FaceDNA: Intelligent Face Recognition System with Intel RealSense 3D Camera

Dan Ye[1], Shih-Wei Liao[1]
[1]National Taiwan University, Department of Computer Science and Information Engineering

Abstract—This paper develops an intelligent face recognition system which has been applied to Intel Realsense 3D camera. The key components include computer vision application, face tracking, personal attributes, FaceDNA, emotion detection, video application modules. Computer vision application can extract rich information from images to categorize and process visual data. FaceDNA is consist of face verification, face identification, face detection, face matching, face grouping. It can detect human faces and compare similar ones, organize people into groups according to visual similarity, and identify previously tagged people in images. Emotion application can analyze faces to detect a range of feelings and personalize your responses. Intelligent video processing stable video output, detect motion, creates intelligent thumbnails, and detects and tracks faces.

Keywords: FaceDNA, Intel RealSense 3D camera, face tracking and recognition.

1 Introduction

The Intel® RealSense™ SDK includes a face tracking module provides a suite of the following face algorithms: Face detection locates a face (or multiple faces) from an image or a video sequence, and returns the face location in a rectangle. You can use this feature to count how many faces are in the picture and find their general locations. Landmark detection further identifies the feature points (eyes, mouth, etc.) for a given face rectangle. The eye location is of a particular interest for applications that change display perspectives based on where on the screen users are looking. Pose detection estimates the face orientation where the user's face is looking. Expression detection calculates the scores for a few supported facial expressions such as eye-closed and eye-brow turning up. Face recognition feature compares the current face with a set of reference pictures in the recognition database to determine the user's identification. Pulse estimation tracks subtle change in face skin color over time and estimates the person's pulse rate. Gaze tracking traces human eye movement and provide estimated eye gaze location on the display screen and angles from the origin.

This paper constructs a novel face recognition system that integrates computer vision application, face tracking, personal attributes, FaceDNA, emotion detection, video application modules. FaceDNA can predict past, future face photo using current face photo which captured by Intel® RealSense™ 3D Camera. According to the changing facial feature various with growing up on same person, infer his or her photoes at different ages. FaceDNA can deduce blood relationship. Analyze family relationship through feature exaction of multiple faces based on human genetic inheritance, FaceDNA can infer relatives among farther, mother, brother or sister. Without the detection of DNA, matching faces of people will return their internal relationship. Through the face photo of parents, FaceDNA can infer photo of their child using image morphing. The attribute festures of face photo can infer the basic information of people such as Name, Age , Gender, race, identity, clothes, gesture, event, memorability. Various emtions can demonstrate their confidence values in each facial expression. Recognize the emotion of face image, record the changes of emtion in the short video, and capture emotion values in real-time.

To our best knowledge, predicting image memorability [1, 2, 3], using deep learning and LaMem, a novel diverse dataset, initiates a novel method to achieve unprecedented performance at estimating the memorability ranks of images, and evaluate memorability maps. New visual materials could be enhanced using the memorability maps approach, to reinforce forgettable aspects of an image while also maintaining memorable ones [4].

The reminder of paper is organized as follows. Section II describes a comprehensive overview of intelligent face recognition system design with face tracker and detection as well as face matching, face verification, face grouping, automatic memorability predictor. Section III depicts that the methodology of predicting face memorability. Implementation on real experiment environment is introduced in section IV. Experiment results on the previous discussed features are presented in section V. Finally, conclusions are reiterated in section VI.

2 System Design

2.1 FaceDNA

Detect one or more human faces in an image and get back face rectangles for where in the image the faces are, along with face attributes which contain machine learning-based predictions [5, 6] of facial features. After detecting faces, you can take the face rectangle. The face attribute features available are: Age, Gender, Pose, Smile, and Facial Hair along with 27 landmarks for each face in the image.

2.2 Face Verification[7]

Check the likelihood that two faces belong to the same person. The API will return a confidence score about how likely it is that the two faces belong to one person. For this application, if providing photos, please use images which contain only a single face.

2.3 Face Identification[8]

Search and identify faces. Tag people and groups with user-provided data and then search those for a match with previously unseen faces.

2.4 Similar Face Searching

Easily find similar-looking faces. Given a collection of faces and a new face as a query, this API will return a collection of similar faces.

2.5 Face Grouping

Organize many unidentified faces together into groups, based on their visual similarity.

