# Photometric Stereo with Relit Images from a Single Frontal Face Image

Junhyeop Lee[1], Kyungjae Lee[2] and Sanyoon Lee[*]
[1, 2, *] Department of Computer Engineering, Yonsei University
134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea
E-mail: {[1]jun.lee, [2]kjaelee, [*]syleee}@ yonsei.ac.kr

**Abstract:** Converting 2D images to 3D models is a long-standing problem. Especially works on photometric stereo method have been reported that it requires multiple images at least three images upto thousands images with varying poses and illumination direction to generate a surface normal map and a depth map. We focus on this issue and propose a solution for photometric stereo with relit face images from a single face frontal image. First, with a relighting method, we can augment face images based on reference images with varying illumination directions. Using the generated virtual faces, relit face images, we can estimate surface normals using the Lambertian model and depth map using surface normals from photometric stereo method.

## 1. Introduction

Many works on 3D reconstruction have been researched. Some works use 3D scanners to reconstruct the objects and the others use photometric stereo method assuming Lambertian reflectance model. The first one needs a device to scan the model such as depth camera or 3D scanner. But the latter one only requires the hundreds or thousands photos, at least 3 or 4 pictures, of the object taken at the same pose and under the different light condition. If the photo is taken at different pose, then after aligning the photo, photometric stereo method is proceeded.

In this paper, we focus on the latter one, using photometric stereo method. Photometric stereo require photo collection of the object, which consists of more than three images[1, 3, 4]. But we introduce another way to reconstruct the 3D object by using a single frontal image.

In order to relight images with different light condition, we need feature points of the input query image and reference images. Finding feature points is also very challenging and there are many fine methods. In this paper, we adopt AAM(Active Appearance Method) algorithm [5].

Here are our contribution of this paper. First, we reconstruct 3D model with a single frontal face image by using photo-metric stereo method. Second, this is easy to implement.

In the next section, we introduce our proposed method. And section 3 presents the experimental setting and its result of our method. Lastly, section 4 concludes our work.

## 2. The Proposed method

In this section, we propose an algorithm which can generate depth map from a single fontal face image. Before all the process is going, we need to detect face featrue points. Without these information, we cannot generate relit imges and, of course, depth map as well. Therefore, we adopt AAM algorithm which is widely used to detect face feature points. And then, for the next process, relighting, we assume that the frontal face image is taken under the frontal light source. With the frontal face image, we can generate virtual images which is relit from different illumination conditions[2]. We can get surface normals using SVD from the image set which contains virtual images and the original image. Then a depth map can be estimated using maximum likelihood estimation method[4]. **Figure 1.** shows the flowchart of our proposed method.

### 2. 1 AAM

AAM is an algorithm which is used in many places, especially face recognition. Although it introduced almosty 10 years ago, it has been adopted in many papers.

AAM is concecptually an extension of Eigenfaces. And it is a generalization of the Active Shape Model (ASM) approach. There is a difference between AAM and ASM. AAM uses appearance of the face in the image, which means that it uses all the information of the face image, rather than juse near modeled shape.

After principle component analysis (PCA) is applied, we can get mean shape from the training image set. Then with the eigenface from PCA approach, we can estimate appearance model.

For testing, given a new face image, we place pre-trained model in the image. Then iteratively, we check the difference between model and the image. By decreasing the error of the distance between them, we update model by fitting the model into the face appearance of the image.

In this paper, we use 179 pictures of 90 subjects in NCKU CSIE Robotics lab face database [6] and 88 pictures of CUHK data set [7] to train a model which we can detect 68 face landmark points with. After finding the landmark feature points, we move onto the next process, relighting. **Figure 2.** (b) is shown as the result of the AAM.

