https://doi.org/10.2352/issn.2169-2629.2019.27.27

Removing gloss using Deep Neural Network for 3D Reconstruction

Futa Matsushita[1)], Ryo Takahasshi[1)], Mari Tsunomura[2)], Norimichi Tsumura[3)]
1) Graduate School of Science and Engineering, Chiba University, CHIBA, JAPAN
2) Department of Information and Image Science, Chiba University, CHIBA, JAPAN
3) Graduate School of Engineering, Chiba University, CHIBA, JAPAN

Abstract

3D reconstruction is used for inspection of industrial products. The demand for measuring 3D shapes is increased. There are many methods for 3D reconstruction using RGB images. However, it is difficult to reconstruct 3D shape using RGB images with gloss. In this paper, we use the deep neural network to remove the gloss from the image group captured by the RGB camera, and reconstruct the 3D shape with high accuracy than conventional method. In order to do the evaluation experiment, we use CG of simple shape and create images which changed geometry such as illumination direction. We removed gloss on these images and corrected defect parts after gloss removal for accurately estimating 3D shape. Finally, we compared 3D estimation using proposed method and conventional method by photo metric stereo. As a result, we show that the proposed method can estimate 3D shape more accurately than the conventional method.

1. Introduction

3D measurement is widely used for the inspection of industrial parts in order to ensure the quality during the production process or in order to maintain the safety of industrial products such as jet engines of airplanes. The 3D measurement technique can be classified into two types, contact and non-contact method. Figure 1 shows the example of the 3D measurement devices for industrial use. The contact 3D measurement uses a detector such as a probe or a cantilever in order to measure target shape [1]. In general, the contact method takes longer time compared with non-contact one. Non-contact measurements are widely used for industrial inspection since it is easy to inspect large target and it has less risk to damage the target.

The non-contact measurement also can be classified into two types, active method and passive method. The active method measured target shape by projecting structured light or slit light onto the target [2]. The passive method measure target shape only using images, such as stereo measurement [3]. In the conventional method, it is difficult to measure target that has gloss. In industrial inspection, the target surface is often smooth and glossy. The gloss makes it difficult to reconstruct target shape because of a lack of information.

A material estimation method based on deep neural network have been proposed by Meka *et.al.*[4] The method extracts specular shading, diffuse shading, and mirror image from input RGB image. The framework can be applied for 3D reconstruction of glossy target.

In this paper, therefore, we propose 3D shape reconstruction pipeline by utilizing framework of deep neural networks for the removal of gross on the target. The rest of this paper is organized as follows. First, we outline the conventional material estimation method. Second, we describe the gloss removal method for 3D reconstruction. Next, we describe the correction method for missing area after gloss removal. Then, we show the experimental result. Finally, we conclude our research.

(CRYSTA-Apex EX, Mitutoyo Corporation)

(a) Contact 3D Measurement [1]

(K-Scan MMDx, Nikon instech Co., Ltd)

(Mentor Visual IQ, General Electric Company)

(b) Non-Contact 3D Measurement [2, 3]

Figure 1. Example of 3D measurement devices for industrial use

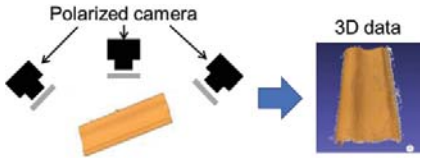


Figure 2.1. 3D measurement using polarized camera [10]

2. Conventional Researches for 3D Reconstruction

Recently, some methods for reconstructing 3D shape using RGB images are proposed. In Section 2.1, we describe conventional methods for 3D shape reconstruction using RGB images. In section 2.2, we describe a deep neural network used to remove gloss.

2.1 3D Reconstruction using RGB images

Many 3D shape reconstruction methods using RGB cameras have been proposed. Among them, binocular stereo is the simplest and easiest method. [5] In the stereo method, two RGB cameras are used to estimate a 3D shape (depth) from the parallax. However, if there is a gloss on RGB images, it is impossible to accurately reconstruct the 3D shape from RGB images of glossy object. There is also photometric stereo method used in this research. [6] The details will be described later in Chapter 5. In the photometric stereo method, camera is fixed, the illumination is applied from various directions and captured, and the three-dimensional shape is estimated from the photographed image group. In the photometric stereo method, the three-dimensional shape can be estimated if the camera position and the position of the illumination are known. In this method, estimation is performed using the diffuse reflection component, and it cannot be estimated well if there is specular reflection that represents gloss. The estimation results are presented in Section 4.

