An Illumination Invariant Face Recognition Approach using Exemplar-based Synthesis Technique

Tele Tan, Thorsten Kühnapfel, Amelyn Wongso and Fee-Lee Lim

Department of Computing, Curtin University of Technology
GPO Box U1987, Perth, Western Australia 6845
{teletan, thorsten, wongsoa, limfl} @cs.curtin.edu.au

Abstract. This paper proposes a new method to solve the problem of face recognition under varying illumination conditions. We introduce an exemplar-based technique to decouple and subsequently recover the canonical face and the illumination functions from the intensity images. The canonical image is equivalent to the reflectance field of the face that is invariant to illumination. We subsequently use the canonical face to synthesize novel face appearances together with a set of lighting models. We then demonstrate the ability of the synthesis approach to improve the performance of the face recognition task.

1 Introduction

The need to further develop robust face recognition techniques to meet real world situations is still an open research challenge. It is widely stated that the two main contributions of poor recognition performances are that caused by variations in face pose and lighting. We will deal with the problem of illumination in this paper. Approaches addressing the illumination-related problems can be broadly classified into two categories; feature-based approach and exemplar- or appearance- based approach. Feature-based approaches aim to define a feature space that exhibits some broad invariance over the lighting variations. Examples of these are [1][10] which uses different image representations like 2D Gabor-like filters, first and second derivatives of the image, and the logarithmic transformation. Although these features may exhibit intensity immunity, none of these are found to be reliable to cope with significantly large variations in illumination changes [9][10].

Exemplar- or appearance- based approaches use a set of sample images taken of a class object (in this case a face) as a basis to compute an intermediary image. The intermediate image can then be used either directly as the probe image or be used to synthesize novel views of the face under different lighting conditions [11]. For example, [2] reported a method to compute the Quotient Image from a small sample of bootstrap images representing a minimum of two class objects. The illumination invariant signature of the Quotient Image enables an analytic generation of the novel image space with varying illumination. However, this technique is highly dependent on the types of bootstrap images used which has the undesirable effect of generating diversely looking Quotient Images even from the same person. Sim and Kanade [3] use a statistical shape-from-shading model to estimate the 3D face shape from a single image. The 3D recovery model is based on the symmetric shape-from-shading algorithm proposed by [4]. They used the 3D face model to synthesize novel faces under new illumination conditions using computer graphics techniques. The approach produce high recognition rate on the illumination subset of the CMU PIE database [5]. However, it was not evident how their synthesis technique can cope with extreme illumination conditions [3]. Debevec in [6] presented a method to acquire the reflectance field of a human face and use these measurements to render the face under arbitrary changes in lighting and viewpoint. However, the need to generate a large sample of images using the light stage is unfeasible for most face recognition systems. A parameter-free method of estimating the bi-directional reflectance distribution of a subject's skin was proposed by Hancock et al in [12]. They estimated the radiance function by exploiting differential geometry and making use of the Gauss map from the surface onto a unit sphere. They demonstrated the approach by applying it to the re-rendering of faces with different skin reflectance models.

As in [2] and [11], we address the problem of class-based image synthesis and recognition with varying illumination conditions. We define an ideal class as a collection of 3D objects that have the same shape but different albedo functions. For recognition purposes, we can broadly assume all human faces to belong to a certain class structure. This assumption was similarly adopted by Shashua [2] and Mariani [11]. Our approach is based on the dual recovery of the canonical face model and lighting models given a set of images taken with varying lighting conditions and from a minimum of two distinct subjects within the class. The canonical image is equivalent to the reflectance field of the face that is invariant to illumination. The lighting model is the image representation of the ambient lighting independent of the face input. We will first formulate the problem with an over-determined set of equations and propose a method in solving them over every pixel location in the image. We will demonstrate the quality of the recovered canonical face for generating novel appearances using both subjective and objective measures.

