

Exploring the Effect of Illumination on Automatic Expression Recognition using the ICT-3DRFE Database

Giota Stratou, Abhijeet Ghosh, Paul Debevec, Louis-Philippe Morency

Institute for Creative Technologies, University of Southern California

Abstract

One of the main challenges in facial expression recognition is illumination invariance. Our long-term goal is to develop a system for automatic facial expression recognition that is robust to light variations. In this paper, we introduce a novel 3D Relightable Facial Expression (ICT-3DRFE) database that enables experimentation in the fields of both computer graphics and computer vision. The database contains 3D models for 23 subjects and 15 expressions, as well as photometric information that allow for photorealistic rendering. It is also facial action units annotated, using FACS standards. Using the ICT-3DRFE database we create an image set of different expressions/illuminations to study the effect of illumination on automatic expression recognition. We compared the output scores from automatic recognition with expert FACS annotations and found that they agree when the illumination is uniform. Our results show that the output distribution of the automatic recognition can change significantly with light variations and sometimes causes the discrimination of two different expressions to be diminished. We propose a ratio-based light transfer method, to factor out unwanted illuminations from given images and show that it reduces the effect of illumination on expression recognition.

Keywords: 3D facial database, illumination effect, image re-lighting, facial expression recognition, ratio images

1. Introduction

One of the main challenges with facial expression recognition is to achieve illumination invariance. Prior studies show that changing the direction of illumination can influence the perception of object characteristics such as 3D shape and location [1]. Relative to common image representations, changes in lighting result in large image differences. These observed changes can be larger even than when varying the identity of the subject [2].

These studies suggest that both human and automated facial identification are impaired by variations in illumination. By extension, we expect a similar impediment to facial expression recognition. This intuition is strengthened by four observations: i) changes in facial expression are manifested as deformation of the shape and texture of the facial surface, ii) illumination variance has been shown to influence perception of shape, which confounds face recognition, iii) most methods for automated expression recognition use image representations, features, and processing techniques similar to face recognition methods [3] which are also confounded by illumination variance, and iv) the training set for most classifiers consists mainly of uniformly lit images.

While most automatic systems for facial expression recognition assume input images with relatively uniform illumination, research such as Li *et al.* [4], Kumar *et al.* [5] and Toderici *et*

al. [6] have worked toward illumination invariance by extracting features which are illumination invariant. To serve this direction of research, facial databases have been assembled which capture the same face and pose under different illumination conditions, and lately the development of 3D facial databases has become of interest, since they allow exploration of new 3D features.

In this paper, we introduce a novel 3D Relightable Facial Expression (ICT-3DRFE) database which enables studies of facial expression recognition and synthesis. We demonstrate the value of having such a database while exploring the effect of illumination on facial expression recognition. First, we use the ICT-3DRFE database to create a sample database of images to study the effect of illumination. We use the Computer Expression Recognition Toolbox (CERT) [7] to evaluate specific Facial Action Units (AU) on that image set and we compare CERT output with a FACS (Facial Action Coding System) expert coder's annotations. We also compare the CERT output of specific expressions under different illumination to observe how lighting variation affects its ability to distinguish between expressions. Second, we present an approach to factor out lighting variation to improve the accuracy of automatic expression recognition. For this purpose, we employ *ratio images* as in the approach of Peers *et al.* [8], to transfer the uniformly-lit appearance of a similar face in the ICT-3DRFE database to a target face seen under non-uniform illumination. In this approach, we use the ICT-3DRFE database to select a matching subject and transfer illumination. We evaluate if "unlighting" a face in this way can improve the performance of expression recognition software. Our experiments show promising results.

Email addresses: stratou@ict.usc.edu (Giota Stratou), ghosh@ict.usc.edu (Abhijeet Ghosh), debevec@ict.usc.edu (Paul Debevec), morency@ict.usc.edu (Louis-Philippe Morency)

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 04 NOV 2011		2. REPORT TYPE		3. DATES COVERED 00-00-2011 to 00-00-2011	
4. TITLE AND SUBTITLE Exploring the Effect of Illumination on Automatic Expression Recognition using the ICT-3DRFE Database				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Southern California, Institute for Creative Technologies (ICT), 12015 Waterfront Drive, Playa Vista, CA, 90094				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT One of the main challenges in facial expression recognition is illumination invariance. Our long-term goal is to develop a system for automatic facial expression recognition that is robust to light variations. In this paper, we introduce a novel 3D Relightable Facial Expression (ICT-3DRFE) database that enables experimentation in the fields of both computer graphics and computer vision. The database contains 3D models for 23 subjects and 15 expressions, as well as photometric information that allow for photorealistic rendering. It is also facial action units annotated, using FACS standards. Using the ICT-3DRFE database we create an image set of different expressions/illuminations to study the effect of illumination on automatic expression recognition. We compared the output scores from automatic recognition with expert FACS annotations and found that they agree when the illumination is uniform. Our results show that the output distribution of the automatic recognition can change significantly with light variations and sometimes causes the discrimination of two different expressions to be diminished. We propose a ratio-based light transfer method, to factor out unwanted illuminations from given images and show that it reduces the effect of illumination on expression recognition.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 10	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

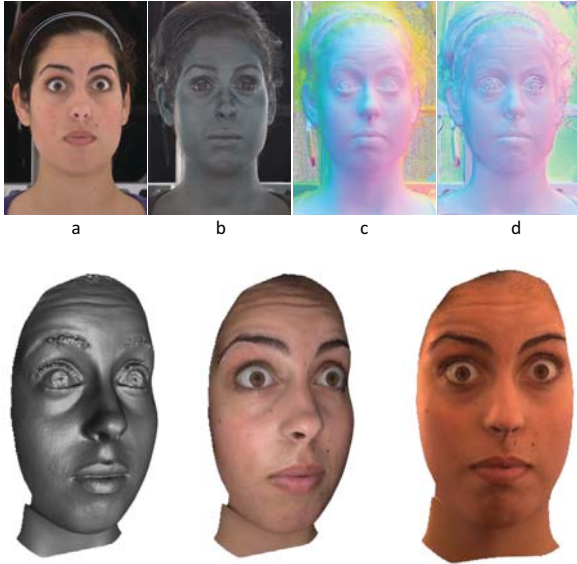


Figure 1: A sample 3D model from ICT-3DRFE and some of its corresponding textures and photometric surface normal maps. First row, a) diffuse texture, b) specular texture, c) diffuse (red channel) normals, d) specular normals. Second row, from left to right: 3D geometry of a subject posing for the “eyebrows up” expression; same pose rendered with texture and simple pointlights; different pose of the model rendered under environmental lighting.