There is also multi-view stereo (MVS) [7]. This method can estimate 3D shape even if the camera position is unknown. Sparse point groups are obtained by Structure from Motion (SfM) [8], which detects feature points and estimates the camera position and attitude from them. Basically, if you know the camera position and attitude, you can obtain dense point group by stereo matching. Although this method has been extensively studied in recent years, the problem is that the object to be measured needs a texture. It becomes difficult to restore the three-dimensional shape for objects with few feature points or those whose feature points disappear due to gloss. Besides, a method of 3D shape reconstruction using deep learning was proposed [9], but the effect of gloss is hardly considered.

Finally, we explain the conventional method of 3D construction by removing the influence of gloss. [10] In this method, a polarized camera is used instead of the RGB camera as shown in Fig.2.1. In this method, a polarized camera is used instead of the RGB camera. Polarization characteristics can remove the influence of gloss, and 3D shape can be estimated accurately. However, it is not general to use polarized camera for 3D reconstruction.

2.2 Removal of gloss based on Deep Neural Network

In this section, we describe the gloss removal method based on Live Intrinsic Material Estimation (LIME) proposed by Abhimitra *et al*. [4]. Figure 2.2 shows the overview of LIME. In this method, the material of the general shape object is estimated in real time from the monochromatic RGB image based on the rendering equation. Deep learning is used to estimate various images and parameters. In the process, the gloss (specular reflection) can be removed.

LIME method is based on rendering equation show shown below:

$$L_o(x, \omega_o) = L_a + \int_\Omega f(x, \omega_o, \omega_i)(\omega_i \cdot n)E(\omega_i)d\omega_i \quad (1)$$

where L_o is observed radiance, x is a point on the object, L_a represents the radiance of ambient illumination, $f(x, \omega_o, \omega_i)$ represents BRDF[11], and an environment map $E(\omega_i)$ is used as incident light on an object. Diffuse reflection and specular reflection can be decomposed and considered by introducing

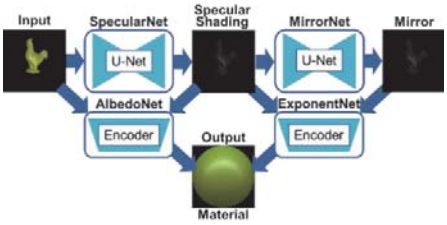


Figure 2.2. Overview of Live Intrinsic Material Estimation(LIME) [4]

Bling-Phong reflection model [12]. The equation for the Bling-Phong reflection model is shown below:

$$\mathrm{BP}(x, n, \omega_i, \omega_o) = m_d(\omega_i \cdot n) + m_s(h \cdot n)^s \quad (2)$$

where m_d represents the diffuse reflection color(diffuse albedo), m_s represents the specular reflectance (specular albedo), n is normal and h is half vector, and exponent s controls the spread of the specular reflection. Replacing BRDF with Eq. (2), Equation (1) is transformed as follows:

$$L_o = L_a + m_d \int_\Omega (\omega_i \cdot n)E(\omega_i)d\omega_i + m_s \int_\Omega (h \cdot n)^s E(\omega_i)d\omega_i \quad (3)$$

The diffuse and specular terms in Eq. (3) can be replaced as shown below:

$$L_o = L_a + m_d D + m_s S \quad (4)$$

Here, diffuse shading D and specular shading S are images that represent the spread of shading respectively. Parameters necessary to estimate the material are diffuse color m_d, specular reflectance m_s, and specular exponent s. Therefore, the material can be estimated by using the network architecture shown in Fig 2.2. We removed gloss using the part of this pipeline of LIME.

The relationship between the input image I after mask and the diffuse reflection color m_d, the specular reflectance m_s, the diffuse shading D, and the specular shading S is expressed by the following equation.

$$I = m_d D + m_s S \quad (5)$$

Here, $m_d D$ represents a diffuse reflection component. The specular reflection is estimated by SpecularNet from the input image I, and if specular reflectance m_s can be estimated using AlbedoNet, a diffuse reflection image is obtained. The diffuse reflection images can be calculated in the process of estimating the material in the LIME method.