2.6 Face relationship[9]

Analyze family relationship through feature exaction of multiple faces based on human genetic inheritance, inference relatives among father, mother, brother or sister. Just like the detection of DNA, matching faces of people will return their internal relationship.

2.7 Analyze attributes

This feature returns information about visual content found in an image. Use tagging, descriptions and domain-specific models to identify content and label it with confidence. Apply the adult/racy settings to enable automated restriction of adult content. Identify image types and color schemes in pictures. Please try vision feature analysis by uploading a local image, or providing an image URL.

2.8 Recognize celebrities

The Celebrity Model is an example of Domain Specific Models. Our new celebrity recognition model recognizes 200K celebrities from business, politics, sports and entertainment around the World. Domain-specific models is a continuously evolving feature.

2.9 Read text in images

Optical Character Recognition (OCR) detects text in an image and extracts the recognized words into a machine-readable character stream. Analyze images to detect embedded text, generate character streams and enable searching. Allow users to take photos of text instead of copying to save time and effort. Please try vision optical character recognition by uploading a local image, or providing an image URL.

2.10 Recognize Emotions in Images

The Emotion application takes an facial expression in an image as an input, and returns the confidence across a set of emotions for each face in the image, as well as bounding box for the face. A user can submit the face rectangle as an optional input. The emotions detected are anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. These emotions are understood to be cross-culturally and universally communicated with particular facial expressions. Emotion application uses world-class machine learning techniques [10] to provide these results. You can also click the open image button or drag-and-drop to upload your own images, or input a URL for a remote image.

2.11 Recognize Emotions in Video

The Emotion application for Video [11] recognizes the facial expressions of people in a video, and returns an aggregate summary of their emotions. You can use this application to track how a person or a crowd responds to your content over time. The emotions detected are anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise.

2.12 Automatic memorability predictor

This section applies a probabilistic framework that models how and which local regions from an image may be forgotten using a data-driven approach that combines local and global images features. The model automatically discovers memorability maps of individual images without any human annotation. We incorporate multiple image region attributes in this algorithm, leading to improved memorability prediction of images.

Made predictions on the basis of a suite of global image features pixel histograms, GIST, SIFT, HOG, SSIM. Running the same methods on current 2/3 data splits. Do better by using our selected features as an abstraction layer between raw images and memorability. We trained a suite of SVRs to predict annotations from images, and another SVR to predict memorability from these predicted annotations. For annotation types, we used the feature types selected by our 100-bit predictive selection on 2/3 training sets. To predict the annotations for each image in our training set, we spilt the training set in half and predicted annotations for one half by training on the other half, and vice versa, covering both halves with predictions. We then trained a final SVR to predict memorability on the test set in three ways: 1) using only image features (Direct), 2) using only predicted annotations (Indirect), and 3) using both (Direct + Indirect). Combining indirect predictions with direct predictions performed best, slightly outperforming the direct prediction method.

3 Predicting Memorability of face

In this section, we explore various features for predicting face memorability and apply a robust memorability metric to significantly improve face memorability prediction. We also note that the task of automatically predicting the memorability of faces using computer vision features.

For predicting memorability, dense global features such as HOG and SIFT significantly outperform landmark-based features such as ‘shape’ by about 0.15 rank correlation. This implies that it is essential to use these features in our face modification algorithm to robustly predict memorability after making modifications to a face. While powerful for prediction, the dense global features tend to be computationally expensive to extract, as compared to shape. Shape is used in this algorithm to parameterize faces so it essentially has zero cost of extraction for modified faces. However, as compared to memorability, the gap in performance between using shape features and dense features is not as large for other attributes. Hence, we use landmark-based features instead of dense global features for the modification of facial attributes.

4 Experiment Implement

Development environment: Windows 10 on MacBook Air, Microsoft Visual studio professional 2015, Inter realsense camera F200, Intel realsense SDK 2016 R1, SDK Runtime Distributable, F200 Camera Driver (Depth Camera Manger), Microsoft Cognitive Services.