2 The Simplified Illumination Function

The intensity of reflected light at a point on a surface is the integral over the hemisphere above the surface of a light function *L* times a reflectance function *R*. The pixel equation at point (x,y,z) can be expressed as

$$I(x,y,z) = \int_t \int_\lambda \int_\theta \int_\phi L(t,x,y,z,\theta,\phi,\lambda) R(t,\theta,\phi,\lambda)\, \mathrm{d}\theta\, \mathrm{d}\phi\, \mathrm{d}\lambda\, \mathrm{d}t \qquad \textbf{(1)}$$

where

x,y,z = the co-ordinate of the point on the surface
ϕ and θ = azimuth and yaw angle from the z axis respectively
t and λ = time and wavelength of the light source

This equation is computationally too complex to solve in many real-time applications. We need to make further simplification of the equation without significantly affecting the goal of our work. Firstly, z, t and λ can be eliminated because we are dealing with the projected intensity value of a 3D point onto a still frame image with grey scale intensity. Additionally, if one considers fixing the relative location of the camera and the light source, θ and ϕ both become constants and the reflectance function collapses to point (x, y) in the image plane. Therefore, the first-order approximation of equation (1) for a digital image $I(x,y)$ can be further written as:

$$I(x,y) \approx R(x,y)\, L(x,y) \qquad \mathbf{(2)}$$

where $R(x,y)$ is the reflectance and $L(x,y)$ is the illumination at each image sample point (x,y). Our approach is to use exemplar images taken over different fixed lighting directions to recover both the reflectance model and illumination source. It is not the scope of this work to accurately model the skin reflectance property according to specificity like the melanin content of the skin, skin hemoglobin concentration and level of perspirations. These are important for visually accurate skin rendering application but less so for face recognition.

3 The Approach

In our case, only the measured intensity images are available. Therefore, there are twice as many unknown data (RHS) as there are known data (LHS) making equation (2) ill-posed. The reflectance surface essentially comprises the combination of the reflectance property associated with the pigmentation of the skin, mouth, eyes and artifacts like facial hair. We define the reflectance model as the canonical face and represent it as a grey level intensity image. We will discuss in this section an approach that we propose to recover the canonical and illumination information from a set of intensity images $I_{ij}(x,y) \approx R_j(x,y)\, L_i(x,y)$, where i and j are indices to the collection of bootstrap[1] faces and illumination directions respectively.

3.1 Defining and Solving the Systems of Equations

As explained in the previous section, equation (2) has more unknown terms than known. In order to make the equation solvable in a least square sense, we need to introduce additional measurements thus making the system of equations ov*er determined*. We further note that the bootstrap image, $I_{ij}(x,y)$ has two variable components. They are the reflectance component which is unique to the individual person and the illumination model which is dependent on the lighting source and direction. Suppose we have M distinct persons which we use in the bootstrap collection (i.e. R_j, $j = 1, \ldots,$ M) and N spatially distributed illumination sources whose direction with respect to

[1] The bootstrap collection comprises of face sample images taken of various person over multiple illumination directions, the relative location of which are fixed.

the person is fixed at all instances (i.e. L_i, $i = 1, \ldots, N$), we will have therefore a total of MxN known terms and M+N unknown terms. These *over-determined* systems of equations can be solved by selecting any values of M and N that are greater than 1. For example, if we use M persons from the bootstrap collection, and collect N images for each person by varying the illumination, we get the following system of equations;

$$
\begin{gathered}
I_{i1}(x,y) \approx R_1(x,y)\, L_i(x,y) \\
\vdots \\
I_{iM}(x,y) \approx R_M(x,y)\, L_i(x,y)
\end{gathered}
\tag{3}
$$

where $i = 1,\ldots,N$. The terms on the left hand side of these equations are the bootstrap images from the M number of persons. If the illuminations used to generate these bootstrap images are the same, the illumination models, L_i will be common for every person as is reflected in equation (3).

