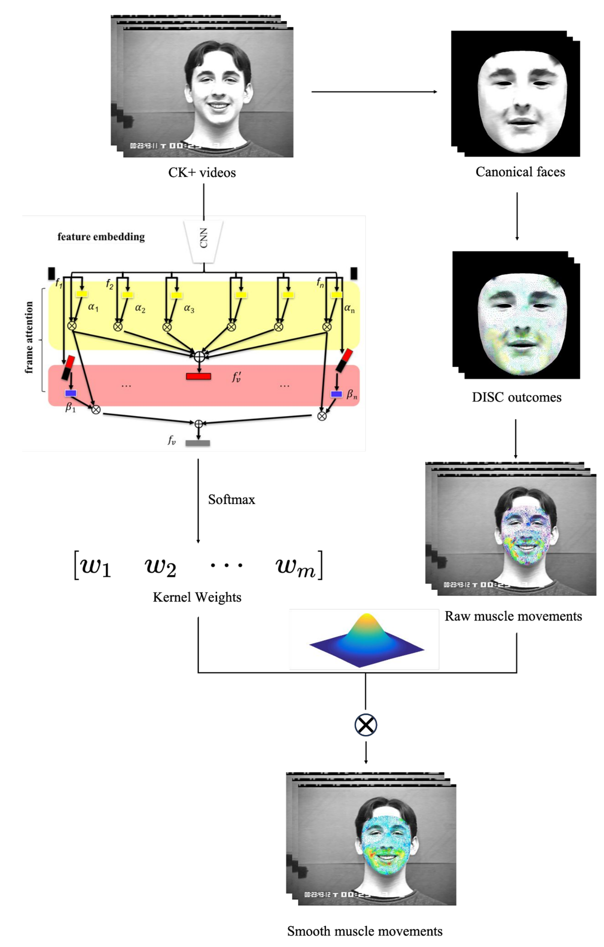
Patrick Lucey, Jeffrey Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. pages 94 – 101, 07 2010.

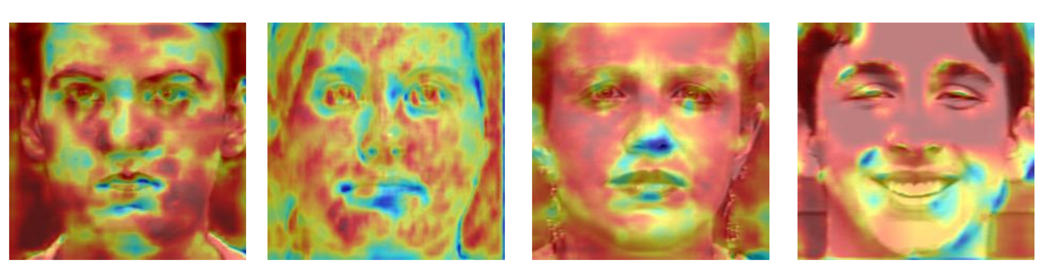


| Model | Average Accuracy |
|---|---|
| Face-GPS without FAN | 85.0% |
| FAce-GPS with FAN | 86.1% |

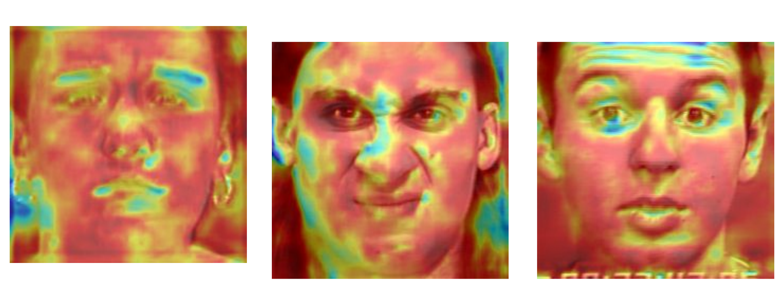
Classification Results of Facial Muscle Features, showing the effectiveness of the kernel smoothing technique to provide informative features for predictive models and analysis.




