ICT Technical Report ICT-TR-02-2010: Specular Normal Synthesis Using Stochastic Super-resolution for Detailed Facial Geometry

Jun Zheng and Abhijeet Ghosh

Abstract

Detailed facial geometry is critical for the visual realism of face models in computer games, movies, and virtual reality applications. The existing face scanning methods, however, are either sacrificing resolution for real-time processing, or requiring expensive high-speed cameras. In this work we propose a new technique for real-time high-resolution facial scanning using spherical gradient illumination. The key elements of the approach are the use of stochastic super-resolution to generate specular normal map based on diffuse normal map, instead of capturing both of them during scanning process. We analyze a training dataset of diffuse normal maps and specular normals of a particular object and learn the mapping from low-frequency components of diffuse normal maps to high-frequency components of specular normal maps of that object. This enables us to infer, for example, the most likely high-resolution specular normal map detail depicting the same person as a low-resolution diffuse normal map given as input. Experimental results show that the proposed algorithm generates high-quality specular normal maps from diffuse normal map inputs.

1 Introduction

Detailed facial geometry contributes significantly to the visual realism of face models in computer games, movies, virtual reality applications, and so on.

In this work we propose a new technique for acquiring specular normal map for high-resolution facial scanning using spherical gradient illumination. The key elements of the approach are the use of stochastic super-resolution to generate specular normal map based on diffuse normal map, instead of capturing both of them. We analyze a training dataset of diffuse normal maps and specular normals of a particular object and learn the mapping from low-frequency components of diffuse normal maps to high-frequency components of specular normal maps of that object. This enables us to infer, for example, the most likely high-resolution specular normal map detail depicting the same person as a low-resolution diffuse normal map given as input. Experimental results show that the proposed algorithm generates high-quality specular normal maps from diffuse normal map inputs.

The main contributions of this work are:

1. acquire specular normal map by synthesizing using super-resolution the specular normals from diffuse normals instead of capturing both of them.
2. transfer specular normal information from one person to other people.
3. transfer specular normal information to unpolarized data.
4. demonstrate that the synthesized normal map can be used for image-based lighting.

2 Related work

Facial Scanning There have been a lot of work on high-resolution facial scanning. The existing methods are based on laser scanning, or structured light scan, or use photometric stereo.

Laser scanning projects a laser dot or line onto an object, and then a sensor measures the distance to the surface, then data is collected and recorded as point clouds which can be converted into a triangulated mesh. One advantage of laser scanning is that Scanning device can be hand-held and portable.

Structured light scanning projects a pattern of light on the subject, and looks at the deformation of the pattern on the subject, and then uses a technique similar to triangulation to calculate the distance of every point on the line [Rusinkiewicz et al. 2002; Zhang et al. 2004; Davis et al. 2005; Zhang and Huang 2006]. These methods can achieve real-time processing, but they fail to provide high-resolution details for facial scanning.

Photometric stereo [Woodham 1980; Woodham 1989; Wenger et al. 2005; Malzbender et al. 2006; Hernández et al. 2007] estimates the surface normals of objects by observing that object under different lighting conditions. The technique was originally introduced by [Woodham 1980]. Recently, Wenger et al. [Wenger et al. 2005] use time-multiplexed and high-speed cameras to capture time-varying reflectance properties, such as surface normals, albedo and ambient occlusions, of a live performance. Hernandez et al. [Hernández et al. 2007] capture and track 3d moving data in real-time using multi-spectral photometric stereo. Malzbender et al. [Malzbender et al. 2006] use photometric stereo to recover per-pixel estimates of surface orientation, and then transform the reflectance based on recovered normal directions to enhance the surface detail. However they require high-speed camera and fast GPU to achieve real-time processing, and they don't account for specularities and shadows.

Ma [Ma et al. 2007] introduced a technique for high-resolution face scanning that combining structured light and photometric stereo scanning. They first obtain base geometry through structured light scan, and then capture photometric surface normal from either diffuse or specular reflectance using four spherical gradient illumination patterns. The surface normals are later used to optimize the base geometry to recover detailed facial features using an embossing process as in [Nehab et al. 2005].

In this work, we go further than previous work for acquiring specular normal map by synthesizing the specular normals from diffuse normals instead of capturing both of them, and then generate detailed facial geometry.

Super Resolution Baker and Kanade [Baker and Kanade 2000; Baker and Kanade 2002] first developed a face hallucination method based on a prior on the spatial distribution of the image gradient for frontal face images. It infers the high-frequency component from a parent structure by recognizing the local features of the training set, and aims to recover extremely high-quality HR images of human faces from LR images. For example, given a LR image of 12×16 pixels only, which could barely be recognized as a face, face hallucination can synthesize an HR image of 96×128 pixels.

A successful face super-resolution algorithm must meet the global constraints, which means that the results must have common human characteristics, and the local constraints, which means that the results must have specific characteristics of a particular face

image [Liu et al. 2007]. To fulfill these two constraints, Liu et al. [Liang et al. 2001; Liu et al. 2007; Liu et al. 2001] introduce a two-step statistical hybrid modeling approach that integrates both a global parametric model and a local non-parametric model. The first step is to derive a global linear model to learn the relationship between HR face images and the corresponding smoothed down-sampled LR ones. The second step is to model the residue between an original HR image and the reconstructed HR image by a non-parametric Markov network. Then by integrating both global and local models, they generate the photo-realistic face images.

However the above methods use probabilistic models and are based on an explicit resolution reduction function, which is usually difficult to acquire in practice [Wang and Tang 2005]. Instead of using a probabilistic model, Wang et al. [Wang and Tang 2005; Wang and Tang 2003] propose a face hallucination model using PCA to represent the structural similarity of face images. They render the new hallucinated face image by mapping between the LR and HR training pairs of face images. In the PCA representation, different frequency components are independent. By selecting the number of eigenfaces, they extract the maximum amount of facial information from the low-resolution face image and remove the noise.

Motivated by the fact that belief propagation converges quickly to a solution of the Markov network, Freeman et al. [Freeman et al. 2002; Freeman et al. 2000] explore a simpler and faster one-pass algorithm, which uses the same local relationship information as the Markov networks but requires only a nearest-neighbor search in the training set for a vector derived from each patch of local image data. Their algorithms are an instance of a general-training-based approach that can be useful for image processing or graphics applications. It can be applied to enlarge images, remove noise, and estimate 3D surface shapes.

3 Method

The facial scanning system captures the face geometry by combining a medium-resolution structured light scan base geometry with high-resolution surface normals. The structured-light stereo uses a stereo pair of cameras plus a projector placed between them. The surface normals are obtained from gradient illumination and by placing individual linear polarizers over each light. This work synthesis the specular normals from diffuse normals instead of capturing both of them.

3.1 Facial Scanning

We capture the face geometry by combining a medium-resolution structured light scan with high-resolution surface normals based on the specular reflectance. The structured-light stereo uses a stereo pair of SLR cameras plus a projector placed between them. The projector projects a sequence of four colored strip patterns and one uniform white pattern. In capturing, we take 8 photos for each gradient pattern under two linear polarization states, and 5 stereo photos for each structures light strip patterns, using Cannon 5D cameras in burst mode which requires just a few seconds to capture data at 12 megapixel resolution [Ma et al. 2007].