The remainder of this paper is arranged as follows: in Section II, we discuss the previous work on automatic facial expression recognition. We also survey the state-of-the-art in facial expression databases and mention other face relighting techniques relevant to facial expression recognition. In Section III, we introduce our new ICT-3DRFE database, discussing its advantages and how it was assembled. Section IV describes our experiment on the effect of illumination on facial expression recognition using the ICT-3DRFE database. Section V describes our illumination transfer technique for mitigating the effects of illumination on expression recognition, showing how this improves AU classification. We conclude with a discussion of future work in Section VI.

2. Previous Work

2.1. Facial Expression Recognition

There has been significant progress in the field of facial expression recognition in the last few decades. Two popular classes of facial expression recognition are: i) facial Action Units (AUs) according to the Facial Action Coding System (FACS) proposed by Ekman *et al.* [10] and ii) the set of prototypic expressions also defined by Ekman [11] that relate to emotional states including happiness, sadness, anger, fear, disgust and surprise. Systems of automatic expression recognition commonly use AU analysis as a low level expression classification followed by a second level of classification of AU combinations into one of the basic expressions [13]. Traditional automatic systems use geometric features such as the location of facial landmarks (corners of the eyes, nostrils, etc.) and spatial relations among them (shape of eyes, mouth, etc.) [3] [12].

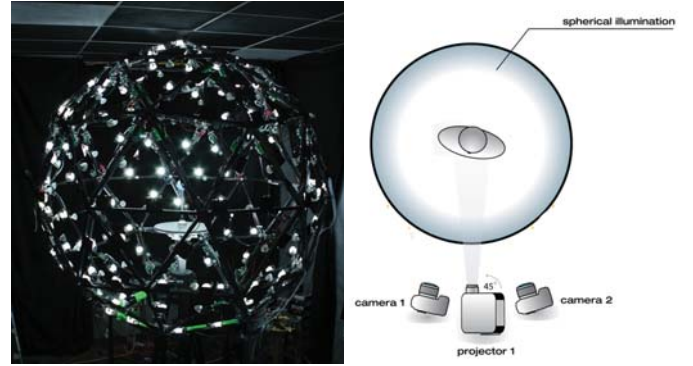


Figure 2: Acquisition setup for ICT-3DRFE. Left: LED sphere with 156 white LED lights. Right: Layout showing the positioning of the stereo pair of cameras and projector for face scanning.

Bartlett *et al.* found in practice that image-based representations contain more information for facial expression than representations based on shape only [14]. Recent methods focus either on solely appearance features (representing the facial texture) like Bartlett *et al.* [14] who use Gabor wavelets or eigenfaces, or hybrid methods, using both shape- and appearance-based features, like in the case of Lucey *et al.* which uses an Active Appearance Model (AAM) [15]. There is also a rising interest in the use of 3D facial geometry to extract expression representations that will be view and pose invariant [13].

2.2. Facial Databases

Facial expression databases are very important for facial expression recognition, because there is a need for common ground to evaluate various algorithms. These databases are usually static images or image sequences. The most commonly used facial expression databases include the Cohn-Kanade facial expression database [16] which is AU coded, the Japanese Female Facial Expression (JAFPE) database [17], MMI database [18] which includes both still images and image sequences, the CMU-PIE database [19], with pose and illumination variation for each subject, and other databases [20]. Since the introduction of 3D into facial expression recognition, 3D databases have gained in popularity. The most common is the BU-3DFE database which includes 3D models and considers intensity levels of expressions [21]. BU-3DFE was extended to the BU-4DFE by including temporal data [22]. The latest facial expression databases are the Radboud Facial Database (RaFD), which considers contempt, a non prototypic expression and different gaze directions [23], and the extended Cohn-Kanade (CK+) database, which is an extension of the older CK, is fully FACS coded and includes emotion labels [24].

Our new ICT-3DRFE database also includes 3D models, considers different gaze directions, and is AU annotated. In contrast to the other databases, however, our ICT-3DRFE database offers much higher resolution in its 3D models, and it is the only photorealistically relightable database.

2.3. Face Relighting

One of our ultimate goals is to factor out the effect of illumination on facial expression recognition. For that, we leverage

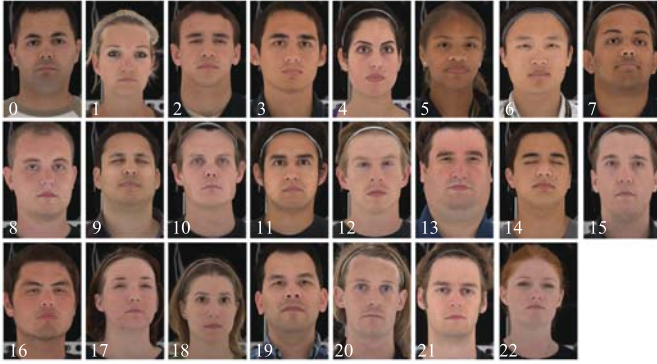


Figure 3: The 23 subjects of ICT-3DRFE database.

image based relighting techniques which have been extensively studied in computer graphics. Debevec *et al.* [26] photographs a face with a dense sampling of lighting directions using a spherical *light stage* and exploits the linearity of light transport to accurately rendering the face under any distant illumination environment from such data. While realistic and accurate, the technique can be applied only to subjects captured in a light stage. Peers *et al.* [8] overcame this restriction through an *appearance transfer* technique based on ratio images [9], allowing a single photograph of a face to be approximately relit using light stage data of a similar-looking subject from a database. Ratio images have also been used to transfer facial expressions from one image to another by Liu *et al.* [25] and for facial relighting [27]. More recent work, has been presented by Chen *et al.* [28] using Edge-preserving filters for face illumination transfer. A few other researchers have explored relighting methods to enhance facial recognition: Kumar *et al.* [5] uses morphable reflectance fields to augment image databases with relit images of the existing set, Toderici *et al.* [6] uses bidirectional relighting and Wang *et al.* [29] use a spherical harmonic basis morphable model (SHBMM).

