ONE COLORED IMAGE BASED 2.5D HUMAN FACE RECONSTRUCTION

Peng Liu, W.L. Woo and S.S. Dlay
School of Electrical, Electronic and Computer Engineering
Newcastle University, Newcastle upon Tyne
England, United Kingdom
Email: peng.liu2@ncl.ac.uk, w.l.woo@ncl.ac.uk, s.s.dlay@ncl.ac.uk

ABSTRACT

2.5D human face is illumination invariant which has a great advantage in face recognition. However, the existing method are linear based and capturing a 2.5D human face involves multi images from the same view point which is impractical in reality. This paper introduces a new nonlinear method for normal Surveillance camera to capture a 2.5D human face data. Only a single image is needed during capturing process by using RGB light sources. The illumination is separated from 2D images by applying ICA (Independent component analysis) method. A nonlinear statistical reflection model is developed through the nonlinear ICA algorithm to compensate nonlinear distortions during image capturing process. The proposed algorithm has achieved excellent features in separating the illumination which yielded very high accuracy of 2.5D human face data recovery.

1. INTRODUCTION

2.5D based face recognition was developed to have excellent characteristics which are invariant to illumination effects. Comparing with traditional 2D based image processing which relates only to gray level information. 2.5D images contain the extra dimension information which is the shape information that is independent to any illumination conditions. Different illumination exerts on the same shape will cause huge variations in the gray level. Gray level images contain part of the shape information and part of the illumination condition. The useful information for recognition is only contained in the shape. Therefore, separate the two and extract the 2.5D shape information is necessary.

This paper is based on photometric stereo approach which was introduced by Woodham [1]. It estimates the local surface orientation by using several images of the same surface taken from the same viewpoint but under different illuminations conditions, which have received wide attention [2-10]. The main limitation of the classical photometric stereo approach is that the light source positions must be accurately known. Hayakawa [11] proposed a method without the pre knowledge of light source direction and intensity. Singular-value decomposition (SVD) was first applied to factorize image data matrix of three different illuminations into surface reflectance matrix and light source matrix based on the Lambertian model. However, it still needs one of the two added constrains for finding the linear transformation between the surface reflectance matrix and the light source matrix. Georghiades [15] extend Hayakawa's method. However, all of the above methods are based on the assumption that human skin surface is a lambertian surface [1], which suggests light is reflected equally in all directions. This is not the case in reality since nonlinear deformation exists and this must be taken into account.

All of the above methods require at least three to seven images taken from the same view angle at different times with various light sources, which is very impractical in reality. For most surveillance cameras, only one picture can be acquired at each time under a certain illumination condition. In this paper, we introduced a single image based reconstruction method which by using RGB light sources illumination condition for each camera shot. A nonlinear statistical based reflection model is also proposed which handles the diffuse components and the specular components during the reflection process. An unsupervised learning adaptation algorithm is used to tune up the proportion of hybrid automatically based on image intensities. The technique of the post-nonlinear Independent Components Analysis (ICA) [12-13] is used to solve the surface normal on each point of an image. The goal of post-nonlinear ICA is to nonlinearly transform the data such that the transformed variables are as statistically independent from each other as possible. Finally, the 2.5-D surface model is reconstructed from the surface normal based on the enforcing integrability method [14].

2. NONLINEAR REFLECTION MODEL

2.1 Lambertian Reflection Model

Reflection of a surface can be generally classified as diffuse reflection and specular reflection. Diffuse reflection is the reflection of light from an uneven or granular surface such that an incident ray is seemingly reflected at a number of angles. Figure 2 illustrates the Lambertian surface.

