INFORMATIK

augmented VISION DFKI

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

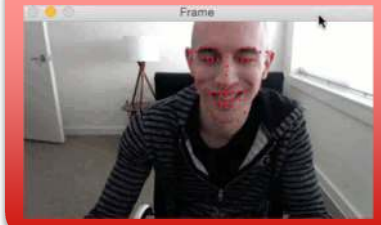
2D Image Processing & Augmented Reality
Winter Semester 2019/2020
Survey on Face Tracking with Deep Learning

Vinay Balasubramanian

v_balasubr18@cs.uni-kl.de

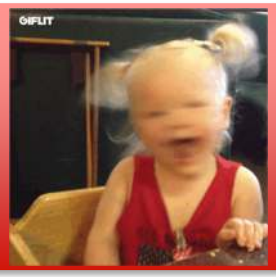
Supervisor: Jilliam Diaz Barros

Outline



Face tracking is a computer vision task that involves tracking a specific number of landmarks on the face detected across all frames of a video.

Applications include Face analysis, Person identification, Activity recognition, Expression analysis, Face modeling etc.



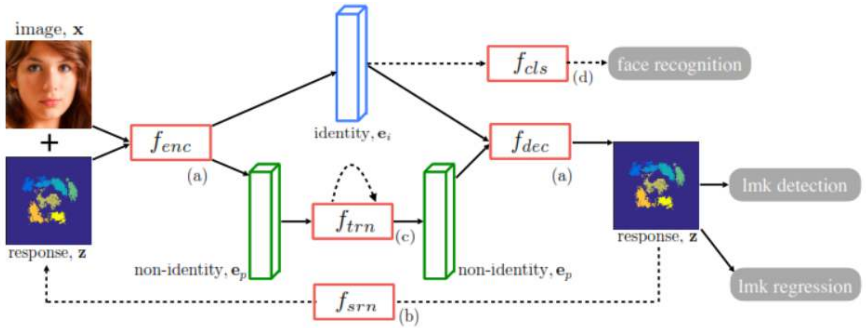
It is a challenging problem as the videos can be captured in unconstrained conditions which may include illumination variations, large head poses, occlusions, etc.

Methods

- **Image-based methods use models trained on still images in each frame.**
- **Video-based methods make use of temporal information to predict facial landmarks in each frame.**
- **Various approaches –**
  - **Regression-based methods**
  - **Video-based face alignment**
  - **Encoder-Decoder Networks**

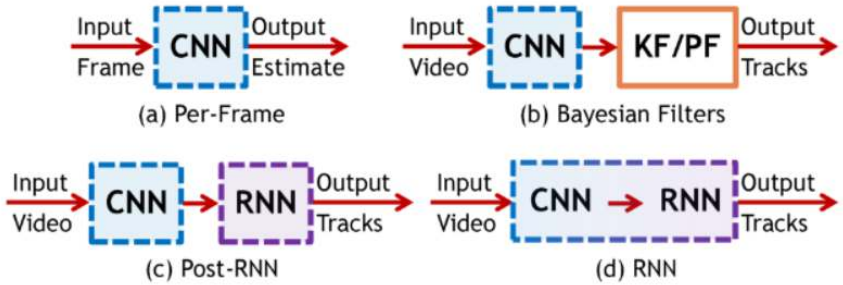
1. Recurrent Encoder-Decoder Network for Video-based Face Alignment

- **Employs recurrent learning at both spatial and temporal dimensions.**
- **At spatial level, face alignment is done in a coarse-fine manner.**
- **At temporal level, Temporal-variant features such as pose and expression are separated from Temporal-invariant features such as facial identity.**



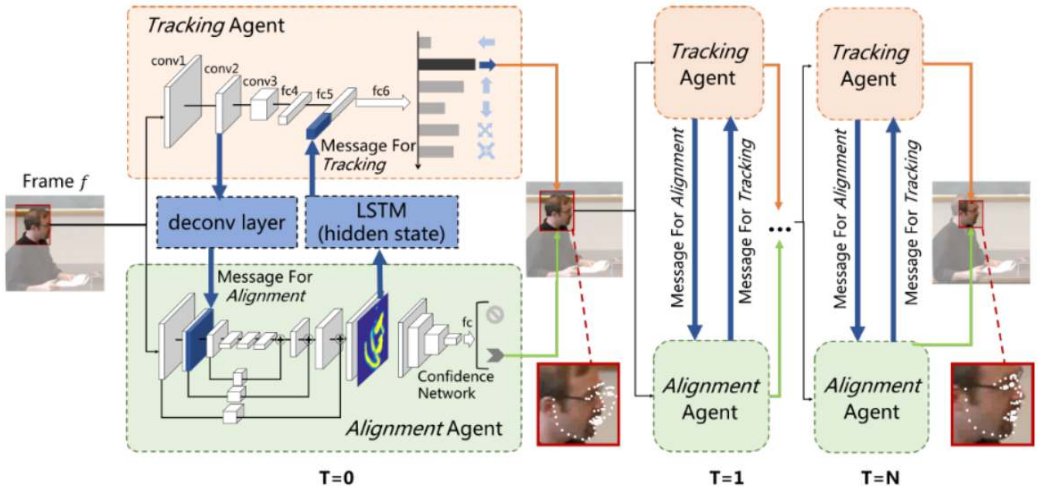
2. Dynamic Facial Analysis: From Bayesian Filtering to Recurrent Neural Network

- Previous approaches for dynamic facial analysis use Kalman/Particle filters.
- Bayesian filters require problem-specific design and tuning.
- This RNN based method avoids tracker engineering by learning from data (sufficient data).
- CNN layers followed by recurrent layers as dense layers.