3. Proposed Method of removing gloss

3.1 Proposed Method

In this section, we propose three methods for removing gloss using the network structure of LIME. The flow of estimation methods is shown in Fig. 3.1. The three gloss removal methods are shown below in detail

(a) The diffuse reflection image is obtained from the image estimated by LIME method.
(b) The specular reflection image is directly estimated using the network structure of SpecularNet, and the diffuse reflection image is obtained by subtraction from the input image.
(c) Directly estimate diffuse reflectance images using network structure of SpecularNet.

In this research, the effect of gloss was removed using these three methods.

3.2 Experimental set up

The data used for learning was downloaded from the LIME website. In this research, we limit the number of shapes and the number of images for learning due to time and machine performance. The shape was limited to a cylindrical type, and the number of images was reduced from 100,000 to 4,000. Also, an image used as learning data and teacher data is shown in Fig 3.1. The resolution of each image is 256 × 256. In addition, shape estimation becomes difficult if there is a bright and overexposed area. Therefore, in this research, learning is performed by removing the overexposed images (RGB values are all 255). As a result, 4,000 to 3744 were selected. In addition to the training data, 26 cylindrical images were prepared as test data.

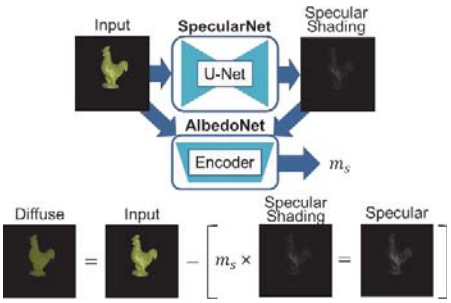
This is also used as an image for learning that is not overexposed. In the experimental environment, we used Keras [13] with Tensorflow as a backend and Geforce GTX 1080 as the GPU. We used ReLu function as activation function, Adam [14] for optimization, and applied Batch Normalization to each layer. According to the paper of LIME, the number of batches was 32 and the learning rate was 1.0×10^{-4}. In this study, unlike LIME, the number of epochs (number of learnings) was set to 300 times.

3.3 Result of gloss removal

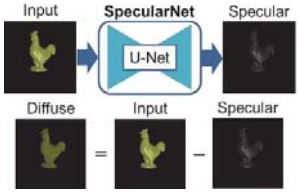
In this section, we present estimation results for the three gloss removal methods described in Section 3.1. Figure 3.2 shows the images of the estimation results of (a), (b) and (c). In addition, Table 3.1 shows the mean square error (MSE) of the diffuse reflection image output by each gloss removal method and the diffuse reflection image used as test data. Table 3.1 shows the average of the results of 3744 training data and 26 test data. In the gloss removal methods (a) and (b), the error is smaller for the results obtained by the training data, and for the gloss removal method (c), the results for the test data are smaller. Basically, the result for the training data is better, but in the gloss removal method (c), the error is large because the color change occurs in the image with pixels whose pixel values are saturated. Since there was no corresponding image in the test data, the error for the test data is smaller. Also, in the gloss removal methods (a) and (b), there is almost no difference in the MSE for the training data.

Table.1 MSE between output image and correct image for each gloss removal method

	(a) LIME	(b) Specular estimation	(c) Diffuse estimation
Learning data	17.20	16.65	134.20
Test data	60.40	34.51	25.95



(a) LIME method for removal gloss



(b) Directly specular image estimation *(c) Directly diffuse image estimation*

Figure 3.1. Propsed method of removal gloss based on Live Intrinsic Material Estimation(LIME)

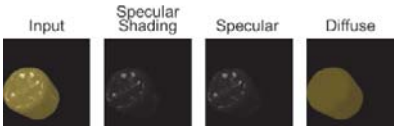


Figure 3.2. Samples for learning and test data

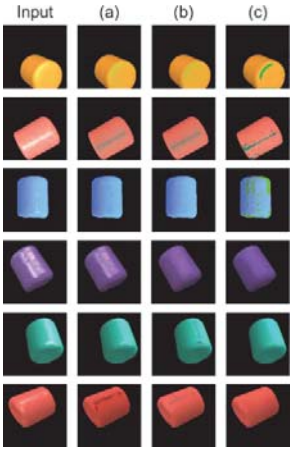


Figure 3.3 Example of Result after gloss removal. (a) gloss removal using LIME architecture, (b) gloss removal by directly estimating specular image using SpecularNet architecture, (c) gloss removal by directly estimating diffuse image using SpecularNet architecture

However, the MSE for test data is better than the gloss removal method (b). This can be seen by comparing the figures with the figures, but the result of the gloss removal method (a) has many images for which the gloss cannot be removed. The reason for this is that the method using LIME is not a network built for gloss

removal. Estimated by SpecularNet and AlbedoNet respectively, the error of result were greatly increased.