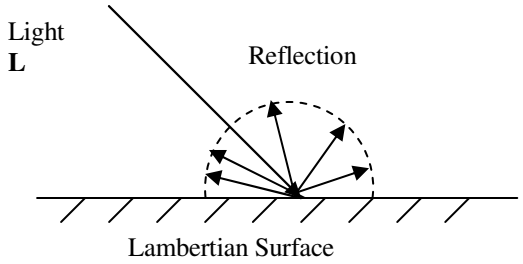


Figure 1. Lambertian Reflection

Lambertian reflection is calculated by taking the dot product of the surface's normalized normal vector and a normalized vector pointing from the surface to the light source, this can be expressed as follow

$$\mathbf{R}(\mathbf{n}(x,y),\alpha(x,y),\mathbf{s}) = L\alpha(x,y)\mathbf{s}^T\mathbf{n}(x,y), \forall x,y \in D \quad (1)$$

In (1), **R** (.) denotes the diffuse component intensity, **n**(x, y) is a 3×1 unit vector indicating the surface normal direction $\alpha(x,y)$ is the diffuse Albedo on position (x, y) of a surface, **s** is a 3×1 vector indicating the light source direction and L is the light strength. However, in reality, the reflection is actually a nonlinear combination of diffuse reflection and specular reflection.

2.2 Nonlinear Model

The nonlinear model in this paper is Non-Lambertian Model which is illustrated by the following equation:

$$\begin{aligned}&\mathbf{R}_{\text{nonlinear}}(\mathbf{n}(x,y),\sigma(x,y),\gamma(x,y))\\&= L\cdot\gamma(x,y)\cdot Lap((\cos^{-1}(\mathbf{a}^T\cdot\mathbf{n}(x,y)),\theta)\end{aligned} \quad (2)$$

In this equation, $\mathbf{R}_{\text{nonlinear}}$ denotes the nonlinear reflectance intensity, **a** is a 3×1 vector which represents the normalised bisector of the light and view direction. **n**(x, y) is the surface normal of the target 3D objects. L is the light strength and $\gamma(x,y)$ is the composite Albedo (the reflection rate) on position (x, y) and Lap indicated the Laplacian distribution. And θ is the parameter of Laplacian pdf, by adjusting which provide relative post nonlinear approximation of the pixel wise reflection.

3. POST NONLINEAR ICA ESTIMATION

Equation (2) can be understood as a linear mixing process embedded inside a nonlinear function. The linear step can be formulated as:

$$\mathbf{m}(x,y) = \mathbf{a}^T\cdot\mathbf{n}(x,y) \quad (3)$$

Equation (3) gives the linear part which is simply the surface normal **n**(x, y) mixed linearly with vector $\mathbf{a}^T$. For three images input, this mixing relationship can be denoted as:

$$\mathbf{m} = \mathbf{A}\cdot\mathbf{n} \quad (4)$$

The mixing matrix **A** linearly mixed surface normal vectors into a new linear combination. This is the reason why ICA is employed to separate the linear mixing problem.

The nonlinear part of the equation is the rest of the equation that generated the intensity images. We can write the nonlinear equation in relationship with the linear mixing variable "m" as following expression:

$$\mathbf{R} = L\cdot\gamma(x,y)\cdot Lap((\cos^{-1}(\mathbf{m})),\theta) \quad (5)$$

Let us define this nonlinear function as $f(\cdot)$ and surface normal **n** is the source signal to be estimated, which can be written as **s**, so we can formulate this simply as

$$r(t) = L\gamma(t)f_i\left(\sum_{j=1}^{3}a_{ij}s_j(t)\right), i = 1,2,...N \quad (6)$$

Equation (6) is the post nonlinear estimation problem which can be solved by first finding a nonlinear function $f^{-1}(\cdot)$ and applying ICA algorithm to separate the source signal **n**(x, y).

$$g(\theta,x(t))) = f^{-1}(\cdot) = Lap^{-1}\left(\frac{r(t)}{L\gamma(t)},\theta\right) \quad (7)$$

The estimation method used is based on maximum likelihood which is driven by the natural gradient method to achieve stable and fast converge. The log likelihood is given by

$$\begin{aligned}&LogL(\mathbf{B},\theta) =\\&\sum_{t=1}^{T}\left\{\sum_{i=1}^{n}\log\left|g'(\theta_i,x_i)\right| + \sum_{j=1}^{n}\log\left[p_j(b_j^Tg(\theta,x(t)))\right] + \log\left|\det\mathbf{B}\right|\right\}\end{aligned} \quad (8)$$

