0429 Update

Paper Selection

- Audio-Gaze Driven Avatar (&Codec Avatar)
  - Richard, A., Lea, C., Ma, S., Gall, J., De la Torre, F., & Sheikh, Y. (2021). Audio-and gaze-driven facial animation of codec avatars. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 41-50).
- Visemenet = JaLi
  - Zhou, Y., Xu, Z., Landreth, C., Kalogerakis, E., Maji, S., & Singh, K. (2018). Visemenet: Audio-driven animator-centric speech animation. *ACM Transactions on Graphics (TOG)*, *37*(4), 1-10.
- D3DExpression
  - Potamias, R. A., Zheng, J., Ploumpis, S., Bouritsas, G., Ververas, E., & Zafeiriou, S. (2020, August). Learning to generate customized dynamic 3d facial expressions. In European Conference on Computer Vision (pp. 278-294). Springer, Cham.
- Meshtalk
  - Zhou, Y., Xu, Z., Landreth, C., Kalogerakis, E., Maji, S., & Singh, K. (2018). Visemenet: Audio-driven animator-centric speech animation. *ACM Transactions on Graphics (TOG)*, *37*(4), 1-10.

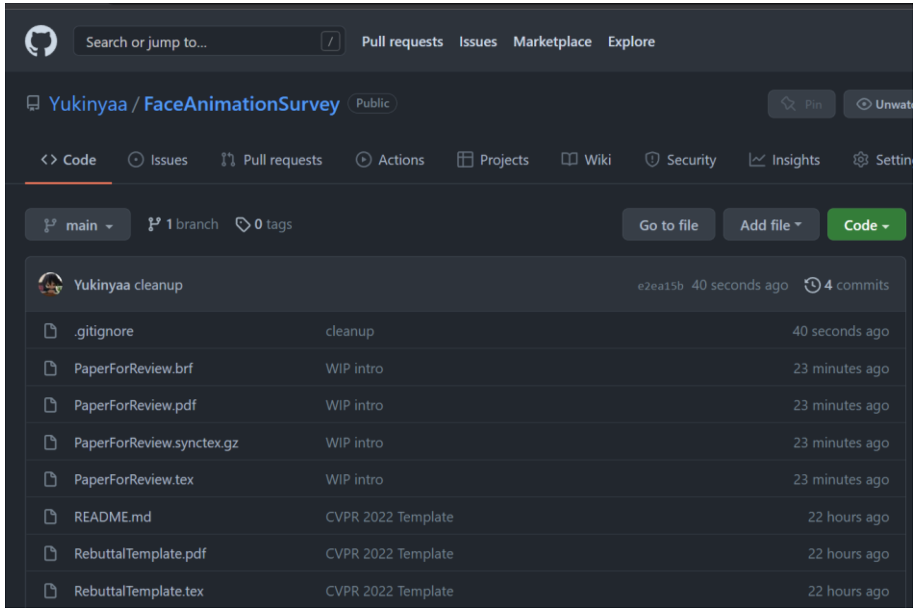
Writing

https://cvpr2022.thecvf.com/author-guidelines

This is the LaTeX template for IEEE/CVF CVPR 2022 submissions, rebuttals, and final versions.

1. Introduction

Modern movie and game renders very realistic 3D face that delivers human emotion and lip motion accurately. But in multiplayer Virtual Reality(VR) environment, capturing human face and delivering such information real-time is a hard task. This comes



- Brief explanation of each method
- Side-by side comparison
- Discussion
    - Limitation and Future direction

TABLE 4. Analysis of MLL, LDL and label enhancement models.

Author	Method	Data	Performance Evaluation	Contribution	Limitation
Zhao et al. [74]	GLMM	ML-JAFFE	Average Precision: 0.9143, Coverage error:2.9381, hamming loss:0.2035, One error:0.1071, ranking loss:0.1466	Model the relationship among FER labels	Model not capture the intensity estimation of the available emotions in FER data.
Ying et al. [71]	EDL + JD	(sJAFFE, sBU-3DFE):	(Kullback leibler:0.0346,0.0402), (Euclidean:0.0957,0.1005), (intersection:0.8998,0.8939), (fidelity:0.9914,0.9898)	present information about emotion intensity in an expression instance	limited to label distribution data.
Xing et al. [76]	LDLogitBoost	s-BU-3DFE	Kullback Leibler:0.0491, Euclidean:0.1263, Fidelity:0.9886, intersection:0.8800	more general entropy model for modeling information distribution in facial images	Not generalised to in-the wild and logical label data, the performance may degrade with large volume data
Xing et al. [76]	AOSO-LDLogitBoost	s-BU3DFE	Kullback Leibler:0.0515, Euclidean:0.1297, Fidelity:0.9874, inersection:0.8764	more general entropy model for modeling information distribution in facial images	Not generalised to in-the-wild and logical label data, the performance may degrade with large volume data
Li and Deng [34]	DBM-CNN	RAF-ML	CLM-Hamming: 0.217, RAkEL-Hamming:0.177, ML-KNN-Hamming:0.168, ML-LOC:0.173, LIFT-Hamming:0.167	Preserves both local affinity and manifold structure of emotion label. Introduction of Adaptation mechanism for data generalisation	computational complexity and resource consumption
Xu et al. [81]	Label enhancement with GLLE (manifold learning)	Bu-3DFE	cheb:1.00, clark:1.13, canb:1.13, cosine:1.00, Interception: 1.07	Label enhancement with consideration given to correlation among labels. Could be applied to data with no distribution label	Not advisable to use on large data size or in-the-wild data. It is Computationally expensive due to implementation of KNN search.
Jia et al. [82]	EDL-LRL + ADMM optimiser	(s-JAFFE, S-BU-3DFE	(cheb: 0.0806,0.0951), (clark:0.3008,0.3556), (cand: 0.6134, 0.7463), (Kullback Leibler:0.0361,0.0694), (cosine: 0.9660,0.9626), (intersection: 0.8970:0.8686)	Preserve correlation among data label locally.	Not generalised to in-the-wild data and data with logical label
Abeere et al. [84]	EDL-LBCNN (CNN + LBC features) KL loss	s-JAFFE	Kullback Leibler:0.0168, CS:0.9842	system performance increases via hybrid convolutional features.	Not generalised to in-the-wild data and data with logical label
Zhang et al. [83]	Cosine similarity + Deep CNN	Oulu-CASIA NIR FER	(Accuracy 81.97% in weak light), (Accuracy: 82.67% in the Dark), (Accuracy in strong light: 84.40%)	the model is immune to illumination variations	not applicable to data with logical label and not generalises to in-the-wild data.
Chen et al. [75]	Auxiliary label (manifold learning) + Deep CNN	posed data(CK+, Oulu-CASIA, CFEE, MMI), wild data (AFFNET, RAF, SFEW)	(Avg. Accuracy: 76.25), (Avg. Accuracy:66.64)	resolve label inconsistency using label enhancement with correlation among labels. Applicable to data without distribution label, not affected by data volume, minimises searching with approximate kNN. Generalises to in-the-wild data and logical data	Time complexity and resource consumption due to auxiliary label space construction.

- Ekundayo, O., & Viriri, S. (2021). Facial Expression Recognition: A Review of Trends and Techniques. *IEEE Access*.