### 2. 2 Relighting method

In order to relight face images, we use Extended YaleB database as reference images. We use two reference images. One is assumed to have similar illumination condition with the input image(A light direction, $\mathrm{I}_r^{\mathrm{A}}$ ). And the other reference image is the target face to be relit(B light direction, $\mathrm{I}_r^{\mathrm{B}}$). A ratio image **R** can be calculated by dividing $\mathrm{I}_{r,warped}^{\mathrm{A}}$ by the $\mathrm{I}_{r,warped}^{\mathrm{B}}$ in pixel-wise. We denote $\mathrm{I}_{r,warped}^{\mathrm{A}}$ and $\mathrm{I}_{r,warped}^{\mathrm{B}}$ as warped reference images by landmark points of the input image. J. Chen proposed to substitute the Cartesian product step with a locally constrained global optimization process[2].

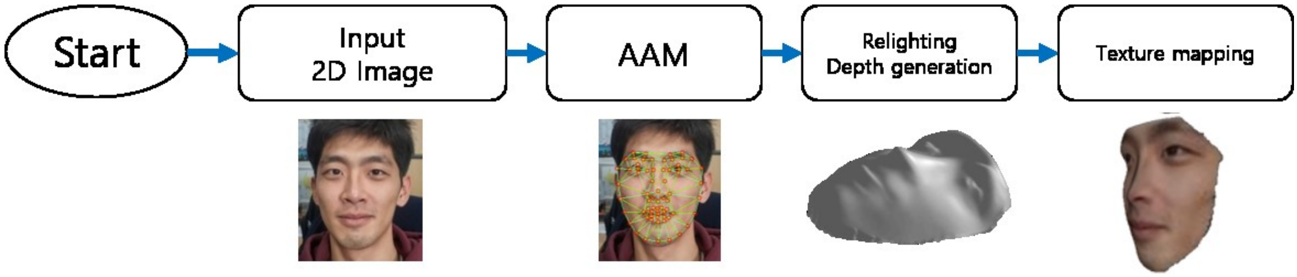


**Figure 1.** Flowchart of the system. First, we find feature points of the face image with AAM algorithm. After that, we apply relighting to generate virtual iamges and estimate depth information. Finally, with the depth map, we add texture information.

We assume the Lambertian reflectance model. Therefore, a face image can be expressed as below equation (1),

$$\mathrm{I} = \rho \mathbf{LS} \qquad \cdots \ (1)$$

which, I is an image, ρ is albedo, **L** is light source direction and **S** is shape normals which contains surface normals. All the images have the same albedo. It means that in a small local window, the images have only different illumination directions.

$$I_{1\omega}(i) = c_i I_{2\omega}(i) \qquad \cdots \ (2)$$

$c_i$ is $l_1/l_2$, nonnegative constant and it is called the local relighting coefficient associated with the local window $\omega$. With the local constraint, objective function can be achieved,

$$\mathrm{M} = \sum_i \left( \sum_{j \in \omega} \left( I_t^B(j) - c_i I_t^A(j) \right)^2 + \lambda (c_i - t_i)^2 \right) \cdots (3)$$

, which $t_i$ is the value of the ratio image at pixel i, and $\lambda$ is a nonnegative weight. After minimizing equation (3) for $c_i$, then we minimize the function for $\mathrm{I}_t^{\mathrm{B}}$.

### 2. 3 Photometric stereo

Photometric stereo assume Lambertian reflectance theory. This is the key of understanding Photometric stereo method.