Because of noise and the limited resolution of the projector, the structured light scan introduces some high frequencies biasing and noise. We have to smooth the structured light scan surface using bilateral denoising, and then create a surface normal map from the smoothed mesh and extract the high frequency details of the estimated normals using high-pass filtering [Ma et al. 2007]. Finally, we optimize the mesh vertices to match this assembled normal map using an embossing process as in [Nehab et al. 2005].

We obtain diffuse and specular normals from gradient illumination for objects whose reflectance is either diffuse or specular, under spherical illumination. The primary light apparatus we use to create the gradient lighting patterns consists of about 150 LED lights placed on the vertices and edges of a twice-divided icosahedron. For polarized patterns, individual linear polarizers are placed over each light. A linear polarizers is mounted on a servomotor in front the camera, which enable to polarizer to be rapidly flipped on its diagonal between horizontal and vertical orientations [Ma et al. 2007].

3.2 Specular Normal Map Synthesis

Given a diffuse normal map as input, to construct the corresponding specular normal map, we first filter the diffuse normal map with a normal map filter similar to Gaussian high-pass filter, and then subdivide the filtered diffuse normal map into patches, which we call the low-frequency patches, by scanning a window across the image in raster-scan order. Similarly, we also filter and subdivide the specular normal map in the training set into patches which we call high-frequency patches.

To synthesize a specular normal map, for each low-frequency patch, a high-frequency patch of the training specular normal map is selected by a stochastic search from the training set based on local diffuse normal details and adjacent, previously determined specular normal map patches. The selected high-frequency patch should not only come from a location in the training images that has a similar corresponding low-frequency appearance, but it should also match at the edges of the high-frequency patch with the overlapping pixels, which we call high-frequency overlap, of its previously determined high-frequency neighbors to ensure that the high-frequency patches are compatible with those of the neighboring high-frequency patches.

In this work we synthesize the specular normal map with an algorithm that is an extension of the one-pass algorithm, proposed by Freeman et al. [Freeman et al. 2002; Freeman et al. 2000]. In the one pass algorithm, we first concatenate the pixels in the low-frequency patch and the high-frequency overlap to form a search vector. The training set also contains a set of such vectors. Then we search for a match by finding the best match in the training set. When we find a match we extract the corresponding high-frequency patch from training data set and add it to the initial diffuse normal map to obtain the output specular normal map.

Mathematically, this process can be described as follows. Suppose we have a training data set

$$\{\langle (x^{(i,j,k)}(R,G,B), y^{(i,j,k)}(R,G,B)), z^{(i,j,k)}(R,G,B)\rangle,$$
$$i = 1,2,\ldots,l; j = 1,2,\ldots,m; k = 1,2,\ldots,n\}$$

where $x^{(i,j,k)}(R,G,B)$ is low-frequency patch at the i^{th} row and j^{th} column of the k^{th} training diffuse normal map, $y^{(i,j,k)}(R,G,B)$ is the corresponding high-frequency overlap and $z^{(i,j,k)}(R,G,B)$ is the corresponding high-frequency patch of the training specular normal map, l is the number of rows of patches in a training image, m is the number of columns of patches in a training image and n is the number of training images.

Given an input low-frequency patch $\overline{x}(R,G,B)$, we need to find an high-frequency patch $z^{(i',j',k')}(R,G,B)$ such that

$$(i',j',k') = \arg\min_{(i,j,k)}\{d(\overline{x}(R,G,B)\cdot\beta, x^{(i,j,k)}(R,G,B)\cdot\beta)$$
$$+\alpha * d(y^{(i,j,k)}(R,G,B)\cdot\beta, y_N^{(i,j,k)}(R,G,B)\cdot\beta)\}$$

where $d(x,y)$ is the Euclidean distance between x and y, $y_N^{(i,j,k)}(R,G,B)$ is the overlap of $z^{(i,j,k)}(R,G,B)$ with the adjacent, previously determined high-frequency patches, α is a user-controlled weighting factor, β balances the contribution of each color channel, and $z^{(i',j',k')}(R,G,B)$ is the selected high-frequency patch.

Based on the fact that we are working on facial scanning, this work uses stochastic search to find the best match by exploiting the fact that face patches maintain relatively tight distributions for shape at successive iterations. For example, suppose we found the best match at iteration t, which is a patch from a left eye taken from training image T; intuitively, there is a high probability that the best match at iteration $t+1$ should also belong to a left eye and come also from T , or from a training image that is similar to T. Therefore the position of the most compatible high-frequency patch at iteration t, z_t, and its history $Z_t = z_1, z_2, \ldots, z_{t-1}$ form a temporal Markov chain, so that the new position is conditioned directly only on the immediately preceding state, independent of its earlier history.

$$P(z_t|z_t) = P(z_t|z_{t-1})$$

Therefore, based on z_{t-1}, we could estimate the most likely positions of z_t and we just need to search these positions instead of the exhaustive search performed by the conventional one-pass algorithm.

Let $z_t^{(i,j,k)}$ be the most compatible patch found at iteration t (for $t = 0$, an exhaustive search needs to be performed). To find the best patch at iteration $t+1$, we generate a set of N candidate patch locations $\{(i',j',k')^1, ..., (i',j',k')^N\}$ around location (i,j) in images that are similar to image k.

Each location $(i',j',k')^q$ will be randomly generated according to the following distributions:

$$\begin{cases} i' = i + \alpha \\ j' = j + \beta \\ k' = \gamma(A(k)) \end{cases}$$

where α and β are normally distributed random variables, $A(k)$ is a list of face normal maps that are similar to normal map k (which includes k itself) and $\gamma(.)$ is a sampling function that randomly selects face normal maps from $A(k)$ with a probability that is directly proportional to their similarity with face normal map k. In a preprocessing stage, we build a directed graph where every vertex contains a node and there is an edge $e(k,v)$ if vertices k and v contain similar face normal maps, according to a mean-squared distance metric. Thus $A(k)$ is the set of vertices that are adjacent to k in this graph.

After generating the N candidate patches, we select as the next patch the best match within the generated point set and repeat the process again.

The proposed algorithm generates a set of N candidates around the point (i,j,k). Suppose we have n images in our training data set and each image has k patches, for exhaustive search, the running time for constructing one local face is $O(nk^2)$; for stochastic search, the running time is $O(kn + kN)$.

4 Experimental results

For experiments we use a high dynamic range (HDR) face dataset which consists of floating-point images of about 1944×1296 pixels with a frontal view of about 10 different expressions under 4 gradient lighting patterns and 5 structured-light patterns. The primary light apparatus we use to create the gradient lighting patterns consists of about 150 LED lights placed on the vertices and edges of a twice-divided icosahedron. For polarized patterns, individual linear polarizers are placed over each light. A linear polarizers is also mounted on a servomotor in front the camera, which enable the polarizer to be rapidly flipped on its diagonal between horizontal and vertical orientations to capture both diffuse and specular reflectance.

In preprocessing we register the normal maps using manually drawn facial masks (Figure 1), so that we can assume that the same parts of faces appear in roughly the same parts of the images. The diffuse and specular normal map size is fixed to 1944×1296 pixels and we use them as the training images.

