[3]DEEP FACIAL EXPRESSION RECOGNITION

- Identifies three main steps required in a deep FER system and describes the related background.
- [3.1]Pre-processing
  - Different background, illuminations, head-poses
  - [3.1.1]Face alignment
    - Given a series of training data, the first step is to detect the face and then to remove background and non-face areas.
      - Viola-Jones face detector
    - Face alignment using the coordinates of localized landmarks can substantially enhance the FER performance.(reduce the variation in face scale and in-plane rotation)
      - Holistic models
        - AAM(Active Appearance Model)
          - Classic generative model that optimizes the required parameters from holistic facial appearance and global shape patterns.
      - Discriminative models
        - Part-based approaches that represent the face via the local appearance information around each landmark.
        - Mixtures of trees structured models(MoT)
        - Discriminative response map fitting(DRMF)
        - Use a cascade of regression functions to map the image appearance to landmark locations and have shown better

results

- ◆ SDM(supervised descent method)

- Deep learning
  - Cascaded CNN
    - ◆ Early work which predicts landmarks in a cascaded way.
  - Multi-task CNN(MTCNN)
    - ◆ Further leverage multi-task learning to improve the performance.
  - Cascaded regression has become the most popular and state-of-the-art methods for face alignment as its high speed and accuracy.

- [3.1.2]Data augmentation
  - ◆ On-the-fly data augmentation
    - Embedded in deep learning toolkits to alleviate overfitting.
    - Training step
      - Input samples are randomly cropped(절단) from the four corners and center of the image and then flipped horizontally
        - ◆ Result in a dataset that is ten times larger than the original training data.
    - Testing step
      - Only the center patch of the face is used for prediction
      - Prediction value is averaged over all ten crops
  - ◆ Offline data augmentation
    - Further expand data on both size and diversity

- Rotation, shifting, skew(왜곡), scaling, noise, contrast, color jittering(이미지 채도 랜덤 noise)
- Enlarge the data size : common noise models, salt & pepper and speckle noise, Gaussian noise
- Combinations of multiple operation can generate more unseen training samples and make the network more robust to deviated and rotated faces.
- Five image appearance filters(disk, average, Gaussian, unsharp, motion filters)
- Six affine transform matrices(회전, 평행이동, scale, 반전(reflection), skew, shearing)
- Deep learning based technology
  - Synthetic data generation system with 3D convolutional neural network(CNN) was created in to confidentially create faces with different levels of saturation in expression.
  - Generative adversarial network(GAN) can also be applied to augment data by generating diverse appearances varying in poses and expressions.

■ [3.1.3]Face Normalization

◆ Variations in illumination and head poses can introduce large changes in images and hence impair the FER performance.

- Illumination normalization
  - IS(isotropic diffusion), DCT(discrete cosine transform), DoG(difference of Gaussian), homomorphic filtering("most consistent results" remove illumination)
  - Combine with histogram equalization results in better face

recognition performance)

- Histogram equalization may overemphasize local contrast.
  - To solve, weighted summation approach to combine histogram equalization and linear mapping.
- GCN(global contrast normalization), local normalization, histogram equalization
  - GCN and histogram equalization were reported to achieve the best accuracy for the training and testing steps.

- Pose normalization
  - Hassner, Sagonas … and very recently GAN-based deep models were proposed for frontal view synthesis(FF-GAN, TP-GAN, DR-GAN)