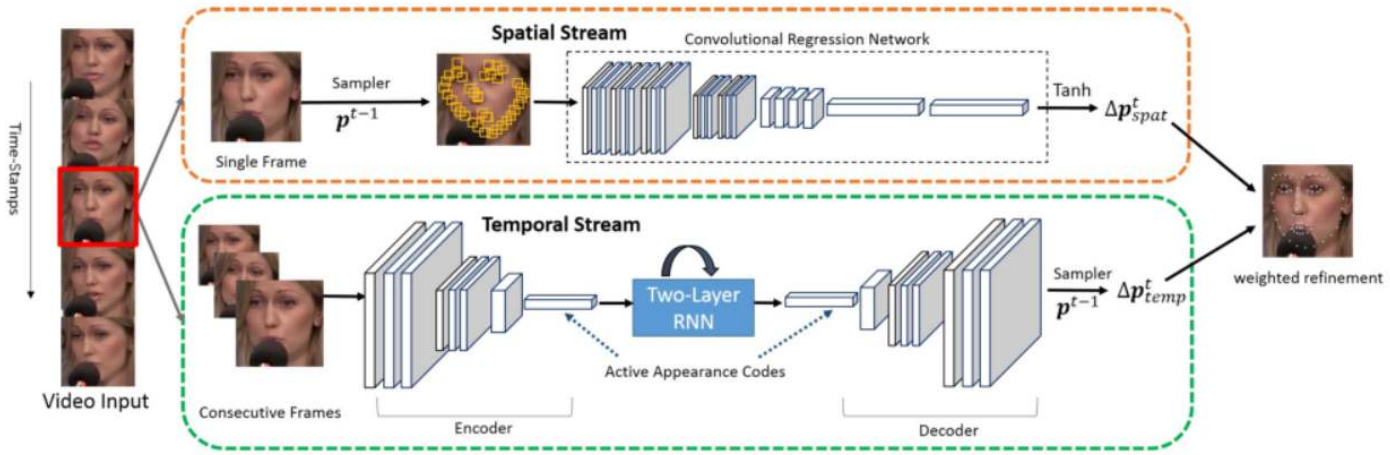
3. Dual-Agent Deep Reinforcement Learning for Deformable Face Tracking

- **Exploits the fact that bounding box tracking and landmark detection tasks are dependent. The accuracy of the latter depends on how good the former is.**
- **The two tasks are modeled in a probabilistic manner by following a Bayesian model.**



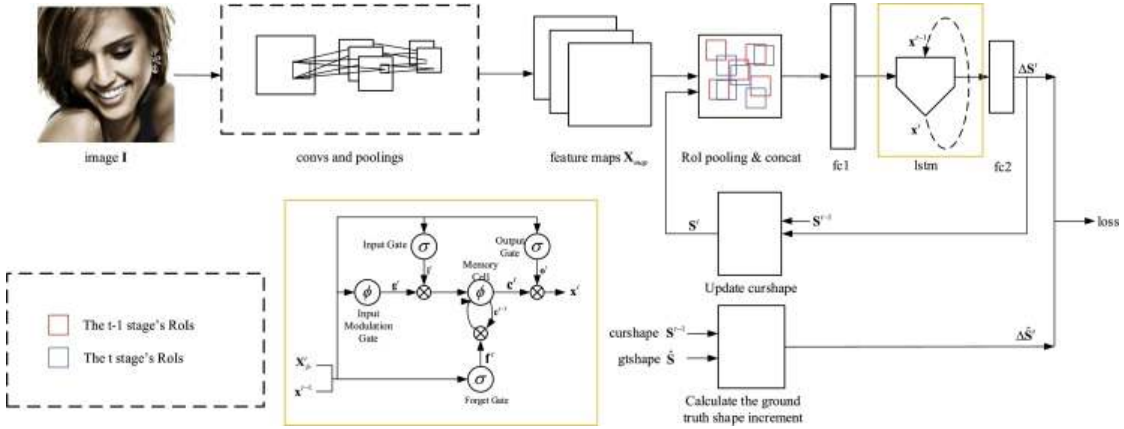
4. Two-stream Transformer Networks for Video-based Face Alignment

- **Two stream deep learning method to capture spatial as well as temporal information.**
- **Facial landmarks are determined by a weighted fusion of spatial and temporal streams.**



5. Face Alignment Recurrent Neural Network

- Recurrent regression based approach
- Uses LSTM to exploit both spatial and temporal information.
- Spatial – The predicted landmark location is used as basis for estimation in the next stage.
- Temporal – The predicted landmark location is used as basis for estimation in the next frame.



Comparison metrics

- **Following are some of the metrics that will be used to compare the various methods in this survey :–**
  - Dataset used for training (In the wild vs constrained)
  - Evaluation metrics
  - Number of landmarks tracked
  - Kind of landmarks retrieved (2D or 3D)
  - Robustness to large pose variations, illumination changes

References

1. Xi Peng, Rogerio S. Feris, Xiaoyu Wang, Dimitris N. Metaxas: RED-Net: A Recurrent Encoder-Decoder Network for Video-based Face Alignment.
2. Jinwei Gu, Xiaodong Yang, Shalini De Mello, Jan Kautz: Dynamic Facial Analysis: From Bayesian Filtering to Recurrent Neural Network.
3. Minghao Guo, Jiwen Lu, and Jie Zhou: Dual-Agent Deep Reinforcement Learning for Deformable Face Tracking.
4. Hao Liu, Jiwen Lu, Jianjiang Feng, Jie Zhou: Two-Stream Transformer Networks for Video-based Face Alignment
5. Qiqi Hou, Jinjun Wang, Ruibin Bai, Sanping Zhou, Yihong Gong: Face Alignment Recurrent Network

augmented VISION DFKI

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Thank You