Where **B** is the de-mixing matrix, which is the inverse of matrix **A.** The maximum likelihood [12] method to solving the above post nonlinear problem is consists of two main steps: (i) linear updating step which updated the mixing matrix and (ii) nonlinear estimation which updates the post nonlinear model. Assume there are M data points or pixels on the images, the gradient of log likelihood function with respect to de-mixing matrix B can be written as:

$$\frac{\partial LogL(\mathbf{B},\theta)}{\partial\mathbf{B}} = (\mathbf{B}^T)^{-1} + E\left\{\sum_{j=1}^{n}h_j(a^Tf^{-1}(\theta,x(t))f^{-1}(\theta,x(t))^T\right\} \quad (9)$$

Therefore the updating rule for **B** is

$$\Delta\mathbf{B} \propto E\left\{\sum_{j=1}^{n}h_j(\mathbf{b}_j^Tg(\theta,x(t)))g(\theta,x(t))^T\right\} + (\mathbf{B}^T)^{-1} \quad (10)$$

$$\mathbf{B}^{k+1} = \mathbf{B}^k + \lambda_{\mathbf{B}}\Delta\mathbf{B} \quad (11)$$

where h_j is the entropy of the source signal, $\lambda_{\mathbf{B}}$ is the step size of the updating process. The bigger value of $\lambda_{\mathbf{B}}$ lead to a fast convergence but less accuracy, the smaller size of $\lambda_{\mathbf{B}}$ provides an accurate result but slow computation. In out program we choose $\lambda_{\mathbf{B}} = 0.001$ for the compromise between the two. The second step is to estimate coefficients

from post nonlinear model. The gradient of log likelihood with respect to coefficient θ can be written as:

$$\frac{\partial L(\mathbf{B},\theta)}{\partial\theta} = E\left\{\frac{\partial \log\left|f_i^{-1}(\theta_i, x_i)\right|}{\partial\theta_i}\right\} + E\left\{\left[\sum_{j=1}^{n} h_j(\mathbf{b}_j^T g(\theta, x(t)))b_{ji}\right]\frac{\partial g_i(\theta_i, x_i)}{\partial\theta_i}\right\} \quad (12)$$

$$\Delta\theta \propto E\left\{\frac{\partial \log\left|f_i^{-1}(\theta_i, x_i)\right|}{\partial\theta_i}\right\} + E\left\{\left[\sum_{j=1}^{n} h_j(\mathbf{b}_j^T g(\theta, x(t)))b_{ji}\right]\frac{\partial g_i(\theta_i, x_i)}{\partial\theta_i}\right\} \quad (13)$$

$$\theta_i^{k+1} = \theta_i^k + \lambda_\theta \Delta\theta_i \quad (14)$$

4. 2.5D RECONSTRUCTION

After calculating the surface normal vectors, the final step is to integrate the normal vector to the 2.5D shape. Due to the existence of system noise which make the surface non integrable, we can not simply integrate surface to acquire the 2.5D reconstruction. The method applied is to enforce integrability by using a classical method [14] for the convenience of comparison with [15]. The source signal estimated in the last section is the surface normal signal which tells which direction the surface normal vector points. The relationship of surface normal vector and the surface depth information can be expressed as below:

$$\mathbf{n}(x,y) = \frac{\begin{bmatrix} -p(x,y) & -q(x,y) & 1\end{bmatrix}^T}{\sqrt{p^2(x,y)+q^2(x,y)+1}} \quad (15)$$

where $q(x,y) = \partial z(x,y)/\partial x$ and $p(x,y) = \partial z(x,y)/\partial y$ are the surface gradients. And z(x, y) is the depth at each (x, y) pixel location, therefore the aim of 2.5 D is to recover the depth information z(x, y) on each location (x, y) of the 2D image. By minimizing the distance of partial derivatives of the estimated non integrable surface $\hat{\mathbf{z}}$ from the partial derivatives of integrable surface $\tilde{\mathbf{z}}$, we finally can achieve integrablity [14]. This distance can be expressed as:

$$d\left\{\left(\hat{z}_x, \hat{z}_y\right), \left(\tilde{z}_x, \tilde{z}_y\right)\right\} = \iint \left|\hat{z}_x - \tilde{z}_x\right|^2 + \left|\hat{z}_y - \tilde{z}_y\right|^2 dxdy \quad (16)$$

After acquiring the surface height data, the last step is to map the texture and reconstruct the 2.5D human face.