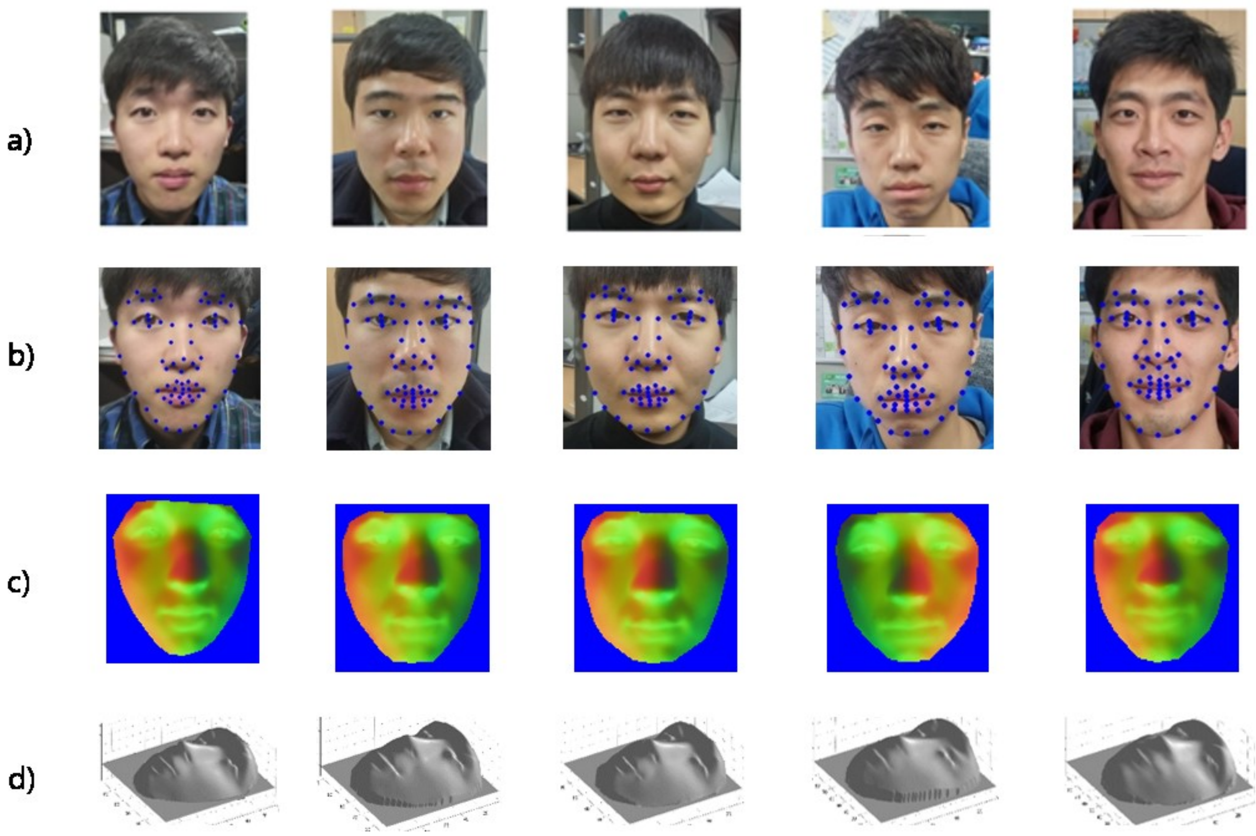


**Figure 2.** the result of the AAM, surface normal and depth map. First row (a) is the input image and the second row (b) shows result of the AAM. The third row (c) shows surface normal as a result of the photometric stereo method with relit images from the input image. Lastly, the fourth row(c) represents depth map estimated from surface normal map.

Images under varying illumination lie in a low-dimensional subspace[3]. Therefore the image set can be divided into equation (1). Using SVM, we can obtain.

$$\mathrm{I} = \mathrm{U\Sigma V^T} = \mathbf{LS} \quad \cdots \quad (4)$$

, where $\mathrm{U(N \times N)}$, $\Sigma(\mathrm{N \times H})$, $\mathrm{V(H \times H)}$ and $\mathrm{I(N \times H)}$, where N is number of images, H is number of pixels in each image. I's each row is the each image reshaped as 1-D row vector from the image set. And $\mathbf{L}$(light direction) can be calculated as $\mathrm{U}\sqrt{\Sigma}$, and $\mathbf{S}$(shape normal with albedo) is expressed as $\sqrt{\Sigma}V^T$. We apply 3D approximation method to get **L** and **S**.