Our approach to factor out the effect of illumination from a target face is similar in principle to that of Peers *et al.* [8] with the difference that while they relight a uniformly illuminated target face to a desired non-uniform lighting condition, our goal is more similar to Wang *et al.* [29], that is, we relight the target face image from a known non-uniform lighting condition to a uniform lighting condition for robust facial expression recognition, and we are especially interested in the case of extreme lighting conditions.

3. ICT-3DRFE Dataset

The main contribution of this paper is the introduction of a new 3D Relightable Facial Expression Database¹. As with any 3D database, a great advantage of having 3D geometry is that one can use it to extract geometric features that are pose and viewpoint invariant. In our ICT-3DRFE database, the detail of the geometry is higher than in any other existing 3D database, with each model having approximately 1,200,000 vertices with

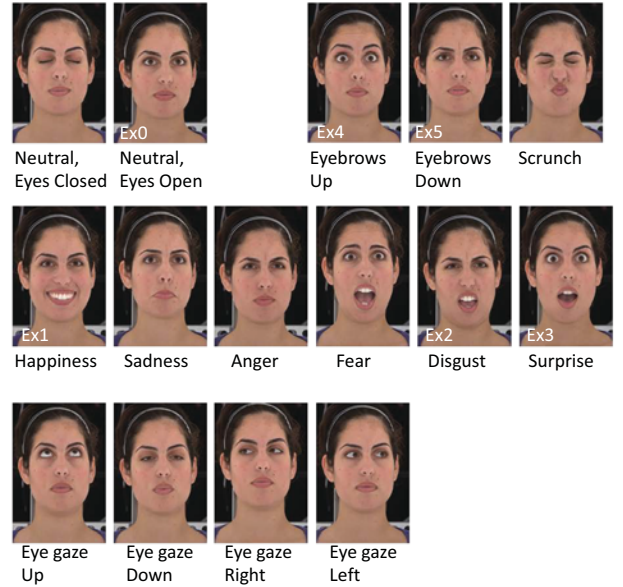


Figure 4: The 15 expressions captured for every subject. The ones annotated with an "Ex" label are the expressions used in our experiments.

reflectance maps of 1296×1944 pixels. This resolution contains detail down to sub-millimeter skin pore level, increasing its utility for the study of geometric and 3D features. Besides high resolution, relightability is the other main novelty of this database. The reflectance information provided with every 3D model allows the faces to be rendered realistically under any given illumination. For example, one could use a *light probe* [32] to capture the illumination in a specific scene to render a face in the ICT-3DRFE database with that lighting. This property, along with the traditional advantages of a 3D model database (such as controlling the pose while rendering) enables many uses. In Section IV, we use our ICT-3DRFE to study the effect of illumination on facial expressions by creating a database of facial images under chosen illumination conditions and poses. Also in Section V, we use the database as a tool for removing illumination effects from facial images. Figure 1 displays a sample 3D model from the ICT-3DRFE database under different poses and illuminations.

3.1. Acquisition Setup

The ICT-3DRFE dataset introduced in this paper was acquired using a high resolution face scanning system that employs a spherical *light stage* with 156 white LED lights (Figure 2A). The lights are individually controllable in intensity and are used to light the face with a series of controlled spherical lighting conditions which reveal detailed shape and reflectance information. An LCD video projector subsequently projects a series of colored stripe patterns to aid stereo correspondence. The face's appearance under these conditions is photographed by a stereo pair of Canon 1D Mark III digital cameras (10 megapixels) (Figure 2B). Computational stereo between the two cameras produces a millimeter-accurate estimate of facial shape; this shape is refined using sub-millimeter surface orientation estimates from the spherical lighting conditions as in Ma *et al.* [30], revealing fine detail at the level of pores and creases.

¹This database is publicly available at <http://projects.ict.usc.edu/3drfe/>

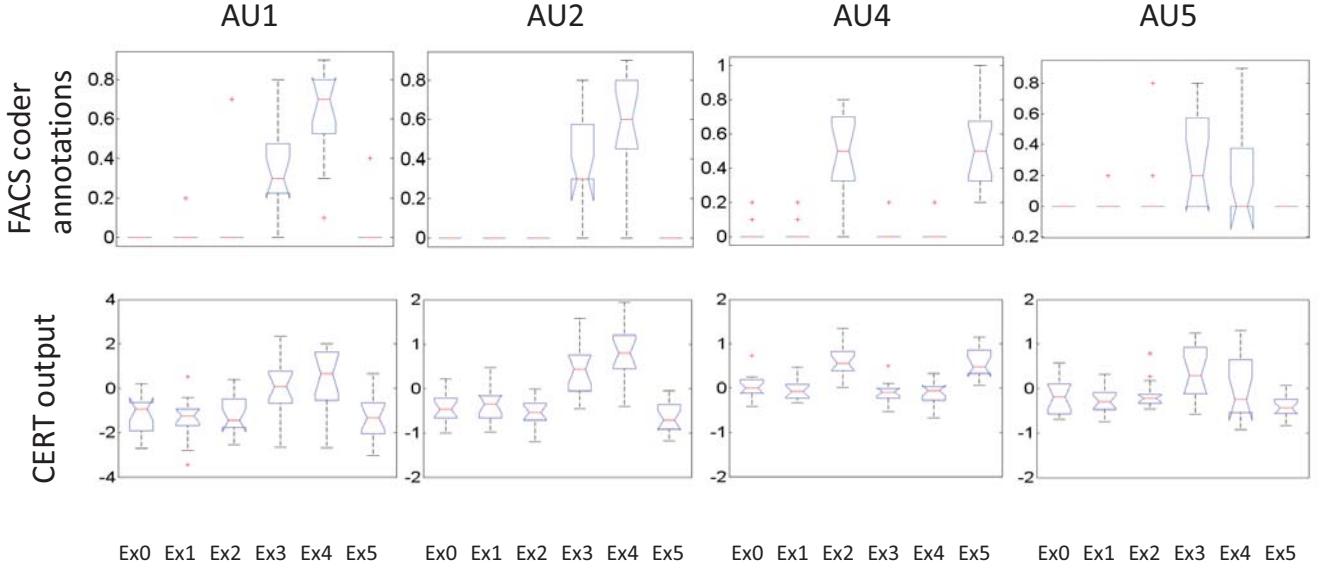


Figure 5: Distribution of AU scores for a selected set of expressions (see Table 2) under uniform illumination. Top: distribution of AU scores, as annotated by expert FACS coder. Bottom: distribution of AU output from CERT, a system for automatic facial expression recognition [7]. From these graphs, it becomes obvious that Ex3 (surprise) and Ex4 (eyebrows-up) have different degrees of eyebrows up (expressed among others by AU1 and AU2), and Ex2 (disgust), Ex5 (eyebrows-down) include a frown (expressed by AU4).