5. SYSTEM DESIGN

In the experiment setting up, three RGB (Red, Blue and Green) light sources are required. Light sources shall be placed three meter away from the subject and with adjustable of orientation and height. The differences of the indent angle between each light source shall be around 10 degree to avoid shading. The image acquiring device is a digital camera which is mounted on a standard Tri Pod which is 1.5 meters high with adjustable orientation and height. Subject should be seated three meter away from the camera. This is shown in Figure2

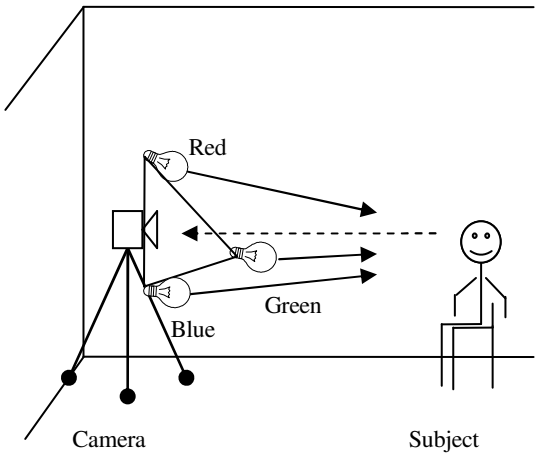


Figure 2. System set up

In this experiment, RGB light are chosen due to the nature independency of the prime RGB colours which can be considered as ideal illumination sources. One digital camera is place in front of the subject. During image capturing process, only one image shall be captured with the mixed RGB light sources from different directions illuminating on the subject at the same time.

6. EXPERIMENT RESULTS

In our experiment, three light sources (Red, Green and Blue - 60Watt each) are applied to the subject from three meter away and lying not on the same plane to avoid singularity of the de mixing matrix. The angle between each light source is bigger than five degree. The light source direction is unknown which will be estimated automatically through proposed algorithm. The acquired RGB input image is shown as following figure:

Figure 3. Single RGB image input

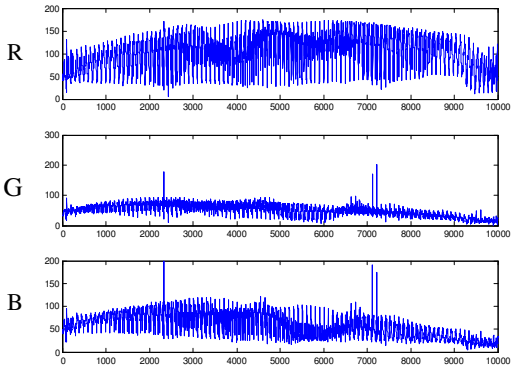


Figure 4. Extracted signal from R G B channels

In Figure4, three signals are extracted from the RGB channels of the input image during pre process stage. R, G and B represent three signals from channels respectively.

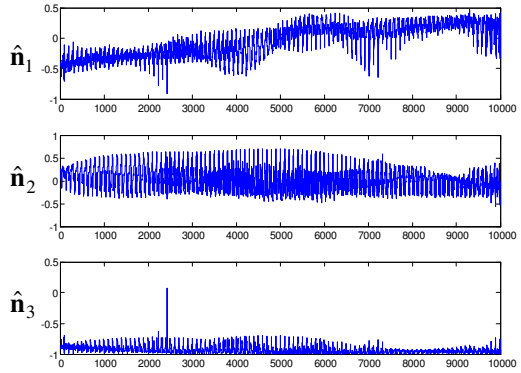


Figure 5. Recovered surface normal signal

Figure 5 is the recovered surface normal signal. $\hat{\mathbf{n}}_1$, $\hat{\mathbf{n}}_2$ and $\hat{\mathbf{n}}_3$ are the represent three surface normal signals in the vector form. Each signal indicated the variation of surface normal along each coordinate direction.