Pre-trained CNN using Caffe deep learning toolbox, it provides the network deploy file, the trained network model, and the train-val file which can be loaded using Caffe. Extract CaffeNet / AlexNet features using the Caffe utility. MemNet trained on first train/test split of LaMem, with FA, rank correlation with 10 crops per image 0.64. Then randomly select 500 different images in emotion from their dataset [12], and collected 80 scores per image. After applying this algorithm to correct the memorability scores, it can obtain a within-dataset human rank correlation of 0.77 (averaged over 25 random splits). Furthermore, it can obtain a rank correlation of 0.76.

5 Experiment Result

Fig.1 demonstrates emotion and gaze detection, face tracking by Intel Realsense 3D camera. Fig.2 indicates FaceDNA extracts various attributes of face photos. FaceDNA includes face detection, verification, face matching, and face grouping modules. Fig.3 depicts emotion detection recognizes the expression of face image, record the changes of emotion in the video, and capture emotion values in real-time. Fig.4 plots the relationship of memorability scores and various emotion image attributes.

6 Conclusions

This paper builds an intelligent face recognition system, which can be applied into Intel Realsense 3D camera. FaceDNA is the new feature in analyzing the face images captured by Intel Realsense 3D Camera. It covers the most advanced technical solutions for face tracking and recognition. The key contribution is to apply a novel face recognition system in realizing cognitive services. Moreover, this system applies computer vision techniques to extract memorability automatically.

7 References

[1].A. Khosla, A. S. Raju, A. Torralba and A. Oliva, Understanding and Predicting Image Memorability at a Large Scale, International Conference on Computer Vision (ICCV), 2015.

[2].Aditya Khosla, Wilma A. Bainbridge, Antonio Torralba and Aude Oliva, Modifying the Memorability of Face Photographs, International Conference on Computer Vision (ICCV), 2013.

[3].Phillip Isola, Jianxiong Xiao, Antonio Torralba, Aude Oliva. What makes an image memorable?, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[4].R. Dubey, J. Peterson, A. Khosla, M.-H. Yang, and B. Ghanem. What makes an object memorable? In International Conference on Computer Vision (ICCV), 2015.

[5].Shengcai Liao, Anil K. Jain, and Stan Z. Li, "A Fast and Accurate Unconstrained Face Detector," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.

[6].X. Wang and X. Tang. Random sampling for subspace face recognition. International Journal of Computer Vision, 70(1), 2006.

[7].N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. Attribute and Simile classifiers for face verification. In Proc. IEEE International Conference on Computer Vision (ICCV), 2009.

[8].M. Guillaumin, J. Verbeek, C. Schmid, I. LEAR, and L. Kuntzmann. Is that you? Metric learning approaches for face identification. IEEE International Conference on Computer Vision (ICCV), 2009.

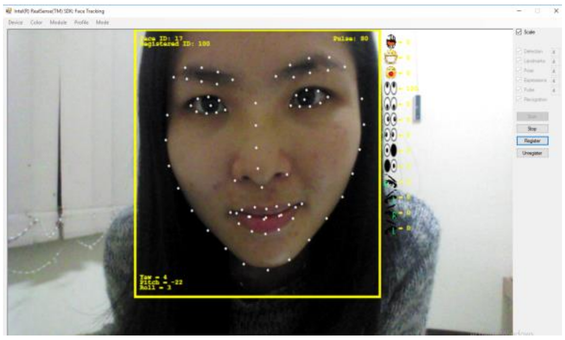
[9].Y. Su, S. Shan, X. Chen, and W. Gao. Adaptive generic learning for face recognition from a single sample per person. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

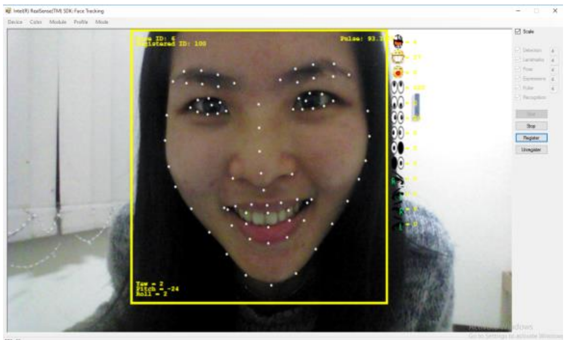
[10].Z. Cao, Q. Yin, J. Sun, and X. Tang. Face recognition with Learning-based Descriptor. In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