Therefore, in this research, the gloss removal is performed using the gloss removal method (b) which is the most accurate in the case of 3D shape estimation using the created CG image group.

The result of the gloss removal method shown in the previous chapter has the problem that a defect occurs when the gloss is removed. In this chapter, we propose a correction method for the defect after gloss removal. Image inpainting is used to compensate for the defects created by gloss removal. Image inpainting, as shown in Fig 3.4, is a method for repairing defects such as scratches on photos and removing unnecessary objects and characters. In this research, we use the image inpainting method [15] proposed by Alexandru et al. This method is a method of calculating the weighted sum from surrounding pixels that are known and filling the restoration point. In image inpainting, basically, a mask image is used to specify a portion to be repaired or corrected. However, in this research, we use the binarized specular reflection image estimated as a mask image as shown in Fig 3.5. However, since the specular reflection image is not binarized, we set the threshold manually in this research.

Figure 3.4 Example of image inpainting [15]

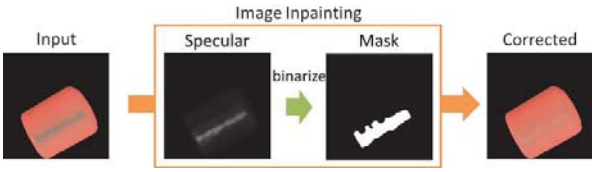


Figure 3.5 Proposed method for defect correction by image inpainting method [15]

4. Comparison of 3D shape estimation results for gloss objects

In this section, we show the image generation with the method of estimating 3D shape, and the result of restoring 3D shape.

4.1 3D estimation by photometric stereo method

In this research, the photometric stereo method is used as a reconstruction method of 3D shape. Figure 4.1 shows an example of the photometric stereo method. In the photometric stereo method, it is possible to estimate the surface normal by using an image in which three or more illumination positions are changed theoretically. In the photometric stereo method, a Lambert diffuse reflection surface is assumed, and the camera is fixed in one direction, and a normal light source is projected from various directions to estimate the normal. When we estimate a normal map from three images, a normal vector obtained when observed luminance $\mathbf{i} = (i_1, i_2, i_3)$ and incident light vector $\mathbf{s} = (s_1, s_2, s_3)$ for a certain pixel. The relationship with $\mathbf{n} = (n_x, n_y, n_z)$ is expressed by the following equation:

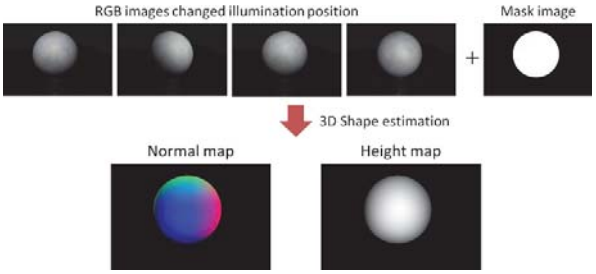


Figure 4.1. Example of Photometric stereo method [6]

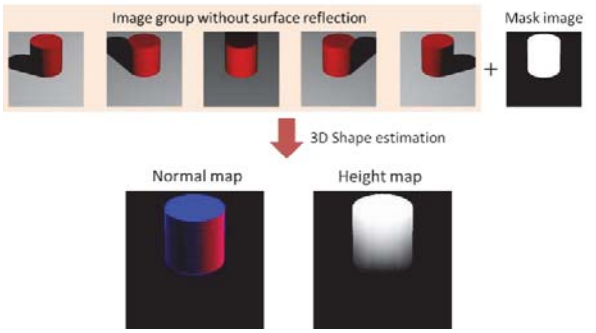
$$\mathbf{i} = \mathbf{s}\rho\mathbf{n} \tag{6}$$

where ρ can estimate the normal vector from the following equation by obtaining the inverse matrix of $\mathbf{s}$ is to determine the diffuse reflectance pseudo-normal vector $\tilde{\mathbf{n}} = \rho\mathbf{n}$.