Numerous non-linear minimization algorithms exist and are usually problem dependent [7]. We chose to use the Levenberg-Marquardt non-linear least square fitting algorithm [8] as it is fast and suited to problems of high dimensionality. The solver takes as input the set of equations shown in (3) to minimize, a Jacobian matrix of derivatives, a set of known data (i.e. I_{ij}) and seed values for the unknowns. We chose to set the seed value to 128 since there are 256 possible grey values for both the reflectance and illumination models. The internal functions of the solver are iterated until the change in computed values falls below a threshold. At this point the algorithm is said to have converged, and the current computed values for the unknown data are taken as the solution. The algorithm is extremely fast and can recover the unknown values (for most practical values of M and N) in near real-time.

3.2 Appearance Synthesis

Once the canonical face and the illumination model are recovered, we can proceed to perform the appearance synthesis using the following principles:

1. New illumination models can be generated by the combination of the subset of the recovered illumination models.
2. Novel appearance views for each person can be generated by the combination of an expanded set of illumination models to closely match the actual illumination conditions.

It is not economical and computationally feasible to store specific illumination models for specific faces. To make this approach viable, we need to define a set of generic illumination models that is suitable for a broad cross section of people with different skin types and bone structures. We estimate this generic illumination models using the weighted average models gathered from a genre of subjects.

4 Experiments and Results

4.1 The Database

For our experiments, we make use of the illumination subset of the CMU PIE database [5]. It comprises 63 people taken under 21 different illumination directions (with all ambient lights switched off) in a controlled environment. All the color images are first transformed to grey-level, pre-processed and the faces cropped. The final size for all images is 110 x 90 pixels.

4.2 Canonical Face Recovery

We use equation (3) to recover the canonical faces with different values of M and N and a subset of them are shown in Fig. 1. In order to measure the quality of the recovered canonical face, we define a set of measures that describes the properties of an acceptable canonical face. The measures are; (1) Elimination of lighting effects like specular reflections and cast shadows. (2) Preservation of the visual distinctiveness of the underlying face. (3) Well-balanced intensity distribution. Based on these measures, we can see that in general the recovery of the canonical faces for different values of M and N are very good. This is a significant improvement over the Quotient Image reported in [2]. To further support the significance of the recovered canonical face, we will next describe a face recognition experiment that will quantitatively show the ability of our approach to deal with illumination variation problem.

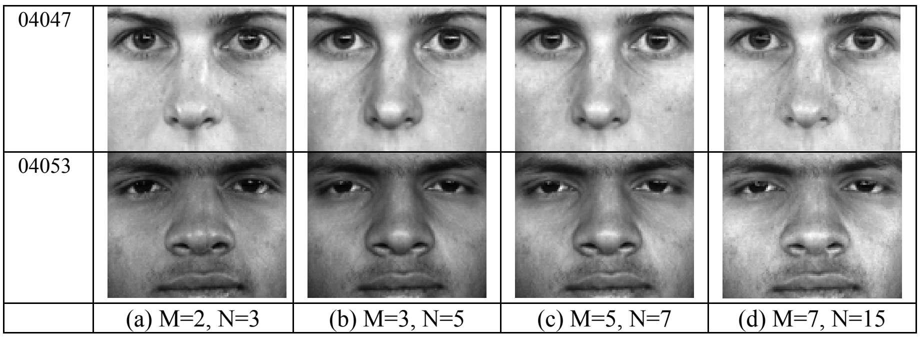


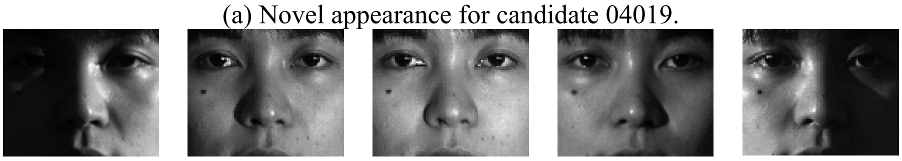
Fig. 1. Canonical faces generated for candidate samples 04047 and 04053 using (a) M=2,N=3, (b) M=3,N=5, (c) M=5,N=7 and (d) M=7,N=15.