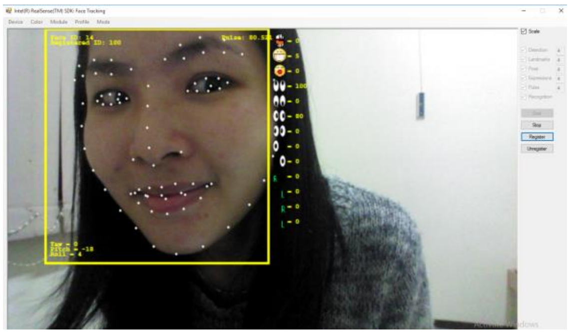
[11].F. De la Torre, W.S. Chu, X. Xiong, F. Vicente, X. Ding, and J. F. Cohn. IntraFace, IEEE International Conference on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia. , 2015.

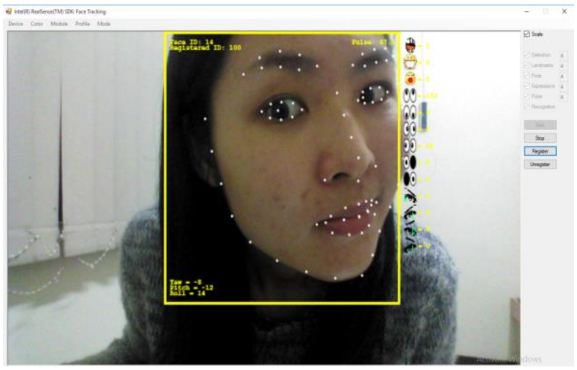
[12].N.Pinto,J.DiCarlo,and D.Cox.How far can you get with a modern face recognition test set using only simple features. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2009.