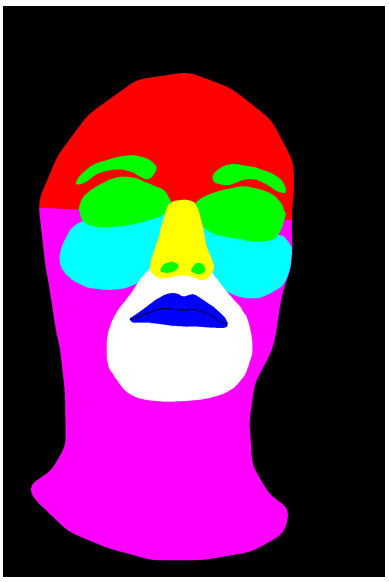


Figure 1: Manually drawn facial masks.

To construct our training dataset, we filter the diffuse normal maps with a normal map filter similar to Gaussian high-pass filter. Then we subdivide the filtered images into low-frequency patches by scanning a 8×8 pixel window across the image in raster-scan order. Then we again filter and subdivide the training specular normal maps into 8×8 pixel high-frequency patches. At each step we also get a 17-pixel overlap of each high-frequency patch with the high-frequency patches above and to the left. Then we create our training vectors by concatenating the low-frequency patches and corresponding high-frequency overlaps. In practice, the size of low-frequency patches and high-frequency patches is not necessarily the same. The parameter α, which controls the trade-off between matching the low-frequency patches and finding the most compatible high-frequency patches, is set to 0.2. The parameter β balances the contribution of each color channel.

Figure 2 shows the specular normal map synthesis results of the same person with different expressions. From the results, we can see that the synthesized specular normal map has much clearer detailed features than diffuse normal map, and almost the same as the captured specular normal map.

Figure 3 shows the geometry optimization results of the same person with different expressions. In this figure, we present the geometry optimized with our synthesized specular normal maps. For comparison, we also present the base geometry captured through structured-light scan without any optimization, and the geometry optimized with the captured specular normal map.

Figure 4 shows the image-based lighting rendering results of the geometry optimized with synthesized specular normal map. The results show that our synthesized specular normal maps can also work with image-based lighting techniques.

Figure 5 to 7 show specular normal map synthesis results of cross-object transfer. The algorithm transfers specular information of our training dataset to another person.

Figure 8 to 10 show specular normal map synthesis results from unpolarized mixed normal map. The algorithm transfers specular information of our training dataset to unpolarized data. The results show that our algorithm can improve the quality of facial scanning without polarization by adding specular normal map information to unpolarized data.

5 Conclusions and future work

In this work we propose a new technique for acquiring specular normal map for high-resolution facial scanning using spherical gradient illumination. The key elements of the approach are the use of stochastic super-resolution to generate specular normal map based on diffuse normal map, instead of capturing both of them. We analyze a training dataset of diffuse normal maps and specular normal maps of a particular object and learn the mapping from low-frequency components of diffuse normal maps to high-frequency components of specular normal maps of that object. This enables us to infer, for example, the most likely high-frequency specular normal map detail depicting the same person as a low-resolution diffuse normal map given as input. Experimental results show that the proposed algorithm generates high-quality specular normal maps from diffuse normal map inputs.

To further reduce the running time of the super-resolution algorithm, we will use stream processing to parallelize the execution. Stream processing permits the execution of data-parallel algorithms with stream processors such as graphic processing units (GPUs), while using the central processing unit (CPU) for other purposes simultaneously. This would enable a conventional PC to run the image super-resolution algorithm in real time.

References

ATKINS, C. B., BOUMAN, C. A., AND ALLEBACH, J. P. 2001. Optimal image scaling using pixel classification. In *2001 International Conference on Image Processing. Volume*, 864–867.

BAKER, S., AND KANADE, T. 2000. Hallucinating faces. In *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*.

BAKER, S., AND KANADE, T. 2002. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 9, 1167–1183.

BÉGIN, I., AND FERRIE, F. P. 2006. Comparison of super-resolution algorithms using image quality measures. In *Proceedings of the 3rd canadian conference on computer and robot vision*, IEEE Computer Society, Washington, DC, USA, 72.

CAPEL, D. 2004. *Image Mosaicing and Super-Resolution (Cphc/Bcs Distinguished Dissertations.)*. SpringerVerlag.

CUTLER, R., AND DAVIS, L. S. 2000. Robust real-time periodic motion detection, analysis, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 8 (August), 781–796.

DAVIS, J., RAMAMOORTHI, R., AND RUSINKIEWICZ, S. 2005. Spacetime stereo: a unifying framework for depth from triangulation. *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI) 27*, 2 (February), 296–302.

DEDEOĞLU, G. 2007. *Exploiting Space-Time Statistics of Videos for Face "Hallucination"*. PhD thesis, The Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania.

DODGSON, N. 1997. Quadratic interpolation for image resampling. *IEEE Transactions on Image Processing 6*, 9 (September), 1322–1326.

FREEMAN, W. T., PASZTOR, E. C., AND CARMICHAEL, O. T. 2000. Learning low-level vision. *International Journal on Compter Vision 40*, 1, 25–47.

FREEMAN, W. T., JONES, T. R., AND PASZTOR, E. C. 2002. Example-based super-resolution. *IEEE Computer Graphics and Applications 22*, 2, 56–65.

GIROD, B. 1993. What's wrong with mean-squared error? In *Digital images and human vision*, 207–220.

GREENSPAN, H., ANDERSON, C., AND AKBER, S. 2000. Image enhancement by nonlinear extrapolation in frequency space. *IEEE Transactions on Image Processing 9*, 6 (June), 1035–1048.

HERNÁNDEZ, C., VOGIATZIS, G., BROSTOW, G. J., STENGER, B., AND CIPOLLA, R. 2007. Non-rigid photometric stereo with colored lights. In *IEEE international conference on computer vision*.

ISARD, M., AND BLAKE, A. 1998. Condensationconditional density propagation for visual tracking. *International Journal of Computer Vision 29*, 1 (August), 5–28.

JIA, K., AND GONG, S. 2005. Multi-modal tensor face for simultaneous super-resolution and recognition. In *IEEE International Conference on Computer Vision*, 1683–1690.

KOMATSU, T., AIZAWA, K., IGARASHI, T., AND SATIO, T. 1993. Signal-processing based method for acquiring very high resolution image with multiple cameras and its theoretical analysis. In *IEE Proceedings of Communications, Speech and Vision, I*, 19–25.

LANDER, K., BRUCE, V., AND HILL, H. 2001. Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology 15*, 1 (August), 101–116.

LI, Y., AND LIN, X. 2004. Face hallucination with pose variation. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 723–728.

LIANG, L., LIU, C., XU, Y., GUO, B., AND SHUM, H.-Y. 2001. Real-time texture synthesis by patch-based sampling. *ACM Transactions on Graphics 20*, 3 (July), 127–150.

LIU, C., SHUM, H.-Y., AND ZHANG, C.-S. 2001. A two-step approach to hallucinating faces: global parametric model and local nonparametric model. In *Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, 192–198.

LIU, W., LIN, D., AND TANG, X. 2005. Hallucinating faces: TensorPatch super-resolution and coupled residue compensation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 478– 484.