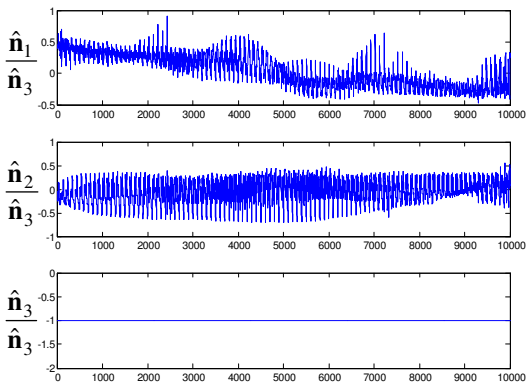


Figure 6. Normalized surface normal

To reconstruct 2.5D depth information, the third signal $\hat{\mathbf{n}}_3$ is normalized to unity as it is shown in Figure 6.

By enforcing Integrability on the normalised surface normal, we can acquire the pure shape information as shown in Figure 7

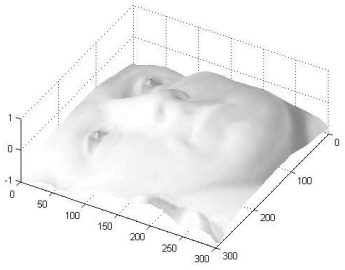
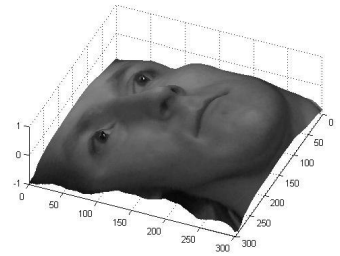
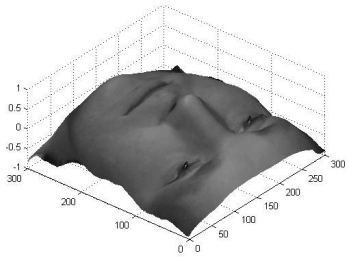


Figure 7. Reconstructed 2.5D shape information without albedo

After mapping texture of each pixel to the recovered shape in figure 7 we can finally achieve 2.5D human face reconstruction with texture attached.



(a)



(b)

Figure 8. Final 2.5D Reconstruction with texture attached

Figure 8 indicated the final reconstructed 2.5D human face with texture rendered. It is more realistic and full of detailed facial ridge and features.

7. COMPARISON AND ANALYSIS

The comparison of proposed algorithm is made by reconstructing synthetic 2.5D sphere object which can be formu-

lated as $z(x,y)=\left|\sqrt{r^2-x^2-y^2}\right|$ where r =100 and $0<x,y<100$. Illumination condition is formulated as three unknown light source directions randomly composed five light source settings (Ls1, Ls2, Ls3, Ls4, Ls5). Our proposed algorithm, Georghiades's method [15] and Hayakawa's algorithm [11] are subject to the test. The comparison result can be made by calculating the mean absolute error (MAE) which is shown in the following figure:

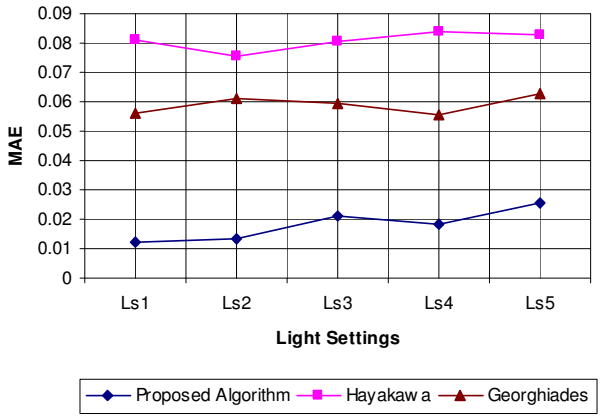


Figure 9. Mean Absolute Error Comparison

It is clearly shown that the MAE of proposed algorithm is around 0.018 which is the lowest, Georghiades's method provide the second lowest MAE, around 0.05. Hayakawa's method is purely a SVD based method therefore ranked the last with MAE around 0.081. Our algorithm has a highest accuracy which far outperformed than other two methods. Our final result only contains 1.8 percent Mean Absolute Error while Georghiades's result contains 5.9 percent and Hayakawa's result contains 8.1 percent Mean Absolute Error. This is due to the post nonlinear model counter balanced the nonlinear distortion.