Fig.1. Face tracker









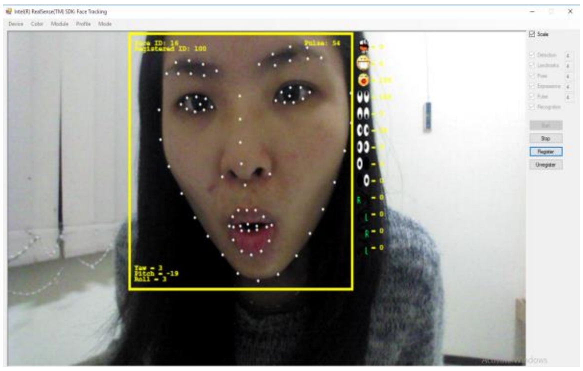
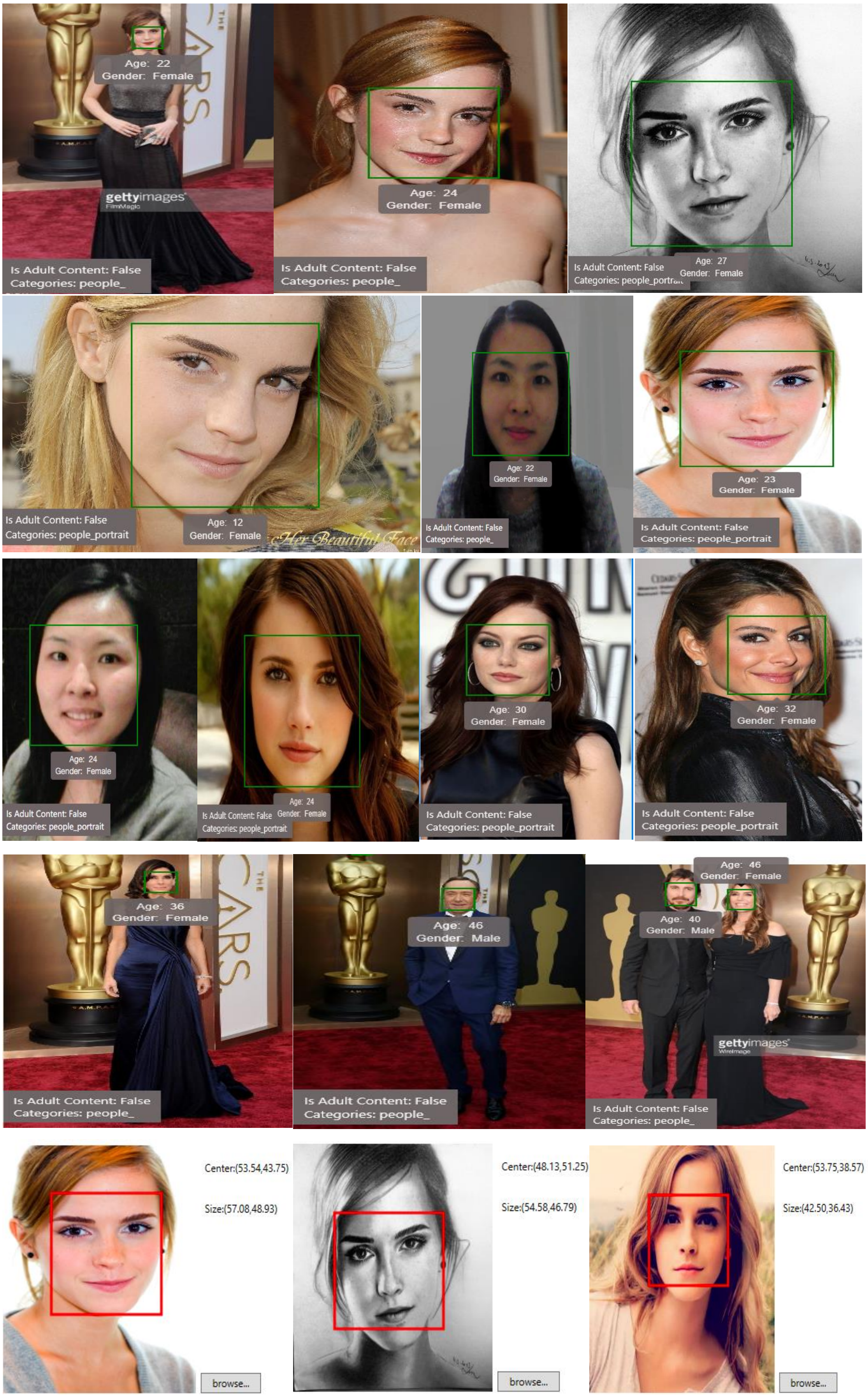
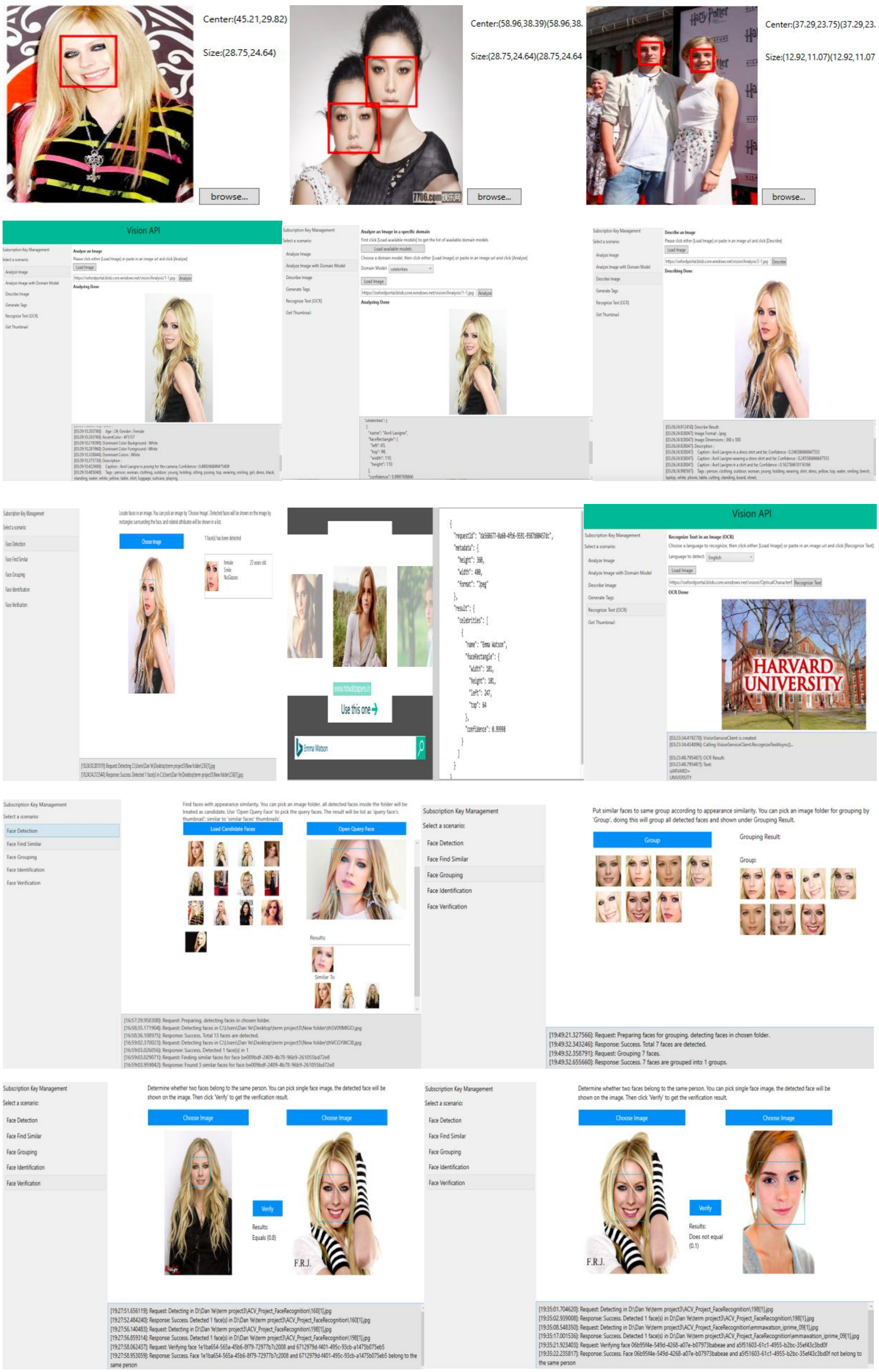