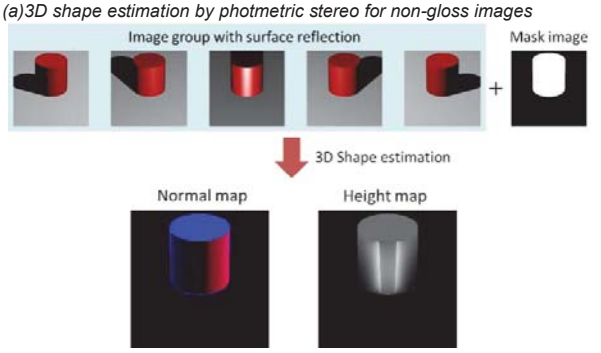
$$\tilde{\mathbf{n}} = \mathbf{s}^{-1}\mathbf{i} \tag{7}$$

This method generates a normal map and a height map from images in which the illumination direction is changed.

An image group used for 3D shape estimation was created by PBRT. The shape is limited to a cylinder, and the material is substrate. As shown in Fig. 4.2, the 5 images with illumination directions changed is created gloss and non-gloss image groups are generated by changing the values that control surface reflection. We compare 3D shape estimation using these image groups.



(a)3D shape estimation by photmetric stereo for non-gloss images



(b)3D shape estimation by photmetric stereo for gloss images

Figure 4.2. 3D shape estimation using 5 images with changing illumination created by PBRT

4.2 Comparison of 3D shape estimation

At the beginning, a 3D normal map and height map are estimated by the photometric stereo method for glossy and nonglossy image groups. The estimated results are shown in Fig 4.2. In this research, we use the result of Fig. 4.2(a) estimated using an object without gloss as the correct image. The normal map indicates that the direction of the normal is different when the colors are different, and the height map indicates that the closer to white the higher the depth direction is.

Next, gloss removal is performed on the glossy image group, and the image inpainting is used to correct the defective portion after the gloss removal. The flow is shown in Fig 4.3. The normal map and height map are estimated for each image group shown in Fig 4.3. A summary of the 3D shape estimation results for each image group is shown in Fig 4.4. Also, Figure 4.5 shows the result of subtraction with the correct image and the result of calculating the mean square error. The height map represents the shape, and the results after defect correction are very accurate compared to results of gloss and gloss removal. With regard to the normal map, there were some parts where the result was a little worse due to the removal of gloss and correction of the defect part. However, the center part after defect correction were corrected correctly, the result is better compared to the other methods. It can be considered that more accurate normal map can be obtained by correcting the missing part correctly. As a result of this research, it was found that applying the gloss removal method to a glossy object and correcting it further improved the estimation result regarding the height.

5. Conclusion and Future works

In this research, we can estimate the 3D shape with high accuracy by removing the gloss from the image group captuerd the glossy object. Section 2 enumerates methods for estimating 3D shapes and describes conventional methods that remove the influence of gloss. Section 3 proposes methods based on deep neural networks used for gloss removal. We compares and verifies the results of the proposed method. As there are defect parts in the image after gloss removal, image inpainting was applied to correct it. In Section 4, we applied the proposed method in this research to compare the conventional method with 3D shape was estimated after gloss removal and defect correction. As a result, it is shown that 3D shape can be estimated with high accuracy compared to the conventional method.

As future problems, since the shape and color are limited, it is possible to restore the three-dimensional shape for various shapes and colors, and to use the Multi-View Stereo method as well as the photometric stereo method. It is necessary to perform 3D shape estimation corresponding to various lighting environments. Finally , we try to automate all the 3D reconstruction.