8. CONCLUSION

This paper proposed a new method for 2.5D human face reconstruction through a single image subjected to RGB lights. A statistical nonlinear model is developed to compensate for the nonlinear distortion during the image capturing process. Through the proposed post-nonlinear ICA algorithm, it has effectively removed the nonlinear illumination factors on 2D images and produced a 2.5D human face depth recovery. Comparing with linear based method, our proposed method gives a substantial higher accuracy. Linear based traditional method assumes the human skin is a lambertian reflection surface while in reality the characteristic of human skin is far from linear which follows random statistical distribution as our model described. Linear based method can only achieve signal recovery along the maximum variance direction, while post-nonlinear ICA method compensates for the nonlinear distortion and recovers the signal with maximum statistical independence.

REFERENCES

[1] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," J. Opt. Eng., vol. 19, no. 1, 1980.

[2] C. Cho and H. Minanitani, "A New Photometric Method Using 3 Point Light Sources," IEICE Trans. Inf. and Syst. vol.E76-D, No. 8, pp. 898-904, August 1993.

[3] E. Angelopoulou and J. P. Williams, "Photometric Surface Analysis in a Triluminal Environment," IEEE International Conf. on Computer Vision, 1999.

[4] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Determining Shape and Reflectance of Hybrid Surfaces by Photometric Sampling," IEEE Trans. Robot. Autom., vol. 6, pp. 418-431, August 1990.

[5] Y. Iwahori, R. Woodham, and A. Bagheri, "Principal Components Analysis and Neural Network Implementation of Photometric Stereo," Proc. of the Workshop on Physics-Based Modelling in Computer Vision, pp. 117-125, June 1995.

[6] F. Solomon and K. Ikeuchi, "Extracting the Shape and Roughness of Specular Lobe Objects Using Four Light Photometric Stereo," IEEE Conf. on CVPR, pp. 466-471, June1992.

[7] O. Drbohlav and A. Leonardis, "Detecting Shadows and Specularities by Moving Light," in Proc. of Computer Vision Workshop, Slovenia, pp. 60-74, 1998.

[8] E. N. Coleman and R. Jain, "Obtaining 3Dimensional Shape of Textured and Specular Surfaces Using Four-Source Photometry," Proc. of Computer Vision, Graphics, and Image, vol. 18, pp. 309- 328, 1982.

[9] H. Rushmeier, G. Taubin, and A. Gueziec, "Applying Shape from Lighting Variation to Bump Map Capture," In Eurographics Rendering Techniques'97, pp. 35-44, St. Etienne, France, June 1997

[10] H. D. Tagare and R. J. P. deFigueiredo, "Simultaneous Estimation of Shape and Reflectance Maps from Photometric Stereo," IEEE International Conf. on Computer Vision, pp. 340-343, 1990.

[11] K. Hayakawa, "Photometric stereo under a light source with arbitrary motion," J. Opt. Soc. Amer.: A, vol. 11, no. 11, 1994.

[12] Jingyi Zhang, W.L. Woo, and S.S. Dlay, "Blind source separation of post-nonlinear convolutive mixture," IEEE Transactions on Audio, Speech and Language Processing, vol. 15, no. 8, pp. 2311-2330, 2007.

[13] W.L. Woo and S.S. Dlay, "Neural network approach to blind signal separation of mono-nonlinearly mixed signals", IEEE Trans. on Circuits and System I, vol. 52, no. 2, pp. 1236-1247, June 2005.

[14] R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 4, pp. 439–451, Jul. 1988.

[15] S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 643–660, Jun. 2000.