Linear polarizer filters on the LED lights and an active polarizer on the left camera allow specular reflections (the shine off the skin) and subsurface reflection (the skin’s diffuse appearance) to be recorded independently, yielding the diffuse and specular reflectance maps (Figure 1) needed for photorealistic rendering under new lighting.

Each facial capture takes five seconds, acquiring approximately 20 stereo photographs under the different lighting conditions. Our subjects had no difficulty maintaining the facial expressions for the capture time, particularly since we used the complementary gradient technique of Wilson *et al.* [31] to digitally remove subject motion during the capture.

3.2. Dataset Description

For the purpose of this dataset, 23 people were captured, as represented in Figure 3. Our database consists of 17 male and 6 female subjects from different ethnic backgrounds, all between the ages of 22-35. Each subject was asked to perform a set of 15 expressions, as shown in Figure 4.

The set of posed expressions consists of the six prototypic ones (according to Ekman [11]), two neutral expressions (eyes closed and open), two eyebrow expressions, a scrunched face expression, and four eye gaze expressions (see Figure 4). For the six emotion driven expressions (middle row), the subjects were given the freedom to perform the expression as naturally as they could, whereas for the action specific expressions the subjects were asked to perform specific facial actions. Our motivation for this was to capture some of the variation with which people express different emotions, and not to force one standardized face for each expression.

Each model in the database contains high-resolution (sub-millimeter) geometry as a triangle mesh, as well as a set of high-

resolution reflectance maps including a diffuse color map (like a traditional ”texture map”, but substantially without ”baked-in” shading), a specular intensity map (how much shine each part of the face has), and several surface normal maps (indicating the local orientation of each point of the skin surface). Normal maps are provided for the red, green, and blue channels of the diffuse component as well as the colorless specular component to enable efficient and realistic skin rendering using the *hybrid normal* technique of Ma *et al.* [30].

3.3. Action Unit Annotations

Our ICT-3DRFE database is also fully AU annotated from an expert FACS coder. Action units are assigned scores between 0-1 depending on the degree of muscle activity. In Figure 5, we show the distribution of the scores for some eyebrow related AU and for the subject/expression set we have chosen for further analysis in this paper.

The displayed AU are: AU1 (inner brow raise), AU2 (outer brow raise), AU4 (brow lower) and AU5 (upper lid raise). The AU score distribution over different expressions demonstrates which AUs are activated in a specific expression and to what degree. For example, from Figure 5, first row, we can tell that expressions Ex3 and Ex4 (surprise and eyebrow-up, respectively) usually employ inner and outer eyebrow raise since they have both AU1 and AU2 activated. Moreover, we can tell that during expression Ex4 subjects tend to raise their inner eyebrow more than during Ex3, because of the distribution of the scores (the degree of AU1 is different between these two expressions). Similarly, among the selected set of expressions, only Ex2 and Ex5 (disgust and eyebrows-down, respectively) include a frown, which is represented by AU4.

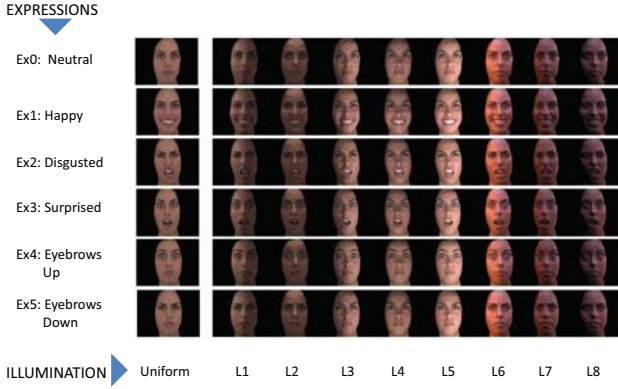


Figure 6: Example of expressions and illumination conditions used in our experiment. Illuminations are the same column-wise, and expressions are the same row-wise.

4. Influence of Illumination on Expression Recognition

In this section, we explore and quantify the illumination effect on expression recognition. For the scope of this study we focus on automatic recognition of facial expressions. We evaluate automatic classification of AU's, since they are the prevailing classification method for facial expressions. We intend to find patterns in the variation of AU response when changing the illumination (either during expression or during a neutral face) and explore which characteristics of illumination affect specific facial AUs.

We decided to focus our first effort on investigating eyebrow facial actions, with the intuition that this area of the face is one of the most expressive ones. Muscle activation along the eyebrows causes big shape and texture variation during expressions.

We set our experiment goals as follows: i) we examine the correlation of our expert FACS coder's annotation with the AU output from automatic expression recognition, ii) we explore the changes in automatic recognition output caused by illumination variation on the neutral face, and iii) we examine if two different expressions, distinguished by different AU scores, remain separable to the same degree when illumination changes.

4.1. Evaluation Methodology

First, we need to create an image set of different facial expressions and under different illumination conditions. Based on the analysis of the FACS annotated AU scores, we chose a set of expressions which activate eyebrow related AUs. Specifically, we picked six expressions for our study, as described in Table 2. Expressions Ex2-Ex5 are chosen because they usually come with intense eyebrow activation, and the first two (Ex0-Ex1) for calibration of what consists of neutral and close to neutral for eyebrow motion, respectively.

For our lighting set, we chose nine different illumination conditions, as seen in Figure 6. They are described in Table 1. The first one (L0) is picked to evaluate the best performance for CERT, since it is a uniform lighting, desirable for automatic facial expression recognition systems. This illumination is uniform. L1-L5 are picked because of the directionality which is

Table 1: Illumination Configurations for Experiments Described in Section IV

Label	Description
L0	uniform
L1	ambient + pointlight at the right side of the head
L2	ambient + pointlight at the top side of the head
L3	ambient + pointlight at the the left
L4	ambient + pointlight at the the bottom
L5	ambient + pointlight at the the bottom left
L6	environmental light
L7	environmental light (modified 1)
L8	environmental light (modified 2)

Table 2: Selected Expressions for Experiments Described in Section IV

Label	Description
Ex0	Neutral - eyes open
Ex1	Happy
Ex2	Disgusted
Ex3	Surprised
Ex4	Eyebrows Up
Ex5	Eyebrows Down

one of the main parameters that impairs shape perception. L6-L8 are picked as representatives of more realistic, environmental lighting conditions that one can actually come across. L7 and L8 are also cases of low illumination intensity.