LIU, C., SHUM, H.-Y., AND FREEMAN, W. T. 2007. Face hallucination: Theory and practice. *International Journal of Computer Vision (IJCV) 75*, 1 (October), 115–134.

LÓPEZ, L. D., AND FUENTES, O. 2007. Color-based road sign detection and tracking. In *International Conference on Image Analysis and Recognition (ICIAR)*.

MA, W.-C., HAWKINS, T., PEERS, P., CHABERT, C.-F., WEISS, M., AND DEBEVEC, P. 2007. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Eurographics Symposium on rendering*, 183–194.

MALZBENDER, T., WILBURN, B., GELB, D., AND AMBRISCO, B. 2006. Surface enhancement using real-time photometric stereo and reflectance transformation. In *Eurographics Symposium on rendering*, 245–250.

NEHAB, D., RUSINKIEWICZ, S., DAVIS, J., AND RAMAMOORTHI, R. 2005. Efficiently combing positions and normals for precise 3d geometry. *ACM Transactions on Graphics 24*, 3, 536–543.

PARK, S. C., PARK, M. K., AND KANG, M. G. 2003. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE 20*, 3 (May), 21–36.

POLANA, R. B. 1994. *Temporal Texture and Activity Recognition*. Ph.d. dissertation, University of Rochester, Rochester, N.Y.

RUSINKIEWICZ, S., HALL-HOLT, O., AND LEVOY, M. 2002. Real-time 3d model acquisition. *ACM transactions on graphics 21*, 3, 438–446.

WANG, X., AND TANG, X. 2003. Face hallucination and recognition. In *Proceedings of the Fourth International Conference on Audio- and Video-Based Personal Authentication (IAPR)*, 486–494.

WANG, X., AND TANG, X. 2005. Hallucinating face by eigentransformation. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews 35*, 3 (August), 425–434.

WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4 (April), 600–612.

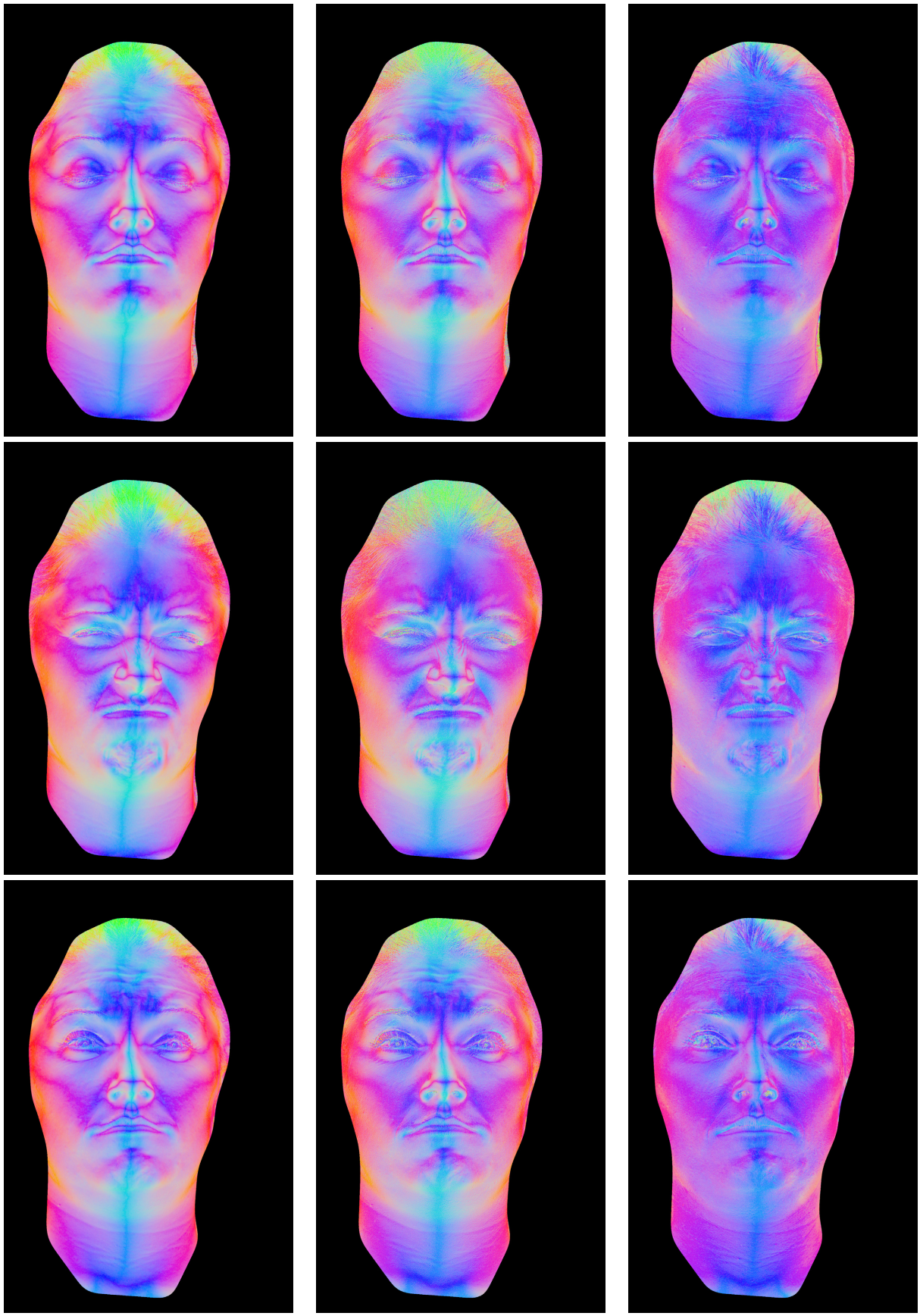
WENGER, A., GARDNER, A., TCHOU, C., UNGER, J., HAWKINS, T., AND DEBEVEC, P. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics 24*, 3, 756–764.

WOODHAM, R. J. 1980. Photometric method for determining surface orientation from multiple images. *Optical Engineering 19*, 1, 139–144.

WOODHAM, R. J. 1989. Photometric method for determining surface orientation from multiple images. 513–531.