The overall pipeline for our method



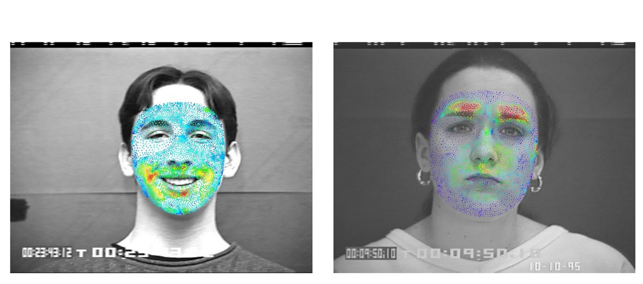
Anger Contempt Fear Happy



Sad Disgust Surprise

Visualizations of FAN using the Grad-CAM package

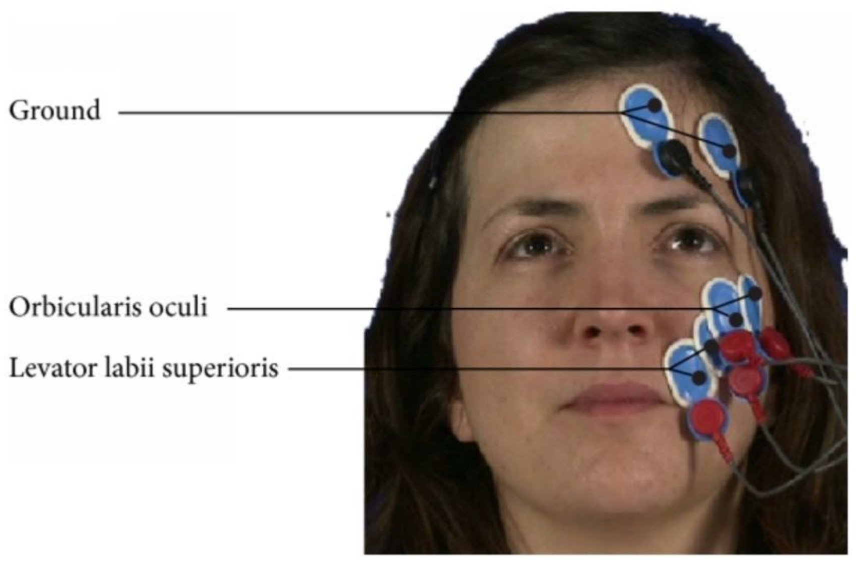
Anger Contempt Fear



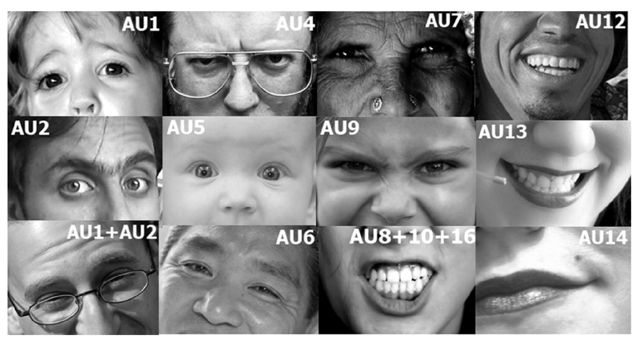
Happy Sad

Disgust Surprise

Facial muscle tracking with our kernel smoothing technique




https://www.researchgate.net/figure/Locations-of-the-EMG-electrodes_fig4_269184843




https://github.com/TadasBaltrusaitis/OpenFace/wiki/Action-Units

# Face-GPS: A Comprehensive Technique for Quantifying Facial Muscle Dynamics in Videos

Juni Kim, Zhikang Dong,
Pawel Polak

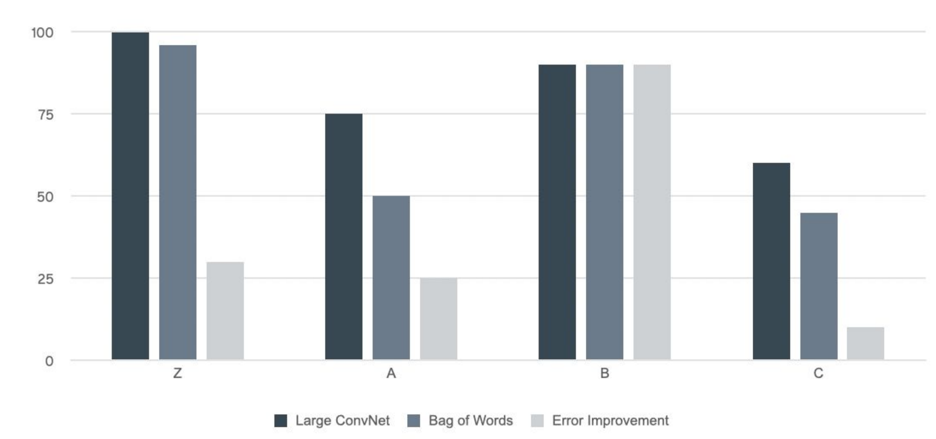
## Headline in 34pt should not extend beyond 2-3 lines

This section is an example of a paragraph. When creating sections, regardless of whether you're putting in text or images, always try to align to the edges of the yellow guidelines. This poster canvas is broken into 3 columns, and aligning to the edges will make it much easier for viewers to differentiate sections and read information. The same is true of horizontal spaces between sections, try to space them equally and with a good amount of breathing room in between each.