To produce our experimental image set for analysis, we used our newly developed ICT-3DRFE database. The image set for one of the subjects can be seen in Figure 6. All 3D models were rendered under the same 6 expressions and 9 illumination conditions. We did that for a subset of fifteen subjects, generating $6 \times 9 = 54$ images for each subject.

For the automatic evaluation of AUs, we used Computer Expression Recognition Toolbox (CERT) [7], which is a robust AU classifier that uses appearance based features [14] and performs with great accuracy. Using CERT we obtained output for some eyebrow related AUs.

4.2. Results

First, we want to evaluate the correlation of CERT output with our FACS coder annotations. AU1, AU2, AU4 and AU5 output evaluated with CERT are shown in Figure 5, second

Table 3: Correlation coefficients between human FACS coder and computer system (CERT) output. In the first column, we look at the scores for all subjects and expressions, whereas in the second column we look at the correlation of the distribution mean over all expressions

Action Unit	Subject-wise Correlation	Distribution mean Correlation
AU1	0.400	0.984
AU2	0.618	0.984
AU4	0.724	0.986
AU5	0.250	0.954
AU9	0.035	0.891
AU12	0.672	0.967

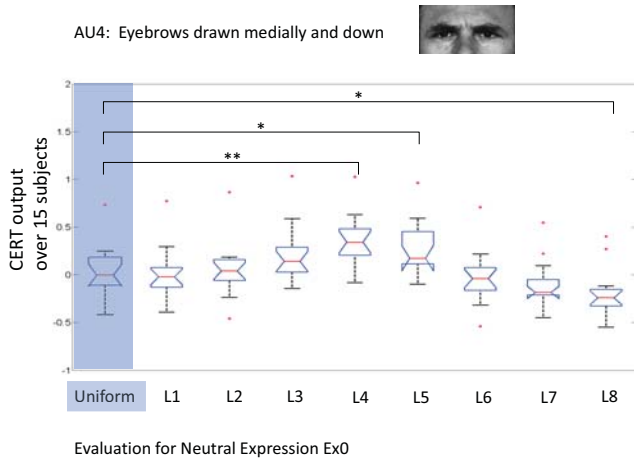


Figure 7: Effect of illumination on automatic facial expression recognition of neutral face. Demonstrate effect on facial action unit AU4. Results from paired

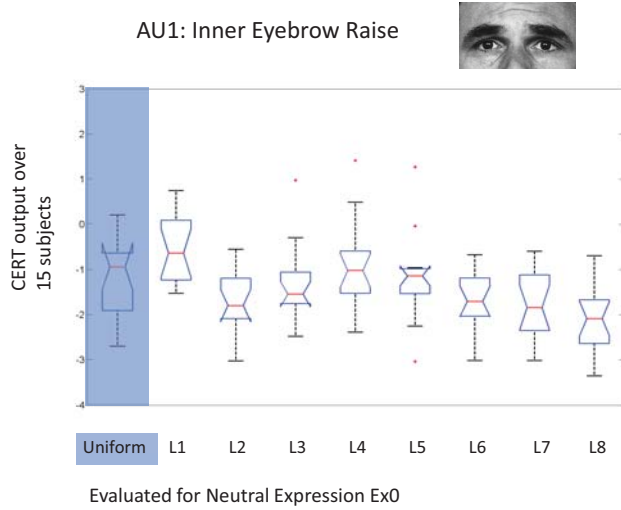


Figure 8: Effect of illumination on automatic facial expression recognition of neutral face. Demonstrate effect on facial action unit AU1

row, below the expert FACS AU annotations. Note that in both cases, uniform illumination condition (L0) was used. CERT output are the Support Vector Machine margins from classification and can be positive or negative [7], whereas the annotated scores range from 0-1 with 1 signifying the highest intensity and 0 meaning that the AU is non existent. Although CERT was trained as a discriminative classifier, Figure 5 shows that its output is directly correlated with the expert FACS coder's annotation.

More specifically, in Table 3 we performed the correlation analysis (Pearson's linear correlation coefficient) between the scores of the expert FACS coder and CERT output (SVM classification margins) for our subject set of 15 people and the 6 expressions (as described in Table 2). The first column shows the correlation of the AU intensities over the data series of all the 15 people $\times$ 6 expressions = 90 values, whereas in the second column, we took the average over all subjects of the distributions of FACS scores and CERT scores per expression (6 values), and calculated the correlation between those series. In the first case, when evaluating over all subjects and expressions, the correlation between human coder and CERT is good for AU2, AU4 and

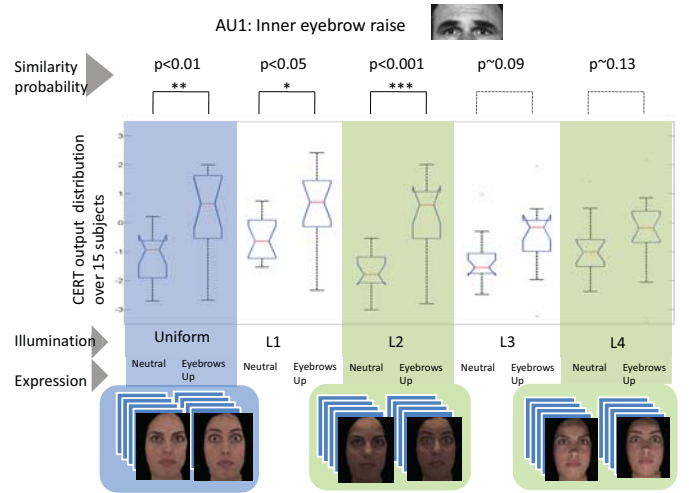


Figure 9: Effect of illumination on separability of neutral expression and eyebrows up expression, with respect to AU1 (inner eyebrow raise).