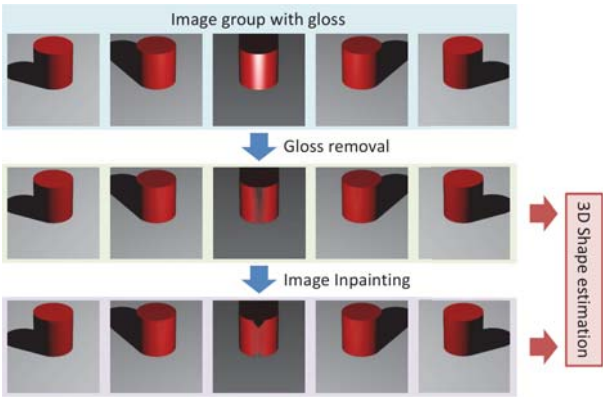


Figure 4.3 Flow of image processing for comparing 3D reconstruction by conventional and proposed method

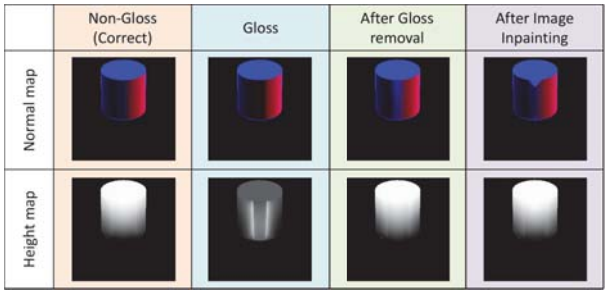


Figure 4.4 Results of 3D estimation for each images after image processing

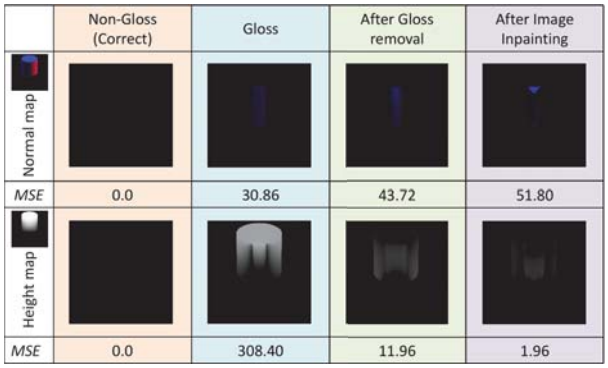


Figure 4.5. Results of subtraction and Mean Squared Error(MSE) between the correct image and the estimated image

Reference

[1] http://jbm.co.jp/cai/Master3DGage/index.html

[2] http://www.nikon-instruments.jp/jpn/industrial-products/3d-metrology/walk-around-scanner/k-scan-mmdx

[3] https://www.geoilandgas.co.jp/it/product#

[4] Abhimitra Meka, Maxim Maximov ,Michael Zollhöfer ,Avishek Chatterjee, Hans-Peter Seidel ,Christian Richardt ,Christian Theobalt, "LIME: Live Intrinsic Material Estimation", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition

[5] D. Scharstein, R. Szeliski, "High-accuracy stereo depth maps using structured light", 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[6] Woodham RJ, "Photometric method for determining surface orientation from multiple images," Optical engineering, vol. 19, no. 1, pp. 191 139‐191 139, 1980.

[7] Yasutaka Furukawa, Jean Ponce, "Accurate Camera Calibration from Multi-View Stereo and Bundle Adjustment", Int J Comput Vis (2009) 84: 257. https://doi.org/10.1007/s11263-009-0232-2

[8] MICHELETTI, N., CHANDLER, J.H. and LANE, S.N., 2015. Structure from motion (SFM) photogrammetry. IN: Clarke, L.E. and Nield, J.M. (Eds.) Geomorphological Techniques (Online Edition). London: British Society for Geomorphology. ISSN: 2047-0371, Chap. 2, Sec. 2.2.

[9] Despoina Paschalidou, Ali Osman Ulusoy, Carolin Schmitt, Luc Gool, Andreas Geiger, "RayNet: Learning Volumetric 3D Reconstruction with Ray Potentials", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition

[10] Zhaopeng Cui ; Jinwei Gu ; Boxin Shi ; Ping Tan ; Jan Kautz, "Polarimetric Multi-view Stereo", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[11] M Ashikhmin, P Shirley, "An anisotropic phong BRDF model", Journal of graphics tools, 2000, Pages25-32

[12] J. F. Blinn. "Models of light reflection for computer synthesizedpictures" Computer Graphics (Proceedings of SIGGRAPH), 11(2):192‐198, 1977.

[13] F. Chollet et al. Keras. https://github.com/fchollet/keras, 2015.

[14] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations (ICLR), 2015.

[15] Alexandru Telea, "An Image Inpainting Technique Based on the Fast Marching Method", Journal of graphics tools, 2004, Vol. 9, No. 1: 25−36