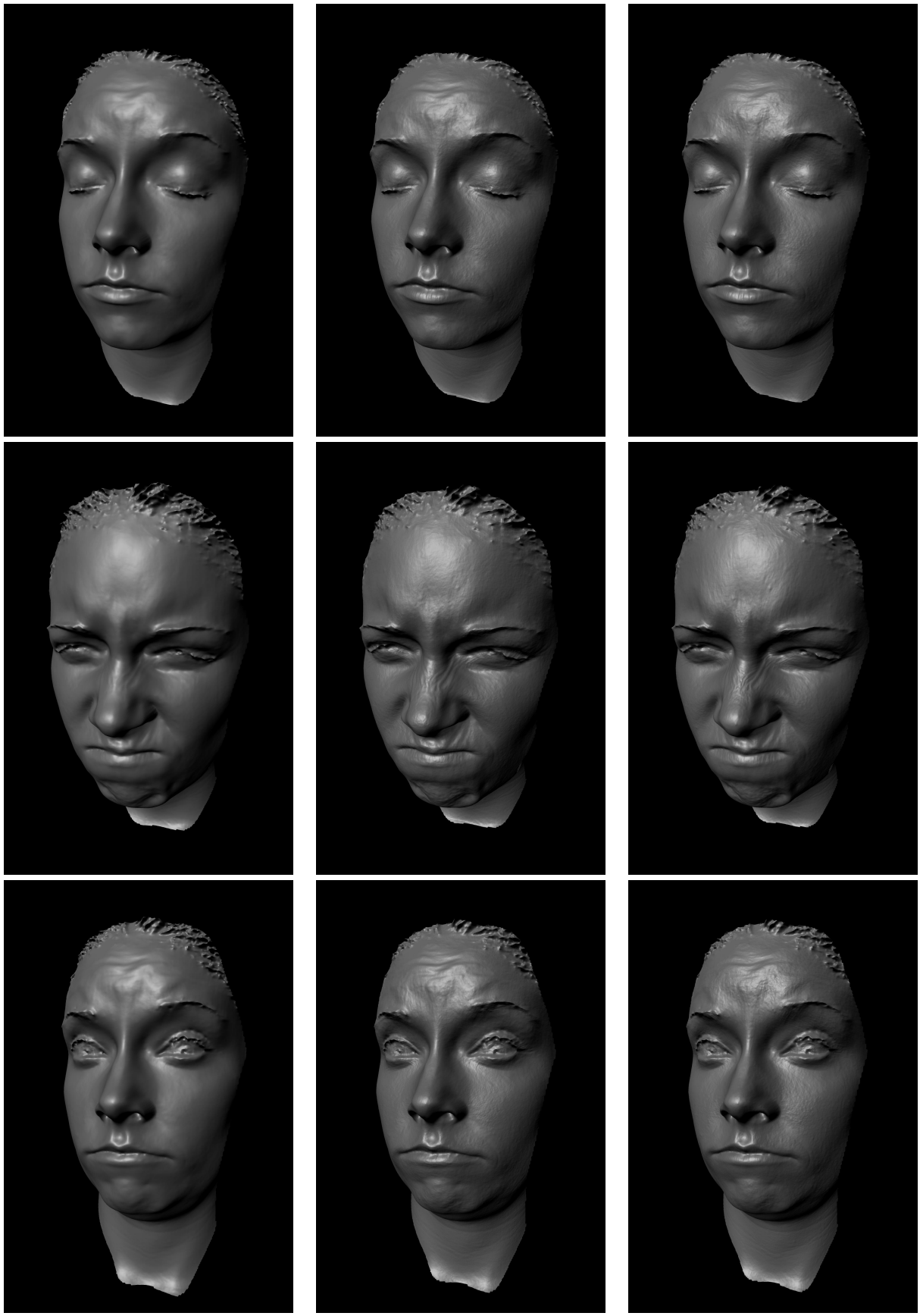
ZHANG, S., AND HUANG, P. S. 2006. High-resolution, real time three-dimensional shape measurement. *Optical Engineering 45*, 12 (December), 123601.

ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. M. 2004. Space-time faces: High resolution capture for modeling and animation. *ACM transactions on graphics 23*, 3, 548–558.



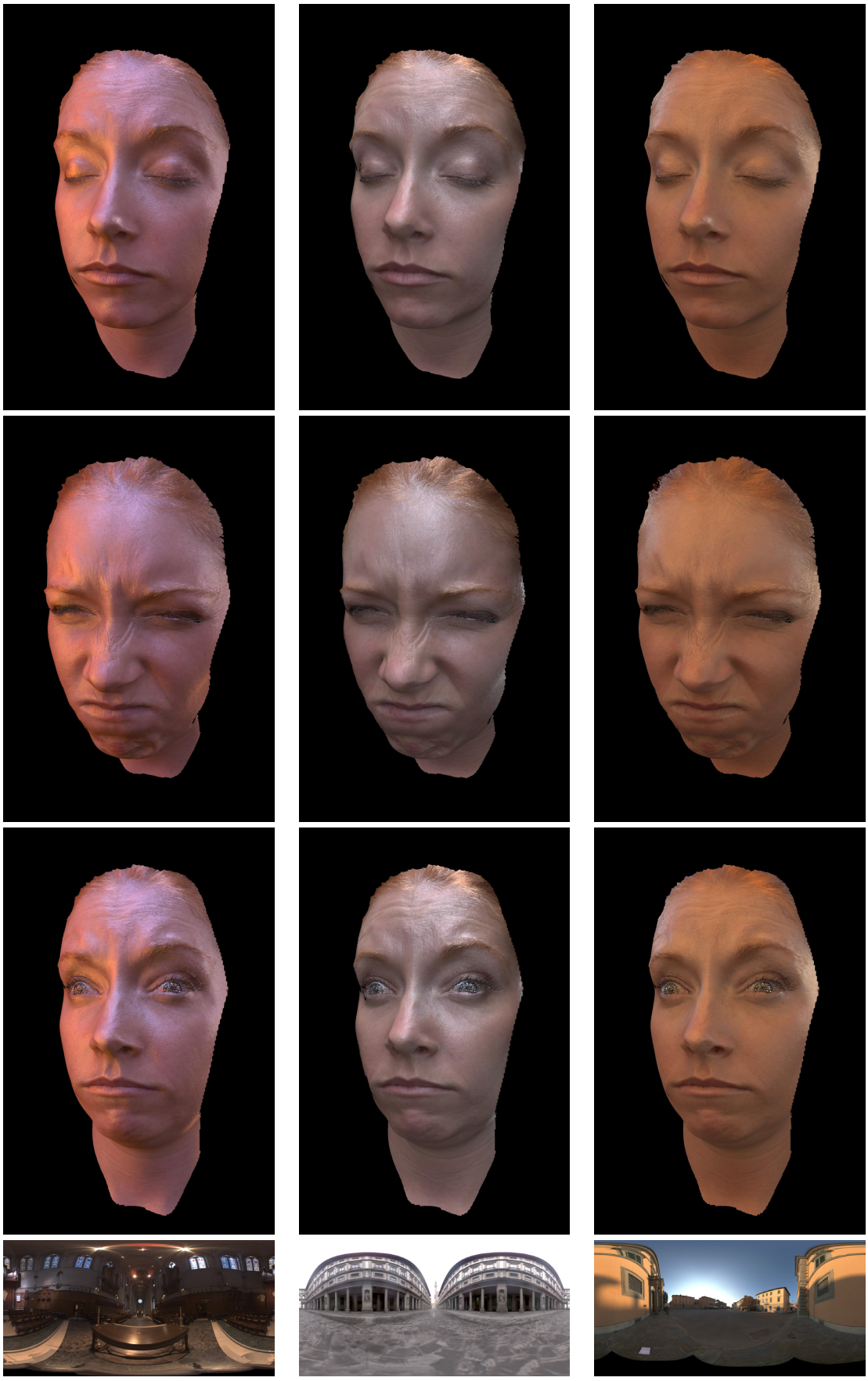
(a) (b) (c)

Figure 2: (a) Diffuse normal map. (b) Synthesized specular normal map. (c) Captured specular normal map. This figure shows the specular normal map synthesis results of the same person with different expressions. The synthesized specular normal map has much clearer detailed features than diffuse normal map, and almost the same as the captured specular normal map.