AU12. This is very good, given that no normalisation has been performed at this stage and given the variability of the scores because of subject properties. Note that the AUs that got lower correlation scores (Table 3) are the ones that were less intense in their activation, thus making easier to confuse inter-subject variance with the variance from different expressions. AU12 which presented itself more intensely in the chosen expression set, shows better correlation score. In the second column, where we compare the distribution means, we observed high correlation values of the average score per expression, something that one can confirm visually from Figure 5. Indeed, the distribution patterns are similar to those of the annotated scores, which validates the CERT output on our image set and certifies that the uniform illumination condition is indeed a suitable input to establish ground truth.

To answer our second question about the AU variation with illumination on a neutral face, we plot the distributions of CERT AU output for the different illumination conditions, evaluated on neutral faces. Figure 7 lays such a plot for AU4 (eyebrows drawn medially and down). The first distribution (first highlighted column on the left) are the scores for uniform light and we consider it to be the ground truth AU score for the neutral face. From Figure 7 we observe that AU4 output changes with illumination and more specifically, illumination conditions L4 and L5 (directionality from the bottom, and bottom left) seem to affect it the most. To analyze the statistical significance of the variation in AU scores, we performed the paired T-Test with a standard 5% significance level, annotated in the Figures with a "**", and with "***" a significance level of 1% .

Similarly, we performed more experiments for some other eyebrow related AUs and we observed that: i) light from the side affects AU1 (inner eyebrow raised) the most (Figure 8), ii) light from the top or bottom affects AU9 (nose wrinkle) the most (Figure not shown). These observations agree with our intuition.

Our third topic of interest is to understand whether different expressions remain distinguishable under different illumination. To answer this, we examine the distributions of a specific AU output under different illumination conditions for an

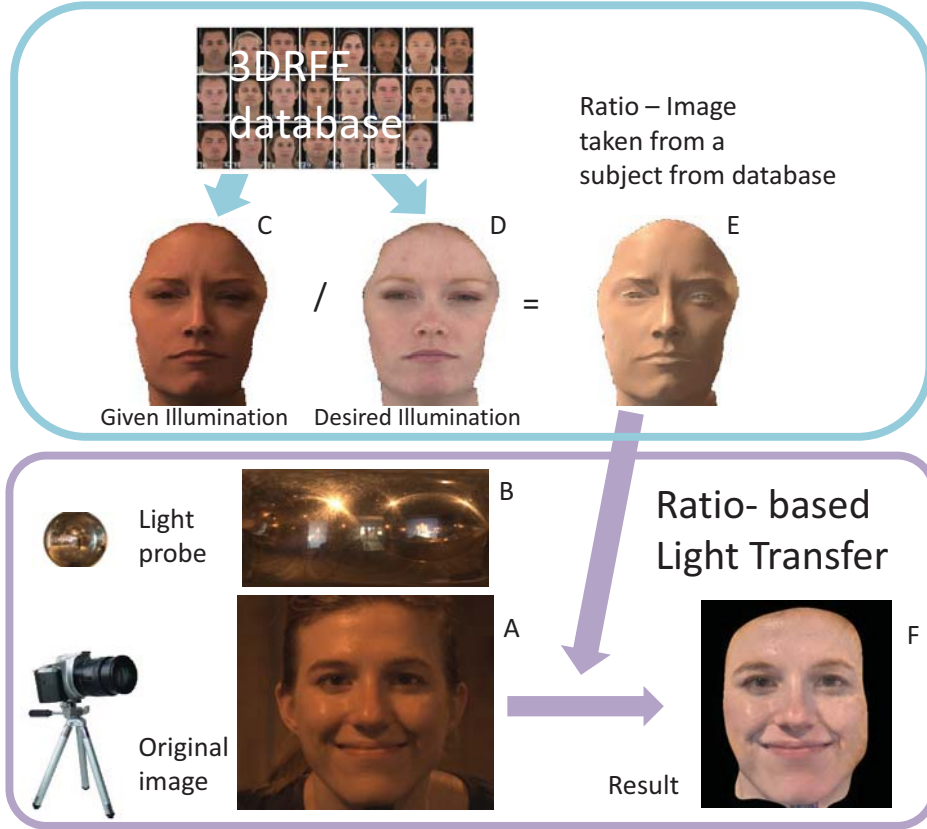


Figure 10: Overview of Ratio-based Light Transfer method: Factoring out a specific illumination from a target image. An original image (A) is taken under known illumination (B). A subject from the ICT-3DRFE database is selected and brought to a similar pose as the target. We then render the database subject under the same illumination as the target (C) and under a desired uniform illumination (D). From these two images (C and D), we obtain the ratio image (E) for the database subject. This ratio image represents the light difference between the two illuminations, the one we have (B) and the one we want (uniform). The ratio image is then used to transfer the desired illumination to our target image as seen in the result (F).

expression that includes this AU and for the neutral face. So this time we are looking at pairs of AU scores, and how their correlation changes with illumination variation.

In Figure 9, we show such analysis for AU1 (inner eyebrow raise), when comparing the neutral expression with the eyebrows-up expression (Ex4). Neutral expression does not include strong AU1 activation, whereas eyebrows-up expression does include high scores of AU1 (reference Figure 5), so the distributions of CERT output for AU1 should be separable as in the first (highlighted left) column of Figure 9, under uniform illumination. However, the discrimination between the two very different expressions is blurred with the change of illumination, as seen in the rest of the columns of Figure 9. Specifically, we observe that under illumination L1, the distinction between neutral and eyebrows up expression becomes a little bit more difficult but still possible. Illumination L2 has the opposite effect, since it makes the neutral and eyebrows up expression even more separable. Illuminations L3 and L4 are making the two expression distributions statistically similar. Also, looking at just the distributions for the neutral illumination, we observe again as mentioned earlier in the result section, that light from the side (L1) causes the distribution of the output to become statistically different from the one under

uniform illumination. Similar observations were made for other AUs (Figures not shown). For example, the expression of disgust (Ex1) is highly distinguishable from the neutral expression (Ex0) under uniform illumination, with respect to AU9 (nose wrinkle). However, neutral expression scores of AU9 become almost similar to those in disgusted expression under illuminations from top or bottom.

5. Ratio Based Illumination Transfer

We discussed in previous sections that state-of-the-art automatic systems for expression recognition demonstrate great performance under ideal (uniform) lighting conditions. We also showed in the previous section that illumination influences the result of one these systems and becomes an impediment to the accurate evaluation of the degree of an expression. In this section we present our approach to reduce the effect of illumination and thus improve the performance of automatic expression recognition systems.