4.3 Face Appearance Synthesis

For each recovered canonical face, the corresponding set of 21 illumination models can then be computed. We further estimated the generic illumination models as defined in Section 3.2 by using 10 candidate samples from the CMU PIE database.

We then use these generic illumination models and the canonical faces from the remaining samples to generate novel appearance faces. Fig. 2a shows the synthesized views of a subject generated using 5 different illumination models. The corresponding images captured by the actual illuminations are shown in Fig. 2b.

(a) Novel appearance for candidate 04019.



(b) Corresponding actual appearance for candidate 04019 from CMU PIE database.

Fig. 2. Novel appearance synthesis results using a subset of the generic illumination models and its comparison with the actual appearance. (derived from camera locations f02, f10, f07, f14 and f17 as defined in the CMU database).

4.4 Recognition Experiment

To demonstrate the feasibility of the face appearance synthesis for recognition, we implement a simple classifier based on template matching. This is equivalent to the nearest neighbor classifier reported by Sim and Kanade [3]. We use only frontal pose faces throughout the experiment. The generic illumination models used here is the same as in Section 4.3. To maintain unbiased recognition outcome, the test samples used for recognition does not come from any of the samples used to produce the generic illumination models. There are 21 persons in the test samples. From each person we compute the canonical representation and use it to synthesize 21 appearances of the person under different lighting conditions. These images collectively form the registry representation of the person in the database. We use actual illumination samples of the PIE database as the test images. There are a total of 441 (i.e. 21x21) test sample images. We construct different registry databases for different combination of M (number of person) and N (number of lighting) values. We then perform the face recognition experiments on the test samples over the different registries. Fig. 3 shows the summary of recognition rate for different values of M and N. We observe several important behaviors. They are:

1. For a fixed value of M, the recognition rate increases monotonically when N increases.
2. However when M increases, N has to consequentially increase for the canonical face to be recovered with reasonable quality. The minimum (M,N) pair needed to establish good recognition rates are (2,3), (4,5), (6,7), (8,9) and (10,11).
3. The recognition rate for N=2 is very poor for all values of M.

4. The range of recognition rates for different values of M and N (ex N=2) are between 83.0% and 90.7%.

As can be seen, the results obtained here is significantly better than [3] which reported an accuracy of 39% with the nearest neighbor classifier on a similar dataset. The general trend of the recognition rates which flatten off as N increases for all values of M suggest a wide perimeter for the choices of these values. However, from the computation, data acquisition and hardware standpoint, it would be effective to keep the M and N values small, without negatively impacting the recognition rate.

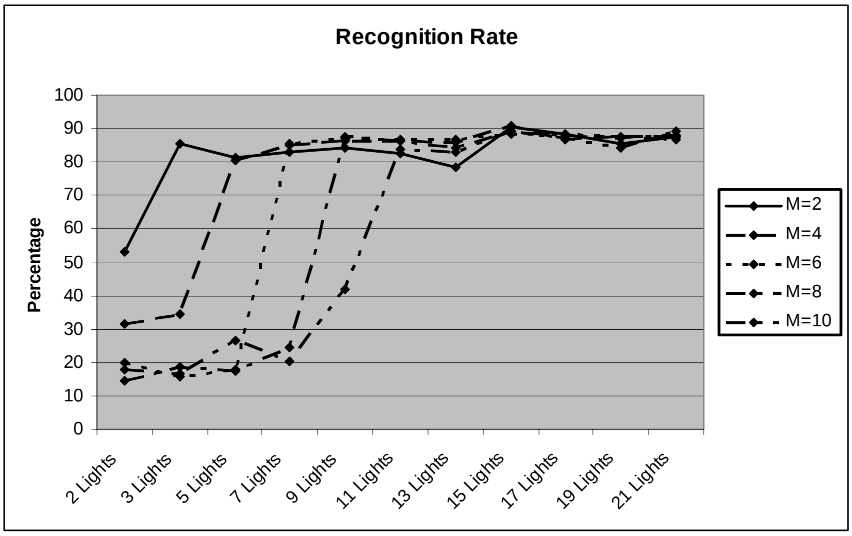


Fig. 3. Recognition rates (in percentage) by varying the values of M and N.