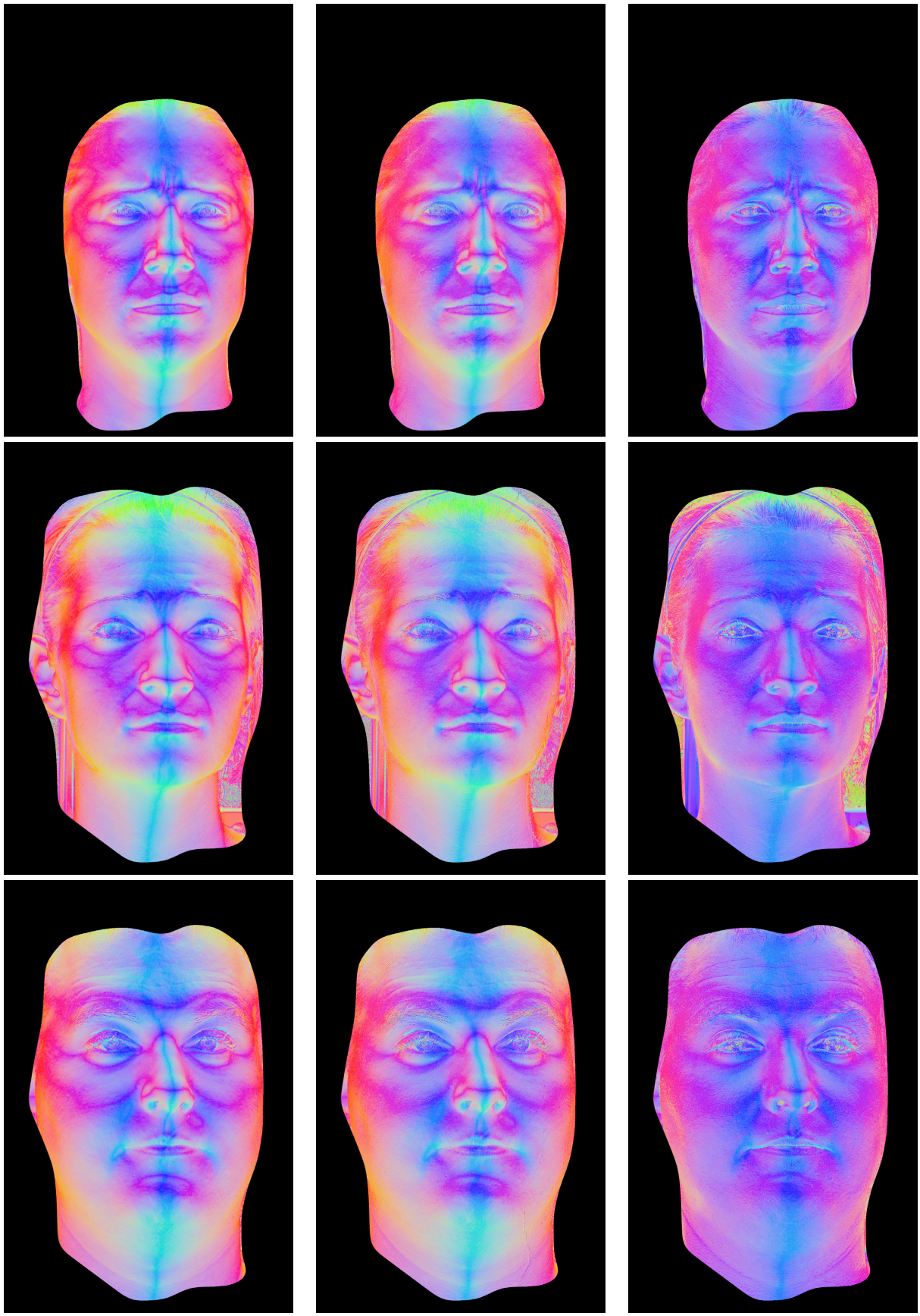
(a) (b) (c)

Figure 3: (a) Base geometry. (b) Geometry optimized with synthesized specular normal map. (c) Geometry optimized with captured specular normal map. This figure shows the geometry optimization results of the same person with different expressions.



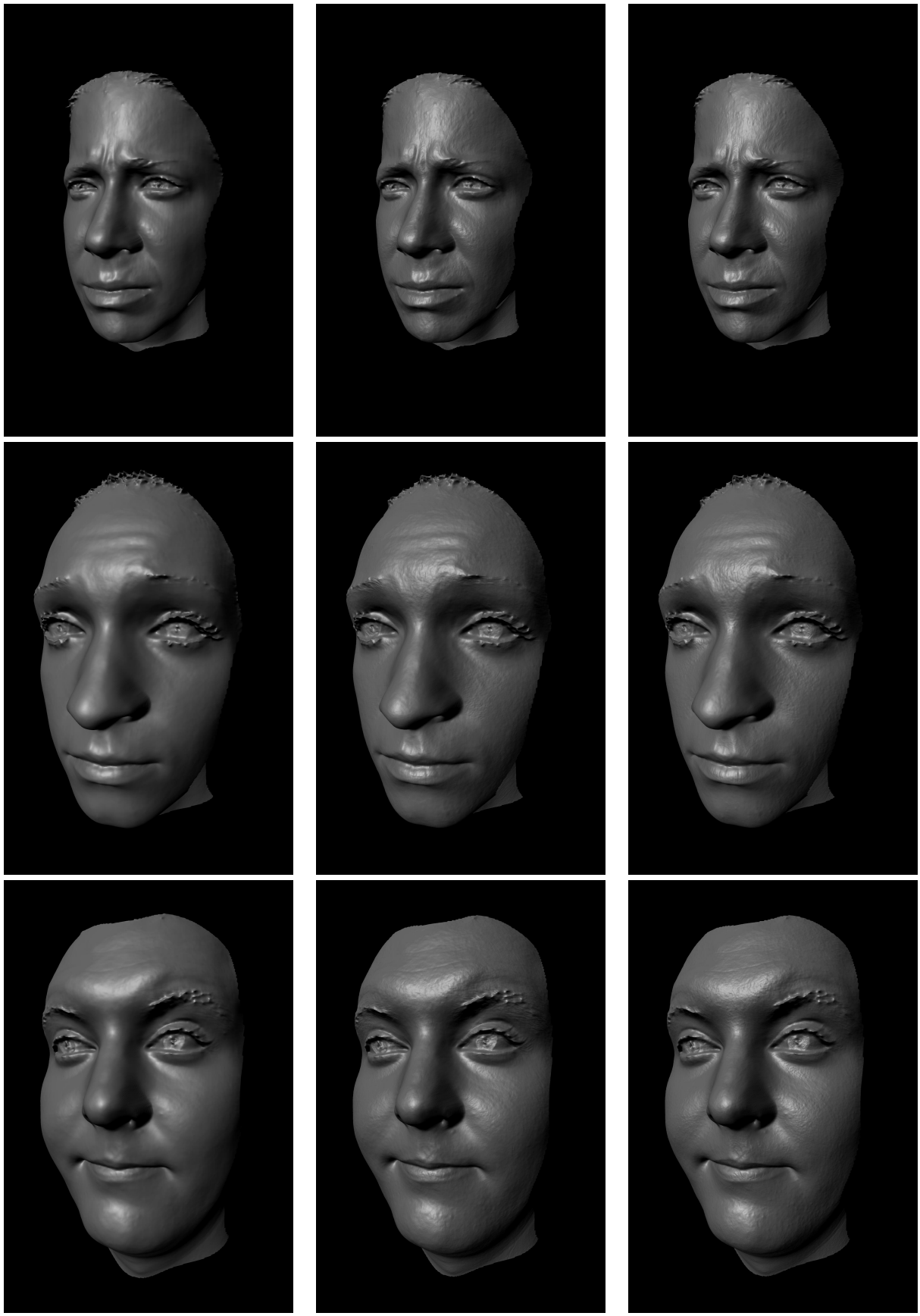
(a) (b) (c)