An overview of our approach is shown in Figure 10. The final objective is to bring a facial image, taken initially under an impairing-to-classifiers illumination condition, into a more uniformly lit illumination that will be an acceptable input to the

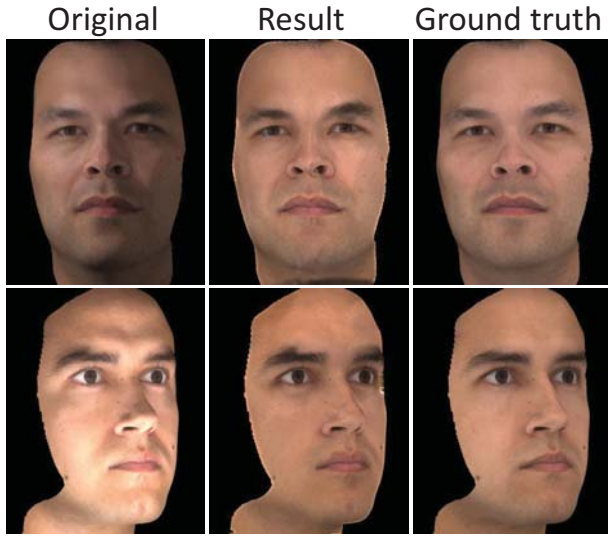


Figure 11: Factoring out known illumination from non frontal pose: A) Original image, with illumination that we want to “neutralize”, B) Output of our method, the target with desired illumination (this image was produced by our image based transfer of illumination method). C) Ground truth for comparison: the target subject illuminated with desired illumination condition (this image is a rendering from the 3D model).

automatic expression recognition systems.

5.1. Method overview

We used ratio images for re-lighting [8]. The overview of our system can be seen in Figure 10. The basic idea behind ratio images is that light can be aggregated or extracted simply by multiplying or dividing the pixel values of the images. So if we have the image of the same person in the same pose under two different illuminations, by dividing these two images we get the difference of the light between the two images [9].

Having a relightable 3D database, is extremely useful in this case, because we can use one of its subjects to match the geometry and pose of the target subject and extract the unwanted illumination from our subject using a ratio image. The ratio image has to be aligned with the target image and for that process we use both optical flow and sparse correspondence using AAM facial points [33]. One of the main differences of our approach and the other approaches that use ratio images for re-lighting, is that other researchers usually transform an image from a smooth illumination condition to a more complex one, whereas we are trying to do the opposite. Effectively, we want to go from a more complex illumination condition to a smoother one. It is also more challenging to perform ratio based light transfer to original images with expressive faces, as opposed to neutral faces.

Some results from our method are shown in Figure 11, where we demonstrate that we can also deal successfully with non frontal poses of faces (second row).

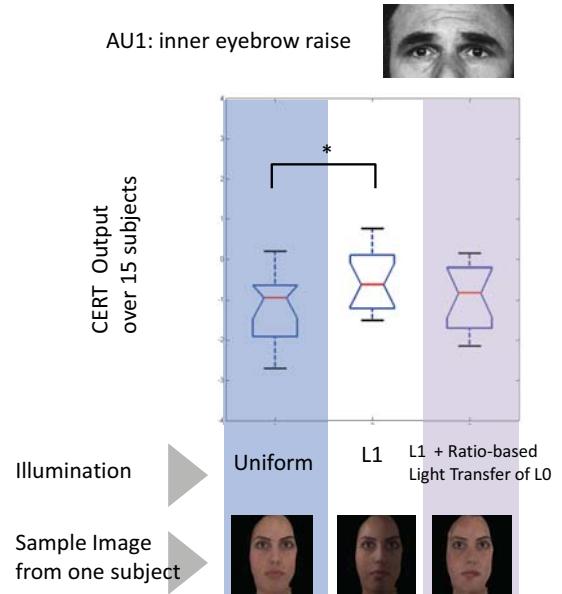


Figure 12: Distributions of AU1 scores for images illuminated with L0: uniform illumination, L1: light source brighter on the left side and images that L1 was factored out from to bring them to L0.

5.2. Results

We applied our method to images from the set used in the previous section, where we demonstrated that illumination affects AU scores. To show our approach, we proceed with the case of AU1, where light coming from the left side (L1) causes CERT output to change significantly as demonstrated in Figure 12, first two columns. We extracted that illumination (L1) from the neutral face of the subjects and changed their images to a more uniformly lit illumination condition (L0), which was used for the definition of the baseline. We evaluated the AU scores of the new “pseudo-L0” set of images using CERT, and the results of the new output are shown in Figure 12, last column.

L1 affects the output of CERT to the point that the distribution of AU1 outputs under L1 becomes statistically different from the one under L0. However, when we process the L1 images with our method of ratio based light transfer and we bring them under a uniform illumination, close to L0, the AU1 output distribution changes correctively toward the expected one, and the statistical difference becomes insignificant.

This is a very encouraging result, given our goal of light invariant AU classification.

6. Conclusions and Future Work

In this paper, we introduced a new database called ICT-3DRFE, which includes 3D models of 23 participants, under 15 expressions, with the highest resolution compared to the other 3D databases. It also includes photometric information which enables photorealistic rendering under any illumination condition. We showed how such properties can be employed in the design of experiments where illumination conditions are

modified to study the effect on systems of automatic expression recognition.

We presented a novel approach towards a light invariant expression recognition system. Using ratio images, we are able to factor out unwanted illumination and in some cases improve the output of AU automatic classification. Our current approach, however, requires that the facial image to be recognized be taken in known (although arbitrary) illumination conditions. For future work, we would like to remove this restriction by estimating the illumination environment directly from the image.

Since our observations generally agree with our intuitions, a goal for future work would also be to study the effect of illumination on human judgment.