Fig.2.FaceDNA



ISBN: 1-60132-442-1, CSREA Press ©


Center:(45.21,29.82)
Size:(28.75,24.64)
browse...
Center:(58.96,38.39)(58.96,38.
Size:(28.75,24.64)(28.75,24.64
browse...
Center:(37.29,23.75)(37.29,23.
Size:(12.92,11.07)(12.92,11.07
browse...
Vision API
Vision API
HARVARD UNIVERSITY
Use this one
Group
Grouping Result:
Choose Image
Choose Image
Verify
Choose Image
Choose Image
Verify

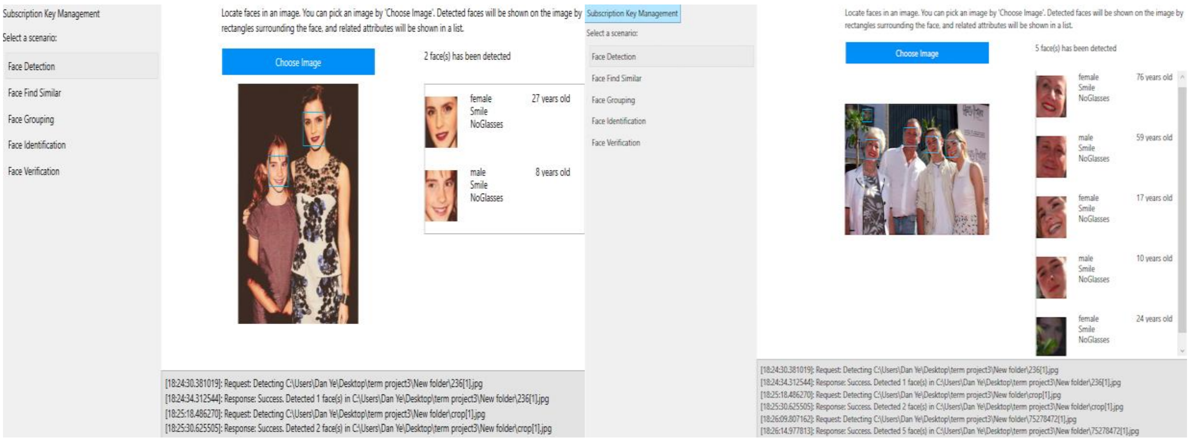
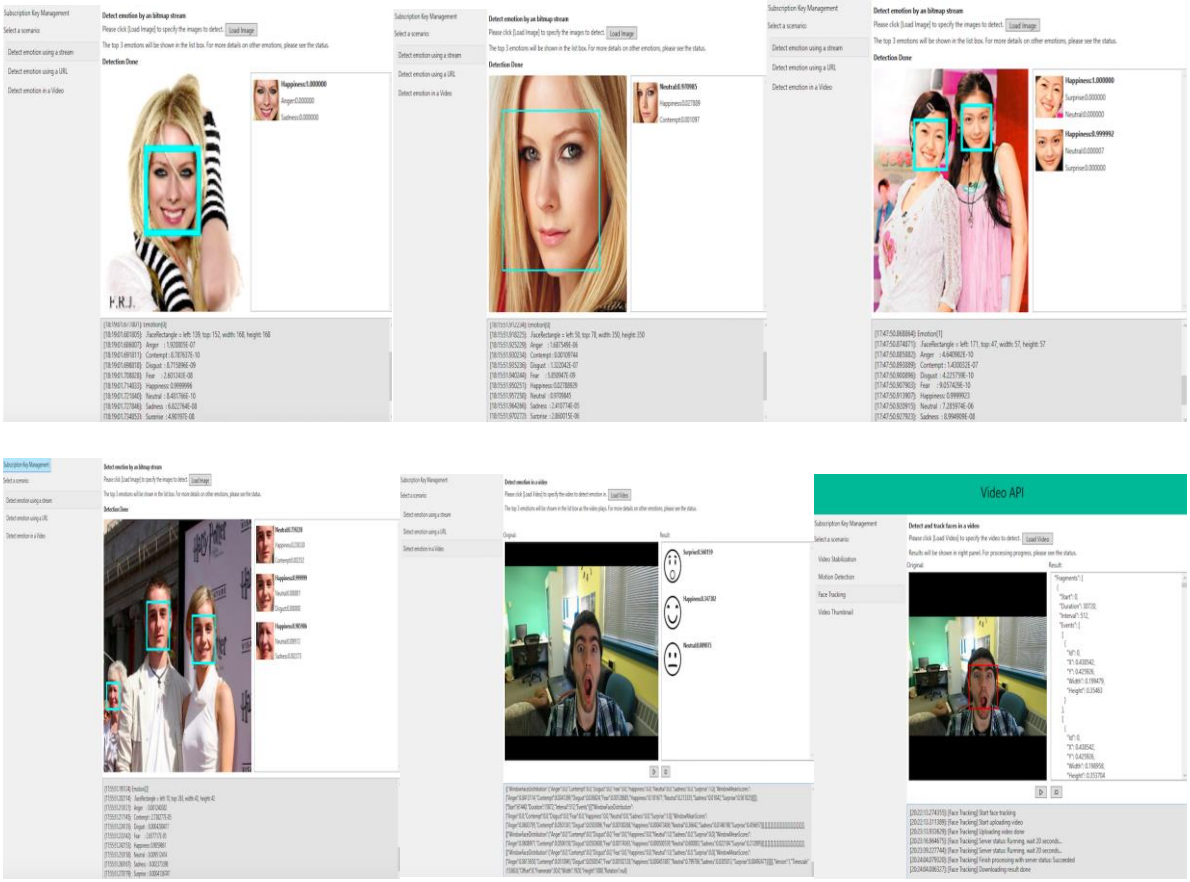


Fig.3. Emotion detection



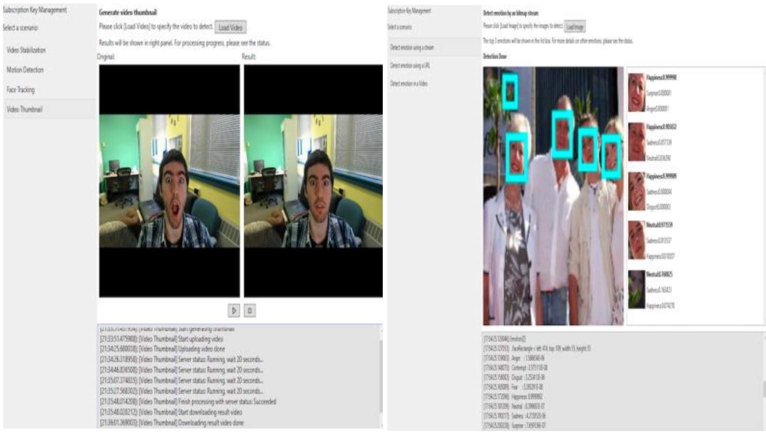


Fig.4. Memorability vs emotions