5 Discussions

The results obtained using the canonical face recovery algorithm is very encouraging. Besides using the images captured by the lighting module as described here, we can explore shape-from-shading techniques to recover the 3D shape of the face as was done in [4]. The range information together with the canonical face are essential for improving the illumination rendering quality and to deal with pose invariant recognition. Although the illumination models recovered using the CMU PIE database generates 21 different variations they are still inadequate as some crucial lighting directions (i.e. especially those coming from the top) are missing. We will next consider using computer graphics tools to develop a virtual light stage that can produce light rendering from any arbitrary lighting directions. These can then be used to extract finer illumination models. We are also in the process of building a scale-down lighting platform to validate our approach further.

6 Conclusion

We have developed an exemplar-based approach aim at recovering the canonical face of a person. The canonical face can either be use as a probe face for recognition or use as a base image to generate novel appearance models under new illumination conditions. We have shown subjectively that the canonical faces recovered with this approach are very stable and not heavily dependent on the types and numbers of the bootstrap images. The strength of the view synthesis algorithm based on the canonical face was further demonstrated by a series of face recognition tests using the CMU PIE images.

Acknowledgement

We would like to thank the Robotics Institute, Carnegie Mellon University for providing the PIE images

References

1. Y. Adini and S. Ullman, “Face Recognition: the Problem of Compensating for Changes in Illumination Direction”, in Proc. of IEEE Trans. on PAMI. Vol. 19, No. 7, 1997, pp. 721-732.
2. T. Riklin-Raviv and A. Shashua, “The Quotient Image: Class based Re-rendering and Recognition with Varying Illuminations”, in Proc. of IEEE Trans. on PAMI. Vol. 23, No. 2, 2001, pp. 129-139.
3. T. Sim and T. Kanade, “Combining Models and Exemplars for Face Recognition: an Illumination Example”, in Proc. of Workshop on Models versus Exemplars in Computer Vision, Dec 2001.
4. Zhao and Chellappa, “Robust Face Recognition using Symmetric Shape-from-Shading”, Technical Report CARTR-919, Centre for Automation Research, University of Maryland, College Park, MD, 1999.
5. T. Sim, S. Baker and M. Bsat, “The CMU Pose, Illumination and Expression Database”, in Proc. of IEEE Trans. on PAMI, Vol. 25, No. 12, 2003, pp. 1615-1618.
6. P. Debevec, T. Hawkins, C. Tchou, W. Sarokin, and M. Sagar, ”Acquiring the Reflectance Field of a Human Face”, in Proc. of SIGGRAPH 2000, pp. 145-156.
7. A. Yeredor, “The Extended Least Squares Criterion: Minimisation Algorithms and Applications”, IEEE Trans. on Signal Processing, Vol. 49, No. 1, Jan 2000, pp. 74-86.
8. J. More, “The Levenberg-Marquardt Algorithm: Implementation and Theory”, in G. Watson, Ed, Lecture Notes in Mathematics, Springer Verlag, 1978, pp. 105-116.
9. B. Manjunath, R. Chellappa and C. D. Malsburg, “A feature based approach to face recognition”, in Proc. of IEEE Computer Society. Confr. On Computer Vision and Pattern Recognition, 1992, pp. 373-378.
10. P. Yang, S. Shan, W. Gao, S. Li and D. Zhang, “Face recognition using Ada-boosted Gabor features”, FG 2004, pp. 356-360.
11. R. Mariani, “A Face Location and Recognition System Based on Tangent Distance”, Multimodal Interface for Human-Machine Communication, Vol. 48. Ed. PC Yuen, YY Tang and PSP Wang, World Scientific, pp 3-31.
12. W.A.P Smith, A. Robles-Kelly and E.R. Hancock, “Skin Reflectance Modelling for Face Recognition”, in Proc. of the Int’l Confr. on Pattern Recognition, 2004, pp. 210-213.