Figure 4: (a) Image-based lighting rendering results (Grace Cathedral). (b) Image-based lighting rendering results (Uffizi Gallery). (c) Image-based lighting rendering results (Pisa courtyard). This figure shows the image-based lighting rendering results of the geometry optimized with synthesized specular normal map. The results show that our synthesized specular normal maps can also work with image-based lighting techniques.



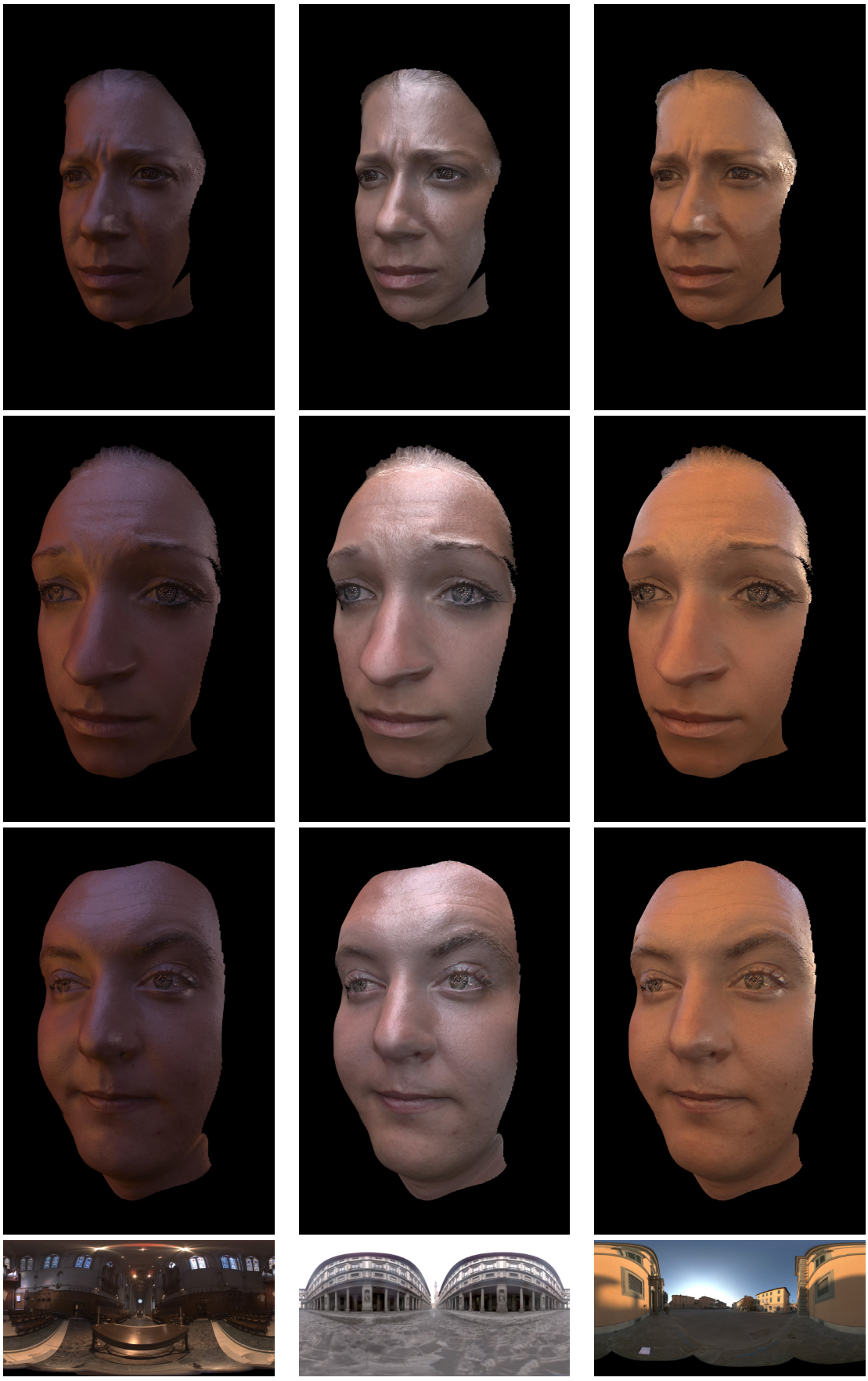
(a) (b) (c)

Figure 5: (a) Diffuse normal map. (b) Synthesized specular normal map. (c) Captured specular normal map. This figure shows the specular normal map synthesis results of cross-object transfer. The algorithm transfers one person's specular information to another person



