[Abstract]

- FER(FACIAL EXPRESSION RECOGNITION)
  - Overfitting
    - ◆ 충분한 training data의 부족
  - Expression-unrelated variations
    - ◆ Illumination, head pose and identity bias
  - Introduce the available datasets
  - Describe the standard pipeline of a deep FER system
  - Review existing novel deep neural networks and related training strategies
  - Competitive performances on widely used benchmarks
  - Review the remaining challenges and corresponding opportunities in this field

[Introduction]

- Prototypical facial expressions
  - Anger, disgust, fear, happiness, sadness, surprise, contempt
    - Culture-specific(not universal)
  - Categorical model
    - VS) affect model : wider range of emotions
    - Most popular(pioneering investigations along with the direct and intuitive definition of facial expression)
- Two main categories(feature representations)
  - Static image FER
    - Feature representation : only spatial information current single image
  - Dynamic image FER
    - Consider the temporal relation among contiguous frames in the input facial expression sequence.
  - Other modalities(multimodal system)
    - Audio and physiological channels
    - To assist the recognition of expression.
- Traditional methods
  - Handcrafted features
  - Shallow learning
    - LBP, LBP-TOP, NMF, sparse learning
- 2013
  - Sufficient training data

- Dramatically increased chip processing abilities(GPU)

- Very recently
  - FER based on deep learning has been surveyed. (static, video images)
- Problems
  - Require a large amount of training data to avoid overfitting
  - Different personal attributes
    - Age, gender, ethnic backgrounds and level of expressiveness
  - Subject identity bias
    - Variations in pose, illumination, occlusions(가려지는 것)

[Facial Expression Databases]

- Frequently used expression databases
- CK+(The Extended CohnKanade)
  - Most extensively used laboratory-controlled database for evaluating FER systems.
- MMI
  - Laboratory-controlled and includes 326 sequences from 32 subjects.
- JAFFE(Japanese Female Facial Expression)
  - Laboratory-controlled image database that contains 213 samples of posed expressions from 10 Japanese females.
- TFD(Toronto Face Database)
  - Amalgamation(합병) of several facial expression datasets.
- FER2013
  - ICML 2013 Challenges in Representation Learning.
- AFEW(Acted Facial Expressions in the Wild)
  - First established and introduced in and has served as an evaluation platform for the annual Emotion Recognition In The Wild Challenge since 2013.
- SFEW(Static Facial Expressions in the Wild)
  - Created by selecting static frames from the AFEW database by computing key frames based on facial point clustering.
- Multi-PIE
  - Contains 755,370 images from 337 subjects under 15 viewpoints and 19 illumination conditions in up to four recording session.

- BU-3DFE(Binghamton University 3D Facial Expression)
  - Contains 606 facial expression sequences captured from 100 people.
- Oulu-CASIA
  - Includes 2,880 image sequences collected from 80 subjects labeled with six basic emotion labels.
- RaFD(Radboud Faces Database)
  - Laboratory-controlled and has a total of 1,608 images from 67 subjects with three different gaze directions.
- KDEF(Karolinska Directed Emotional Faces)
  - Originally developed for use in psychological and medical research.
- EmotioNet
  - Large-scale database with one million facial expression images collected from the Internet.
- RAF-DB(Real-world Affective Face Database)
  - Real-world database that contains 29,672 highly diverse facial images downloaded from the Internet.
- AffectNet
  - Contains more than one million images from the Internet that were obtained by querying different search engines using emotion-related tags.
- ExpW(Expression in-the-Wild Database)
  - Contains 91,793 faces downloaded using Google image search.

[DEEP FACIAL EXPRESSION RECOGNITION]

- Identifies three main steps required in a deep FER system and describes the related background.
- Pre-processing
  - Different background, illuminations, head-poses
  - 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.

- 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.