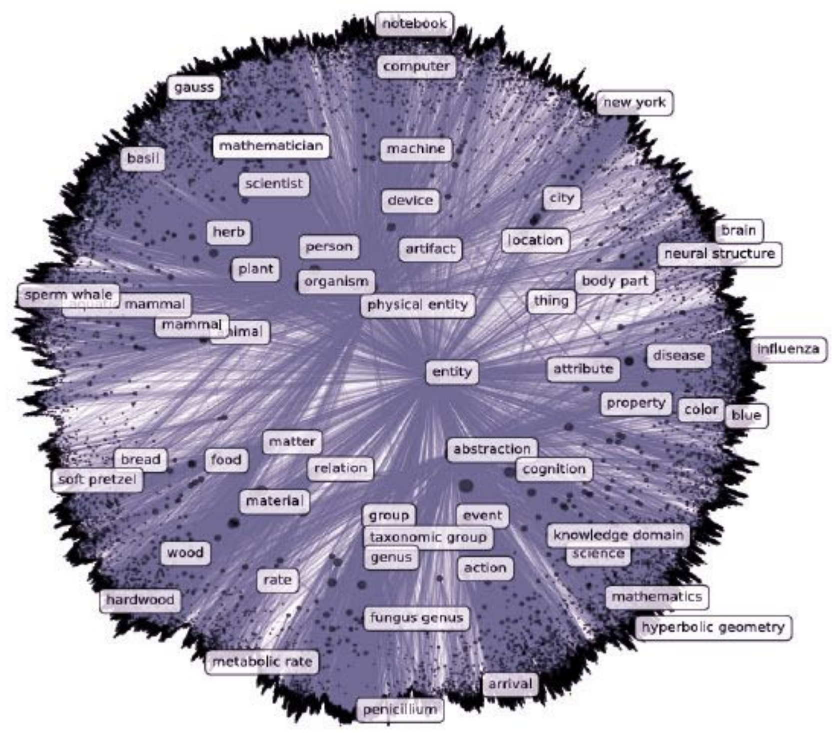
- Bulleted list item
- Bulleted list item
- Bulleted list item
- Bulleted list item
- Bulleted list item
- Bulleted list item

## Section header in 34pt font

Optional section descriptor in 21pt font




This section is an example of a paragraph. When creating sections, regardless of whether you're putting in text or images, always try to align to the edges of the yellow guidelines. This poster canvas is broken into 3 columns, and aligning to the edges will make it much easier for viewers to differentiate sections and read information.



Optional caption for images, charts, and graphs

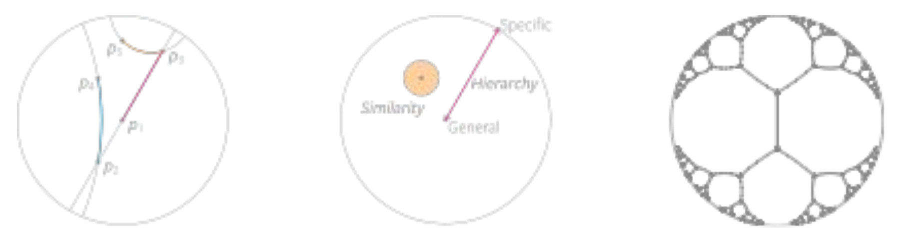
## Section header in 34pt font

Optional section descriptor in 21pt font

This section is an example of a paragraph. When creating sections, regardless of whether you're putting in text or images, always try to align to the edges of the yellow guidelines. This poster canvas is broken into 3 columns, and aligning to the edges will make it much easier for viewers to differentiate sections and read information.



Optional caption for images, charts, and graphs

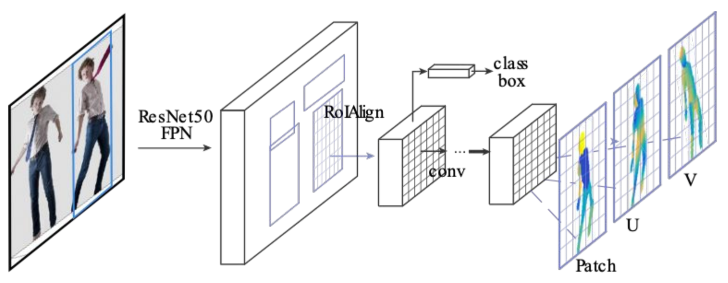


| Dimensionality | | 5 | 10 | 20 | 50 |
|---|---|---|---|---|---|
| Euclidean | Rank | 3542.3 | 2286.9 | 1685.9 | 1281.7 |
| | Map | 0.024 | 0.059 | 0.087 | 0.140 |
| Translational | Rank | 205.9 | 179.4 | 95.3 | 92.8 |
| | Map | 0.517 | 0.503 | 0.563 | 0.566 |
| Poincaré | Rank | 4.9 | 4.02 | 3.84 | 3.98 |
| | Map | 0.823 | 0.851 | 0.855 | 0.86 |

## Section header in 34pt font

Optional section descriptor in 21pt font

This section is an example of a paragraph. When creating sections, regardless of whether you're putting in text or images, always try to align to the edges of the yellow guidelines.




### References

References in 14pt font

Homer W Simpson (2013). "Donuts taste good." In: IEEE 13th Internation Conference on Data Mining. IEEE, pp. 405-409

Marge Simpson (2010). "Blue hair looks nice.". In: Nature communications 1, p. 622.

Bart Simpson (2013). "Hello". In: IEEE Simpsons.

Marge Simpson et al. (2013). "Lorem Ipsum." In: Advances in Neural Information Processing Systems 26. Ed. by Christopher J. C. Burges et al., pp. 27–29.