S contains surface normal at z, x, and y axis with albedo. We estimate depth map from the surface normals. There are several method to estimate depth map. General way is, written like below called gradients estimates:

$$\mathrm{p} = -n_x/n_z \quad \cdots \quad (5)$$

$$\mathrm{q} = -n_y/n_z \quad \cdots \quad (6)$$

$$\begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{pmatrix} = \begin{pmatrix} D_x \\ D_y \end{pmatrix} \boldsymbol{z} + \boldsymbol{\epsilon} \quad \cdots \quad (7)$$

Equation (7) is a linear regression problem, where **z** is the depth map and $\mathrm{D_x}$ and $\mathrm{D_y}$ are finite difference operators respectively and . $\boldsymbol{\epsilon}$ represents noise vector. We assume that the vocariance matrix of the noise is large and sparse. Therefore equation (7) is a generalized least square problem. This method is called as linearized maximum likelihood depth-map estimation and is robust to Gaussian noise. And we adopt this method to estimate depth-map[4]. **Figure 2 (c)** and **(d)** shows the result of the photometric stereo method, surface normal map and depth map respectively.

## 3. Result

For the experiments, we use five images from the google image search. And we implement the approach on MATLAB 2013a. In order to warp the reference images and implement the relight method, AAM (Active Appearance Model) is used to detect 68 landmark points of the frontal face image.

**Figure 2** and **3** shows the result of our approach. **Figure 2** represents results of the AAM and photometric stereo. **Figure 3** shows texture mapped 3D models generated from the depth maps. The results seem very reasonable although we only use a single frontal face image. At the same time, there is an issue that they cannot give us accurate depth maps. If we look at the texture mapped 3D model in **Figure 3 (b)** and **(c)**, it seems to be similar with the original input image. However, in **Figure 2 (d)**, all the depth map seems not to have discriminative features. The reason is that it is dependent of the reference image's light information. For example, because of the shadow on the side of the reference face's nose, almost of the 3D model have the similar height of the nose. Except that limitation, these results are qualitatively acceptable and reasonable with one single image.


**Figure 3.** texture mapping on the depth map from **figure 2**. First row (a) is the input image and the second row (b) shows the texture mapped 3D model by using protometric stereo from the first row's images. Lastly, the thrid row (c) is as well as (b) at the different angle.

## 4. Conclusions

In this paper, we introduced a solution to reconstruct 3D model from a single frontal face image by using photometric stereo. There is limitation which are inaccurate height of the model because we use only one image. It is clear that depth estimation using a single image has less information than using multiple images more than three. In future work, we improve the relighting method to give us more precise depth information.


## Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01061315)


## References


[1] Yuille, Alan L., et al. "Determining generative models of objects under varying illumination: Shape and albedo from multiple images using SVD and integrability." *International Journal of Computer Vision* 35.3 (1999):203-222

[2] Chen, Jiansheng, et al. "Face image relighting using locally constrained global optimization." *Computer Vision–ECCV 2010*. Springer Berlin Heidelberg, (2010). 44-57.

[3] Epstein, Russell, Peter W. Hallinan, and Alan L. Yuille. "5/spl plusmn/2 eigenimages suffice: an empirical investigation of low-dimensional lighting models." Physics-Based Modeling in Computer Vision, 1995., Proceedings of the Workshop on. IEEE, (1995).

[4] Harrison, Adam P., and Dileepan Joseph. "Maximum likelihood estimation of depth maps using photometric stereo." Pattern Analysis and Machine Intelligence, IEEE Transactions on 34.7 (2012): 1368-1380.

[5] Cootes, Timothy F., Gareth J. Edwards, and Christopher J. Talyor. "Active appearance models" IEEE Transactions on Pattern Analysis & Machine Intelligence 6 (2001): 681-685.

[6] Nat, Cheng Kung Univ., Tainan, Taiwan, "NCKU CSIE Robotics Lab face database", http://robotics.csie.ncku.edu.tw/Database.htm, (2006).

[7] Wang, Xiaogang, and Xiaoou Tang., "Face photo-sketch synthesis and recognition", Pattern Analysis and Machine Intelligence, IEEE Transactions on 31.11 (2009): 1955-1967.