7. ACKNOWLEDGMENTS

The authors would like to thank the following collaborators for their help on this paper: Ning Wang for the AU annotation of the database, Simon Lucey from CSIRO ICT Center (Australia) for the face tracking code used in this paper for sparse correspondence during image alignment, Cyrus Wilson and Jay Busch for help with processing of the acquired data. This material is based upon work supported by the U.S. Army Research, Development, and Engineering Command (RDECOM). The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

- [1] Braje, W.L., Kersten, D., Tarr, M.J. and Troje, N.F. Illumination effects in face recognition. (1998) *Psychobiology*, 26 (4), 371-380.
- [2] Adini, Y., Moses, Y., Ullman, S. (1995). Face recognition: the problem of compensating for changes in illumination direction (Report No. CS93-21). The Weizmann Institute of Science.
- [3] Chibrelushii, C.C., Bourel, F.: Facial expression recognition: A brief tutorial overview. In: Fisher, R. (ed.) *CVonline: On-Line Compendium of Computer Vision* (January 2003)
- [4] Li, H., Buenaposada, J.M.; Baumela, L.; , "Real-time facial expression recognition with illumination-corrected image sequences," *Automatic Face and Gesture Recognition*, 2008. FG '08. 8th IEEE International Conference on , vol., no., pp.1-6, 17-19 Sept. 2008
- [5] Kumar, R., Jones, M., Marks, T.K., "Morphable Reflectance Fields for enhancing face recognition," *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on , vol., no., pp.2606-2613, 13-18 June 2010
- [6] Toderici, G., Passalis, G., Zafeiriou, S., Tzimiropoulos, G., Petrou, M., Theoharis, T., Kakadiaris, I.A., "Bidirectional relighting for 3D-aided 2D face recognition," *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on , vol., no., pp.2721-2728, 13-18 June 2010
- [7] Bartlett, M., Littlewort, G., Tingfan Wu, Movellan, J., *Computer Expression Recognition Toolbox*, Univ. of California, San Diego, CA, *Automatic Face and Gesture Recognition*, 2008
- [8] Peers P., Tamura N., Matusik W., Debevec P., (2007). Post-production Facial Performance Relighting using Reflectance Transfer. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*, 26(3).
- [9] Riklin-Raviv T., Shashua A., (1999). The quotient image: class based recognition and synthesis under varying illumination conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 02, 262265.
- [10] Ekman, P., Friesen, W.V., Hager, J.C., *Facial Action Coding System (FACS): Manual*, Salt Lake City (USA): A Human Face, 2002.
- [11] Ekman, P., *Emotion in the Human Face*, Cambridge University Press, 1982
- [12] Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(12), 1424-1445 (2000)
- [13] Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S., "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.31, no.1, pp.39-58, Jan. 2009
- [14] Bartlett M.S., Littlewort G.C., Frank M.G., Lainscsek C., Fasel I., and Movellan J.R., (2006). Automatic recognition of facial actions in spontaneous expressions., *Journal of Multimedia*., 1(6) p. 22-35.
- [15] Lucey, S., Ashraf, A.B., Cohn, J.F., Investigating Spontaneous Facial Action Recognition through AAM Representations of the Face, *Face Recognition*, K. Delac, and M. Grgic, eds., pp. 275-286, I-Tech Education and Publishing, 2007.
- [16] Kanade, T., Tian, Y., Cohn, J.F., "Comprehensive Database for Facial Expression Analysis," fg, pp.46, Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), 2000
- [17] Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J. (1998), Coding facial expressions with Gabor wavelets, In 3rd International Conference on Automatic Face and Gesture Recognition.
- [18] Pantic M, Valstar M, Rademaker R and Maat L (2005), Web-based Database for Facial Expression Analysis, Proc. of IEEE Intl Conf. Multimedia and Expo (ICME05).
- [19] Sim, T., Baker, S., Bsat, M., "The CMU Pose, Illumination, and Expression (PIE) Database," *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, May, 2002.
- [20] Anitha C., Venkatesha M.K., Adiga B.S., A Survey on Facial Expression Databases, *International Journal of Engineering Science and Technology*, Vol. 2(10), 2010, 5158-5174
- [21] Yin L, Wei X, Sun Y, Wang J, Rosato M J (2006), A 3D Facial Expression Database For Facial Behavior Research, 7th International Conference on Automatic Face and Gesture Recognition (FGR06).
- [22] Yin L, Chen X, Sun Y, Worm T, Reale M (2008), A High-Resolution 3D Dynamic Facial Expression Database, The 8th International Conference on Automatic Face and Gesture Recognition (FGR08).
- [23] Langner O, Dotsch R, Bijlstra G, Wigboldus D H J, Hawk S T, and Van Knippenberg A (2010), Presentation and validation of the Radboud Faces Database, *Cognition and Emotion, Psychology Press*.
- [24] Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I. , "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010 IEEE Computer Society Conference on , vol., no., pp.94-101, 13-18 June 2010
- [25] Liu, Z., Shan, Y., Zhang, Z. 2001. Expressive expression mapping with ratio images. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques (SIGGRAPH '01)*. ACM, New York, NY, USA
- [26] Debevec P., Hawkins T., Tchou C., Duiker H.-P., Sarokin W., Sagar M., (2000). Acquiring the reflectance field of a human face. In *Proceedings of ACM SIGGRAPH, Computer Graphics Proceedings, Annual Conference Series*, 145156.
- [27] Zhen Wen; Zicheng Liu; Huang, T.S.; , "Face relighting with radiance environment maps," *Computer Vision and Pattern Recognition*, 2003. Proceedings. 2003 IEEE Computer Society Conference on , vol.2, no., pp. II-158-65 vol.2, 18-20 June 2003
- [28] Xiaowu Chen; Mengmeng Chen; Xin Jin; Qingping Zhao; , "Face illumination transfer through edge-preserving filters," *Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE Conference on , vol., no., pp.281-287, 20-25 June 2011
- [29] Yang Wang, Lei Zhang, Zicheng Liu, Gang Hua, Zhen Wen, Zhengyou Zhang, Dimitris Samaras. Face Re-Lighting from a Single Image under Arbitrary Unknown Lighting Conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 31, Number 11, Pages 1968-1984, November 2009
- [30] Ma W.-C., Hawkins T., Peers P., Chabert C., Weiss M., Debevec P., (2007). Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Spherical Gradient Illumination. *Proc. Eurographics Symposium on Rendering*, p. 183-194.
- [31] Wilson C.A., Ghosh A., Peers P., Chiang J., Busch J., Debevec P., (2010). Temporal Upsampling of Performance Geometry using Photometric Alignment. *ACM Transactions on Graphics*, 29(2).

- [32] Debevec P., (1998). Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. Proceedings of the 25th annual conference on Computer graphics and interactive techniques (SIGGRAPH '98), p. 189-198.
- [33] Lucey S., Wang Y., Saragih J., Cohn J.F., 2010. Non-rigid face tracking with enforced convexity and local appearance consistency constraint. *Image Vision Comput.* 28, 5 (May 2010), 781-789.