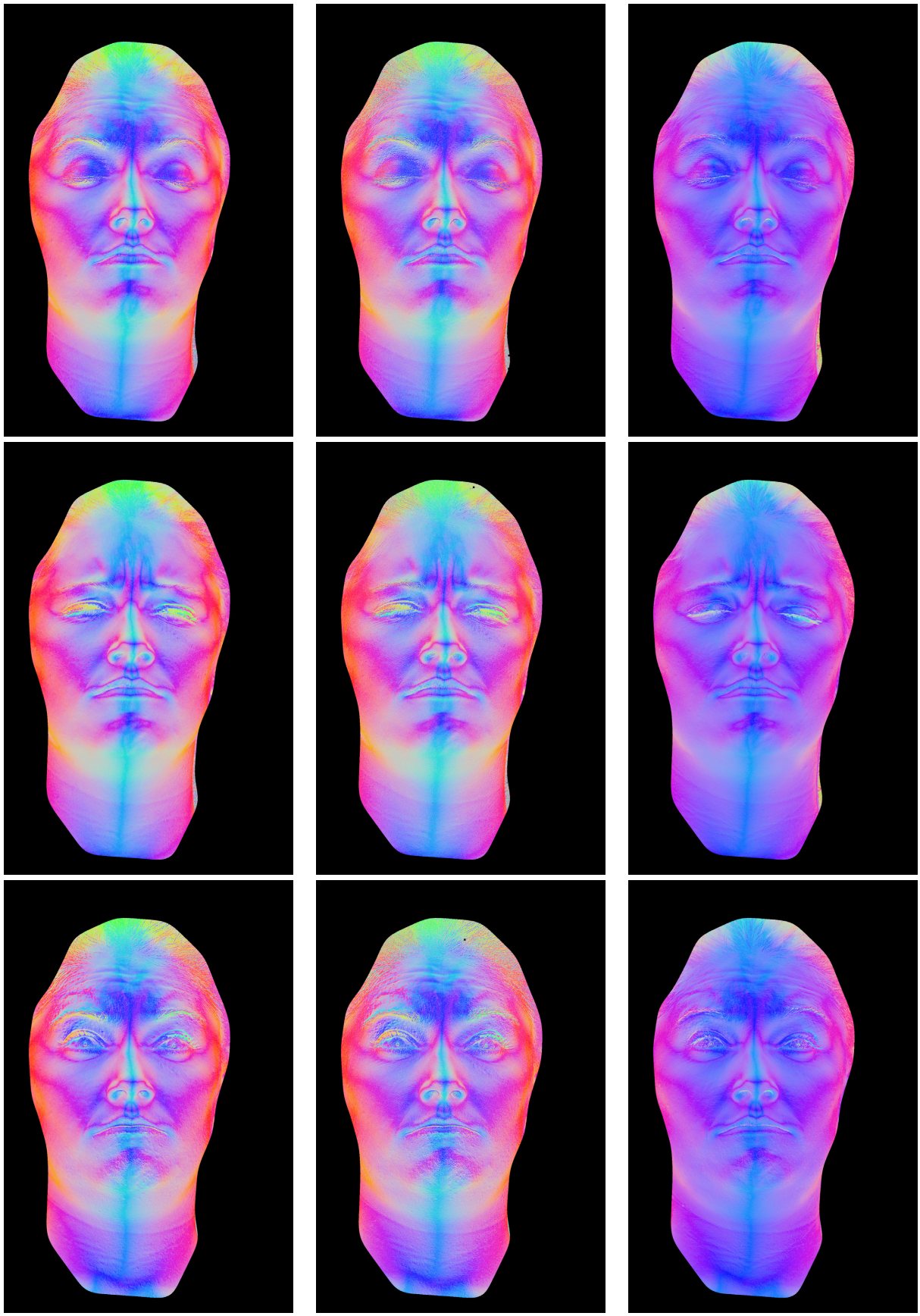
(a) (b) (c)

Figure 6: (a) Base geometry. (b) Geometry optimized with synthesized specular normal map. (c) Geometry optimized with captured specular normal map. This figure shows the geometry optimization results of cross-object transfer.



(a) (b) (c)

Figure 7: (a) Image-based lighting (Grace Cathedral). (b) Image-based lighting (Uffizi Gallery). (c) Image-based lighting (Pisa courtyard). This figure shows the image-based lighting rendering results of the geometry optimized with synthesized specular normal map which is transferred from another person.

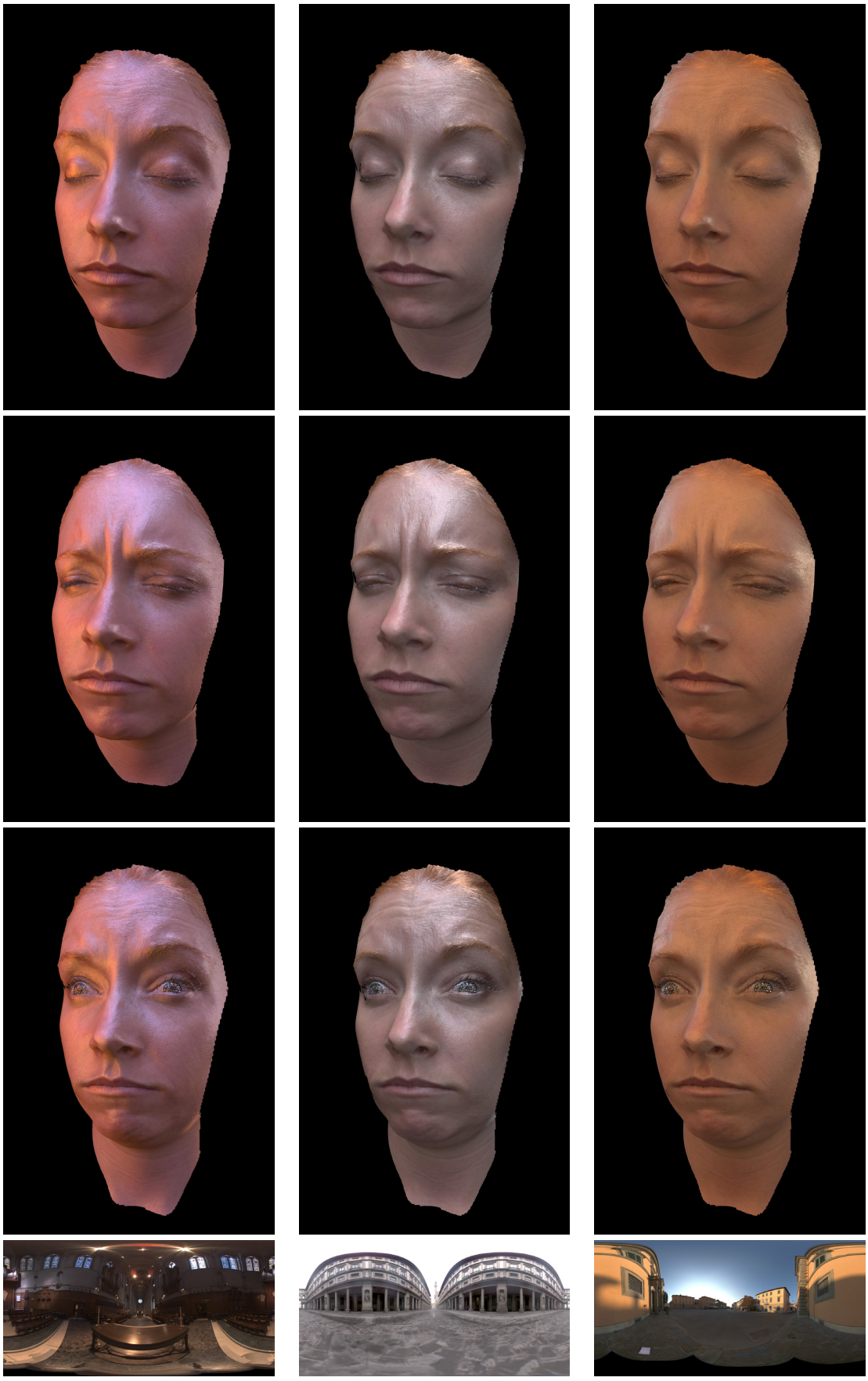


(a) (b) (c)

Figure 8: (a) Mixed normal map. (b) Synthesized specular normal map. (c) Captured specular normal map. This figure shows the specular normal map synthesis results of the same person with different expressions. The synthesized specular normal map are synthesized from unpolarized mixed normal map.

(a) (b) (c)

Figure 9: (a) Base geometry. (b) Geometry optimized with synthesized specular normal map. (c) Geometry optimized with captured mixed normal map. This figure shows the geometry optimization results of the same person with different expressions.



(a) (b) (c)

Figure 10: (a) Image-based lighting rendering results (Grace Cathedral). (b) Image-based lighting rendering results (Uffizi Gallery). (c) Image-based lighting rendering results (Pisa courtyard). This figure shows the image-based lighting rendering results of the geometry optimized with synthesized specular normal map. The results show that our synthesized specular normal maps can also work with image-based lighting techniques.